Deliberation Dialogues for Reasoning about Safety Critical Actions

Pancho Tolchinsky[1]   Sanjay Modgil[2]   Katie Atkinson[3]
Peter McBurney[3] Ulises Cortés[1]

Knowledge Engineering & Machine Learning Group,
Technical University of Catalonia, Spain[1];

Department of Informatics, King's College, London, UK[2]
Department of Computer Science, University of Liverpool,
Liverpool, UK [3]

Abstract

In this paper we present the argument-based model *ProCLAIM*, intended to provide a setting for heterogeneous agents to deliberate over safety critical actions. To achieve this purpose *ProCLAIM* features a Mediator Agent with three main tasks: *1)* guiding the participating agents in what their valid dialectical moves are at each stage of the dialogue; *2)* deciding whether submitted arguments should be accepted on the basis of their relevance; and finally, *3)* evaluating the accepted arguments in order to provide an assessment of whether the proposed action should or should not be undertaken. The main focus in this paper is the proposal of a set of reasoning patters, represented in terms of argument schemes and critical questions, intended to automatise deliberations on whether a proposed action can safely be performed. Framed within the *ProCLAIM* model, we we aim to motivate the importance of these schemes and critical questions for: *a)* the Mediator Agent's guiding task that allows for a highly focused deliberation; *b)* the effective participation of *heterogeneous* agents; and *c)* enabling the reuse of previous similar deliberations in order to evaluate arguments on an evidential basis.

1   Introduction

Safety critical actions such as transplanting an organ to a particular patient or to spill an industrial wastewater discharge require an extra obligation to ensure that no undesirable side effects will be caused, as these side effects may well be the death of the patient or a severe impact on the effluvial ecosystem. To minimise harm, the choice of safety-critical actions is usually governed by guidelines and regulations that direct decision makers on what to do. However, strict adherence to such domain consented knowledge may not always be convenient. For

instance, in the transplant domain, strict adherence to conventional guidelines, regarding the criteria for donor and organ eligibility for transplantation, results in a progressive undersupply of available organs with the result of significantly extended waiting times and increased mortality of those on waiting lists [29]. Domains such as organ transplantation or wastewater management are highly complex and rapidly evolve; thus common consented knowledge cannot always be expected to be up to date and account for all possible circumstances[1]. Hence, decision makers that are experts in these domains, should be allowed to deviate from guidelines, in so far as their decisions are well justified and supported by empirical evidence.

Furthermore, some safety-critical actions require the participation of several agents – experts in different aspects of the problem – for deciding whether or not their performance is safe. For example, an organ available for transplantation is better judged as suitable or not for a given recipient, if experts at the donor site jointly take a decision with the experts at the recipient site, which may be located in a different hospital [27]. Despite the added value of joint decision making among experts, this requirement cannot always be met. Without the appropriate support, the deliberation among experts on whether a proposed action is safe or not is time consuming and has no guarantee of a successful outcome. Thus, any decision support systems intended to assist experts in deciding whether a safety-critical action can be performed without causing severe undesirable side effects, must take into account that:

- Decisions on whether or not to perform a safety-critical action should be well justified.

- Guidelines and regulations are important, but strict adherence to them does not always warrant safety or determine the best decision.

- Empirical evidence plays an important role in safety-critical decision making.

- Decision makers may be experts in the domain. While their decision should be subjected to guidelines, they should be able to deviate from conventional guidelines in some special circumstances.

- Several experts may be required to participate in the deliberation on whether the proposed action is safe or not. Not only because they may have complementary knowledge about the problem at hand, but also because they may represent potentially-diverse interests. In this situation one should take into account that:

  - Decision makers may be in disagreement about whether the action can safely be performed or not.

---

[1]For example, Transplant organisations periodically publish the consented organ acceptability criteria. However, these criteria rapidly evolve because of the researchers' effort in extending them to reduce organ discards. Hence, the more advanced transplant professionals deviate from consented criteria.

- Decision makers, especially human experts, may not be able to maintain long intensive deliberations.

- Participant agents are expected to be heterogeneous. Some agents may be humans while others may be artificial. Furthermore, artificial agents may well be diverse in their implementation given that different agents may be implemented by different developers. In general, in such a setting, a uniform underlying logic cannot be assumed for all participants [36]. It should be noted that by *heterogeneous* agents we do not imply the deliberation occurs in an open environment. Quite the opposite, we expect a highly regulated environment.

In this paper we present an argumentation-based model called *ProCLAIM* that provides a principled way for addressing the above introduced problem. *ProCLAIM* is therefore proposed as a model intended to support experts in a collaborative decision as to whether or not a safety-critical action can be performed. Central to this model is the definitions of policies for the agents' overall interaction, where policies define what can be argued about and how, at each stage of the deliberation. The main purpose and contribution of this paper is the proposal of an argumentation process, based on argument schemes and critical questions, intended to drive the deliberation over safety-critical action among human and artificial agents in a manner which is structured and orderly, and which elicits all the information needed to make such decisions jointly and rationally, even when this information is possessed only by some of the participating agents. A second contribution is that participants in the dialogue need not have any specialised knowledge of argumentation theory, because the framework embeds domain expertise (*e.g.* medical or environmental) in a natural way using application-specific reasoning patterns.

The work presented in this paper builds on and substantially extends works in a number of earlier papers. The key ideas were sketched out in [55], and proposed in the context of a medical agent-based organisation (CARREL) [58] intended to facilitate the offer and allocation of human organs for transplantation. In [56] we proposed the use of scenario specific argument schemes and critical question, tailored for medical applications, to define a protocol-based exchange of arguments which models the agents' deliberation. In [57] we introduced the *ProCLAIM* model and focused on the role of a Case-Based Reasoning component. In [54] we presented a mature version of the above mentioned medical application, and in [53] we described its prototype implementation as the main large scale demonstrator system of the FP6 European project ASPIC[2]. Subsequent work focused on generalising *ProCLAIM* so as to be applicable to domains other than the medical. Key to this was the generalisation of [56]'s scenario-specific schemes and critical questions, sketched in [51], and used in the application of *ProCLAIM* to the environmental domain [52].

In the following subsections we introduce the theoretical context of this work. In §1.1 we introduce some basic concepts in argumentation, particularly focusing

---

[2]http://www.argumentation.org/carrel.htm

on the use of argument schemes and critical questions as a means to define the argument-based interaction among agents. In §1.2 we briefly describe Walton and Krabbe's influential characterisation of the different types of dialogues, where we focus on the deliberation dialogues and their collaborative nature. In §1.3 we introduce the notion of dialogue games, used to define agents' interaction in a dialogue. We conclude this introductory section presenting this paper's organisation.

1.1   Argumentation and Argument Schemes

Recent years have witnessed a growing interest in the use of argumentation techniques for defeasible (non-monotonic) reasoning and conflict resolution in automated systems [44, 8, 43]. Requirements for these modes of reasoning arise when information is incomplete or uncertain, and when different agent perspectives yield choices that may rationally be acceptable to one agent but not to another. In such situations argumentation techniques define the construction of arguments (supporting reasons) for possibly conflicting conclusions. Such arguments are constructed on the basis of underlying knowledge bases or theories. Then, based on the conflict based relations between the constructed arguments, those that are 'winning' or' justified' are evaluated, where the claims of the latter identify the inferences from the underlying theories.

To model the process of argumentation in automated reasoning systems, requires methods that enable our reasoning agents to both generate arguments and proposals about what to believe and what to do, and methods to enable reasoning agents to assess the relative worth of the arguments pertinent to a particular debate, *i.e.*, which arguments are the most convincing and why. Here we set out the main mechanisms that we will use for these purposes: argument schemes and argumentation frameworks.

Argument schemes were introduced in the informal logic literature as a method for argument representation. In particular, the association of schemes with *critical questions* (CQ) that enable systematic identification of how to attack arguments instantiating schemes, was pioneered by Walton [60]. Argument schemes represent stereotypical patterns of reasoning whereby the scheme contains premises that presumptively licence a conclusion. The presumptions need to stand in the context in which the argument is deployed, so they can be challenged by posing the appropriate critical questions associated with the scheme. In order for the presumptions to stand, satisfactory answers must be given to any such questions that are posed in the given situation.

Argument schemes and CQ have been applied in a wide variety of works in AI. Of particular relevance to the work in this paper, is the use of argument schemes and CQ for the definition of a persuasion dialogue game for reasoning about action proposals [4]. Computational accounts of the schemes and CQ approach to argumentation over action have a number of advantages. The schemes and CQ effectively map out the *relevant* space of argumentation, in the sense that for any argument they identify the valid attacking arguments from amongst those that are logically possible. They also provide a natural

basis for structuring argumentation based dialogue protocols. This later work is used as a starting point for defining *ProCLAIM*'s protocol-based exchange of arguments. This is discussed in §5 where we present a structured set of schemes and CQ tailored for deliberating over safety critical actions.

Whilst argument schemes provide us with a means to generate arguments and question them, we also need a mechanism that will enable us to automatically evaluate the arguments and challenges generated in order to determine the ones that are acceptable. For this we make use of Dung's abstract argumentation theory [16]. This theory has proven to be an influential approach to conflict resolution and non-monotonic reasoning over the past decade. The underlying idea is that one is given a directed graph –a so called *abstract argumentation framework (AF)*- consisting of abstract arguments (*i.e.*, no commitment is made to their internal structure) related by a binary attack or defeat relation. The justified status of arguments is then evaluated based on their interactions. This evaluation is in turn based on the notion of an argument being acceptable with respect to a set of arguments if it is not attacked by a member of that set, and all its attackers are attacked by a member of that set.

Numerous subsequent works ([2],[7], [34]) have extended the basic framework so that an attack from an argument A to an argument B can be disregarded if B is for some reason stronger than or preferred to A. Then, the justified arguments are evaluated based only on the *successful attacks* (that are usually referred to as *defeats*). This allows us for example to resolve local disputes between mutually (symmetrically) attacking arguments, so that if A attacks B and B attacks A, then a relative preference over these arguments will determine that one asymmetrically defeats the other.

Here we recall the following basic concepts that were introduced by Dung in [16]:

An Argumentation Framework (AF) is a pair $AF = <AR, Attack>$, where AR is a set of arguments and $Attack \subseteq AR \times AR$ is the attack relationship for AF. A pair $<x,y>$ is referred to as "*x attacks (or is an attacker of) y*" or "*y is attacked by x*". For R, S, subsets of AR, we say that:

- $s \in S$ is attacked by R if there is some $r \in R$ such that $<r,s> \in A$.

- $x \in AR$ is *acceptable* with respect to S if for every $y \in AR$ that attacks x, there is some $z \in S$ that attacks y (*i.e.* z, and hence S, defends x against y).

- S is *conflict free* if no argument in S is attacked by any other argument in S.

- A conflict free set is *admissible* if every argument in S is acceptable with respect to S.

- S is a *complete extension* if S is a subset of AR, S is admissible, and each argument which is defended by S is in S.

- S is a *grounded extension* if it is the least (with respect to set inclusion) complete extension.[3]

As stated above, such frameworks can be depicted as argument graphs. In section §2 we demonstrate how, for our particular application, the interacting arguments define a tree, and how we use such trees to evaluate arguments and attacks generated by our system, and thus ultimately decide on whether or not an action should be performed. From hereon, when we say an argument is *justified* we assume it is under the grounded semantics. One advantage of the grounded semantics is that computing its extension is a linear problem and that the extension always exists, though it may be empty. For this reason the grounded semantics has been argued to be too skeptical. While skepticism may yield well for reasoning over safety-critical actions it may be too restrictive for other applications [17].

Let us suppose $< AR, Attack >$ is an argumentation framework and G the grounded extension. Then, if $x \in AR$:

- x is said to be **justified** if $x \in G$.

- x is **defeated** if there exist an argument $y \in \mathbf{G}$ such that $(y, x) \in Attack$.

- Otherwise, x is said to be **defensible**.

Now, if the argument proposing the safety critical action is evaluated as *justified*, the action is recommended as safe. If evaluated as *defeated*, the action is deemed unsafe. Otherwise, no conclusive recommendation can be given. However, as we discuss in §7, a *ProCLAIM* proposed solution will highlights the relevant issues that must be resolved before taking the final decision.

The approaches to argument modelling described above form the basis for some of the elements in our system that are used to generate and evaluate arguments. However, we also need to specify how these models will be used within the context of a dialogue, as discussed next.

1.2 Deliberation Dialogues

As defined by Walton and Krabbe in their influential classification of human dialogues [61], deliberation dialogues involve participants deciding what action or course of actions should be undertaken in a given situation. Typically participants in such a dialogue believe that they share a responsibility for deciding what to do, which provides a collaborative context to the dialogue. The Walton and Krabbe classification is based on the prior knowledge and goals of the individual participants, and includes five other major types of dialogues. Thus, in **Information-Seeking Dialogues** a questioner seeks to discover an answer to a factual question from another participant, whom is believed by the questioner to know the answer to the question. In **Inquiry Dialogues** participants

---

[3]We refer the reader to [16] for definition of extensions defined under stable and preferred semantics

jointly seek to answer a factual question whose answer may not be known to any of them beforehand, or may require sharing of the partial knowledge each participant has. In **Persuasion Dialogues**, a participant, typically called the *proponent*, seeks to persuade his or her fellow participant(s) to endorse some proposition. Participants in a **Negotiation Dialogue** seek to agree a division of a scarce resource (which may be the participants' time) between potentially conflicting claims over it. If participants each seek to maximise their share of the resource, then their dialogue will require some form of bargaining or compromise in order to reach an agreement. And finally, in Eristic Dialogues, participants seek to vent perceived grievances with one another, and the dialogue may act as a substitute for physical fighting. Negotiation and Deliberation Dialogues involve discussions over actions, while Information-Seeking and Inquiry Dialogues involve discussions over beliefs. Persuasion Dialogues may concern either beliefs or actions.

In addition to this focus on action, all Deliberation Dialogues normally share some other characteristic features. The first is the absence of any fixed positions by the participants at the start of the dialogue. Unlike Persuasion Dialogues, for example, participants in a Deliberation Dialogue do not (at least, initially), seek to persuade another participant to endorse some statement or proposal for action. Indeed, the focus, or governing question, may change in the course of a Deliberation Dialogue, as participants explore the space of possible actions and examine the features and consequences of these actions. During the course of a Deliberation Dialogue, of course, participants may engage in a Persuasion Dialogue (or indeed any one of the Walton and Krabbe types), which may motivate to model such a Persuasion Dialogue as embedded inside the Deliberation Dialogue. A second important feature of a Deliberation Dialogue is the mutual focus, which distinguishes these dialogues from, say, Negotiation Dialogues. Participants in a Deliberation Dialogue do not necessarily seek to agree a course of action which accommodates or reconciles their various different interests. It may be that the participants *do* in fact aim to do this, and it may even be that they succeed in doing so, but these aspects are not essential features of a Deliberation Dialogue. Although in a Deliberation Dialogue, participants exchange proposals and express their positions about what is to be done, their shared goal is to jointly reach agreement on the best or most sensible action to decide to do (in these circumstances, at this time, by the designated actors, under these constraints), rather than to defend a particular position. For a fuller discussion of the characteristic features of Deliberation Dialogues, see [31].

1.3 Dialogue Games

Dialogue games are interactions between two or more participants who 'move' by uttering locutions, according to certain rules. They were first studied by Aristotle [3] and then by argumentation theorists and logicians in the post-war period (eg, [23, 28]). Over the last decade they have been applied in various areas of computer science and artificial intelligence, particularly for rule-governed interactions between software agents; for a recent review of such applications,

see [33]. A dialogue game may be specified by listing the legal locutions, together with the rules which govern the utterance of these locutions, the opening and termination of dialogues, and the rules for manipulation of any dialogical commitments incurred by the participants during a dialogue [32].

Our work is informed by the dialogue-game framework for ideal Deliberation Dialogues proposed by McBurney, Hitchcock and Parsons in [31]. In that framework, deliberation dialogues may proceed through eight successive stages: *Open, Inform, Propose, Consider, Revise, Recommend, Confirm* and *Close Stages.* The goals of participants in these dialogues change according to the stage of the dialogue. It is for this reason that stages are marked explicitly, so as to better enable participants to know what is expected of them at each stage. In this framework, there are some constraints on the order of the stages, and some stages may be repeated, in an iterative fashion, before the dialogue is completed. This approach is taken as a starting point for defining *ProCLAIM*'s dialogue game which we introduce in §4.

1.4   Document Organisation

In the following section we introduce the *ProCLAIM* model. In §3 we introduce one of our case studies – human organ transplantation – which will be used throughout this paper to illustrate the agents' interaction; particularly the exchange of arguments. In §4 we present the model's deliberation dialogue game that will define the context for the agents' overall interaction. In section §5 we focus on the argumentation, we first define the internal structure of *ProCLAIM*'s arguments in §5.1, to then in §5.2 introduce the model's protocol-based exchange of arguments based on schemes and CQs tailored for deliberating over safety critical actions. In §6 we illustrate how the schemes can be further specialised for a particular application and how these more specialised schemes and CQ are used to guide agents in their deliberation facilitating the interaction for both human and artificial agents. In §7 we briefly discuss the model's argument evaluation. In §8 we discuss how the defined argument schemes and CQs facilitate the reuse of previous deliberations for evaluating arguments on an evidential basis. Finally, we conclude in §9.

2   The *ProCLAIM* Model

The *ProCLAIM* model is intended to provide a setting for heterogeneous agents to efficiently and effectively deliberate over whether a safety-critical action can safely be performed.*ProCLAIM* can be regarded as defining a centralised medium through which heterogeneous agent are directed in a deliberation over the safety of a proposed action. This medium is tailored for the deliberation purpose and is intended to focus the discussion on the relevant matters, while keeping track of the participants' submitted arguments and evaluating them to propose a solution to the addressed problem. This centralised medium is embodied by a Mediator Agent (*MA*) whose role is to ensure the success of the

deliberation process, enabled by the MA's access to a number of knowledge resources, which we now introduce.

The setting *ProCLAIM* provides for participant agents to efficiently deliberate is best described by defining the mediator agent's tasks:

- **Direct participants** on what argument-based moves (argument schemes or critical questions) they can submit at each point of the deliberation. Thus, for each argument a participant wants to reply to, she is given a set of schemes that she can instantiate and submit as a valid attack, in so far as the instantiation is appropriate. A participant may also challenge some of the submitted arguments and, in turn, a participant may answer to these challenges with the instantiation of the appropriate argument scheme.

- **Validate the incoming arguments** in order to exclude arguments that may jeopardise or disrupt the course of the deliberation. While agents are given the schemes to instantiate, there is no guarantee that the instantiation will be relevant for the discussion. Thus, one of MA's tasks is to discern between relevant and non-relevant instantiations, so as to keep the deliberation highly focused on only the important matters. Each validated argument and challenge is added to a tree of interacting arguments whose root is the argument proposing the initial action.

- **Submit additional arguments** that introduce new factors not taken into account by the participants but that either guidelines and/or evidence associated with previous similar cases indicate as relevant. Thus, for example, if α is taken as a fact, and guidelines indicate that α is a contraindication for performing the proposed action, but, for some reason no participant highlights this, the MA will submit an argument against the action proposal indicating that there is a contraindication α for its performance. This argument will be added to the tree of interacting arguments.

- **Evaluate the tree of interacting arguments** so as to propose a solution to whether the proposed action is safe or not. A solution is proposed by means of assigning a preference between conflicting (mutually attacking) arguments and then evaluating the justified arguments as defined by Dung's theory.

In order to perform the above introduced tasks, the MA references four knowledge resources, as shown diagrammatically in Figure 1 and also described below:

Argument Scheme Repository (ASR): In order to direct the participant agents in the submission and exchange of arguments, the MA references a repository of argument schemes and their associated critical questions. The schemes and critical questions are instantiated by agents to construct

arguments, and effectively encode the full 'space of argumentation', *i.e.*, all possible lines of reasoning that should be pursued w.r.t a given issue. The repository is structured in such a way that it defines the protocol-based exchange of arguments. Thus, given an argument (that instantiates a scheme in ASR) the repository returns the schemes that agents can instantiate in reply to the former argument (as well as the critical questions used to challenge it).
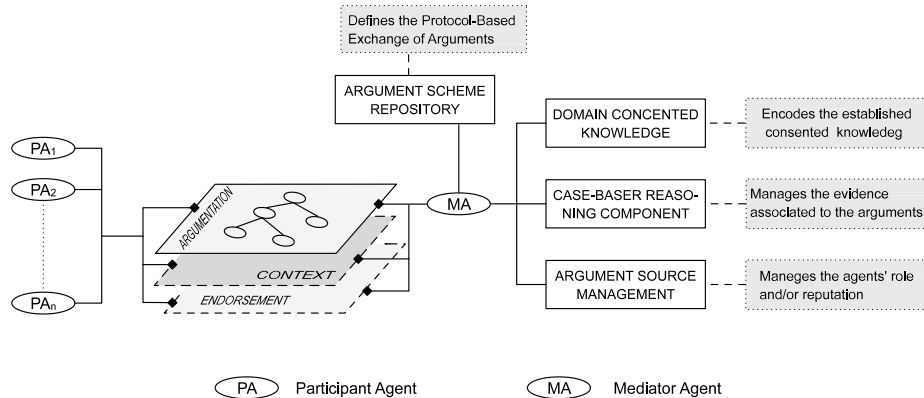

Figure 1: *ProCLAIM*'s architecture.

Domain Consented Knowledge (DCK): This component enables the MA to check whether the arguments comply with the established knowledge, by checking what the valid instantiations of the schemes in ASR are (the ASR can thus be regarded as an abstraction of the DCK). This is of particular importance in safety critical domains where *1)* one is under extra obligation to ensure that spurious instantiations of argument schemes should not bear on the outcome of any deliberation; and *2)* guidelines usually exist in such domains and so should be taken into account when evaluating the submitted arguments. The MA also references the DCK in order to check whether any known factor is not being addressed by the participants (experts) in spite of being deemed relevant from the view point of the guidelines. In such a case, the MA uses the DCK in order to submit additional arguments, which account for these neglected, but relevant, factors. In this last sense, the MA can be regarded as a participant expert in guidelines.

Case-Based Reasoning Component (CBRc): This component enables the MA to assign a preference relation between mutually attacking arguments (*i.e.* resolve conflicts amongst pairs of arguments) on the basis of their associated evidence gathered from past deliberations. The CBRc also provides additional arguments that were deemed relevant in previous similar

situations and are applicable in the current target problem. Again, in this last sense, the MA plays the role of an expert, or specialist in collecting evidence from previous deliberations.

Argument Endorsement Manager (AEM): Depending on who endorses an argument, the strengths of arguments may be readjusted by the MA. Thus, this component manages the knowledge related to, for example, the agents' roles and/or reputations.

Broadly speaking, a deliberation in $ProCLAIM$ begins with one of the agents[4] submitting an argument proposing an action (*e.g.* transplant an available organ to a particular recipient). The MA will then guide the participant agents in the submission of further arguments that will attack or defend the justification given for the proposed action. Each submitted argument (or challenge) instantiates a scheme (or critical question) of the ASR. Hence the MA references the ASR in order to indicate which are the schemes or critical questions the participants can instantiate in reply to each of the submitted arguments or challenges.
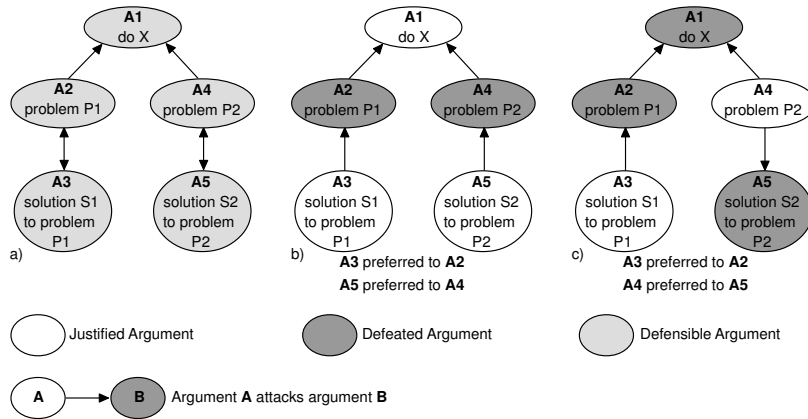

Figure 2: Resolving a tree of interacting arguments in order to decide whether or not to perform X. In figure *a)* no solution can be proposed since it is still undecided as to whether the respective solutions address the respective problems, as indicate by the symmetric attacks between A2 and A3, and A4 and A5. In figures *b)* and *c)* the solutions are, respectively, to perform action X and not to, depending on the arguments' preference assignment.

A submitted argument, if legal (instantiates an appropriate scheme), is evaluated by the MA in order to determine whether the instantiation of the scheme is a valid one. This is done by the MA referencing the DCK, CBRc and AEM. If an argument is compliant with guidelines, *i.e.* validated by the DCK, the argument is accepted and added to the tree of interacting argument which we

---

[4]This agent may well be the MA if the action is a default one.

denote as $\mathbb{T}$. If the incoming argument is not validated by the DCK it may still be accepted if either the CBRc indicates the argument has previously been accepted or the AEM indicates that the submitter is sufficiently reliable so as to exceptionably accept this argument. In either case the argument is added to the tree of arguments $\mathbb{T}$ and the MA broadcasts this new argument to all participants together with the schemes they can instantiate in reply. If the argument is not accepted by the knowledge resources, the MA informs the submitter of the reasons for it being rejected. Other approaches have considered the relevance of an argument in terms of its impact, or potential impact, on the dialectical status of acceptability of the already posed arguments [40, 39, 30]. Broadly speaking, if an argument to submit will have little or no effect on the acceptability status of the existing arguments, it may be deemed irrelevant and so may not be submitted[5]. This approach to the relevance of an argument could be used in *ProCLAIM* for example to stop the deliberation once it is clear that there are enough arguments to deem the proposed action unsafe.

The agents' interaction is described in §4 in terms of set of stages and interaction layers. For instance, the exchange of arguments is defined within the *Argumentation Layer* of the *Deliberation Stage*. As depicted in Figure 1, besides the exchange of arguments, within the *Deliberation Stage* we define other interaction layers, such as the *Context Layer* in which participants submit the facts and complementary actions they believe to be potentially relevant for the deliberation. At the *Endorsement Layer*, participants indicate which arguments in $\mathbb{T}$ they endorse.

Once all participants have no more arguments to exchange, have submitted all the facts they deem relevant and have informed of the arguments in $\mathbb{T}$ they endorse, the deliberation moves into the *Resolution Stage* in which $\mathbb{T}$ is evaluated. As depicted in figure 2a) $\mathbb{T}$ may contain arguments that mutually attack each other preventing a definitive solution. Thus, to evaluate $\mathbb{T}$ MA has to assign a preference relation between the mutually attacking arguments, and so change the symmetric attacks into asymmetric ones. Once this is done, MA applies Dung's evaluation of the justified arguments to propose a solution (Figure 2). In order to assign this preference between mutually attacking arguments the MA again references the DCK, the CBRc and the AEM. From each resource the MA derives a preference assignment. These may all be in agreement (*e.g.* $A3$ preferred to $A2$) or not, *i.e.* some prefer one argument while another knowledge resource prefers the other argument. The MA's task, is to provide a solution that accounts for each of the knowledge resources' preference assignment. So a solution in which not all resources agree could be of the type: -While guidelines indicate that $S2$ is not a solution to problem $P2$, trustworthy experts argue that $S2$ is a solution to $P2$ and this position is weakly backed up by evidence. On the basis of this information, the person responsible for the final decision will decide whether or not to perform action X.

Many works addressing cooperative environments (*e.g.* production and lo-

---

[5]Also interesting is the work presented in [14], where a pruning of which arguments to account for is made in order to compute dialectical trees efficiently.

gistics [18, 49]) feature an agent, or set of agents, dedicated to coordinate the tasks performed by the different working components or agents. In argumentation, the role of the mediator agent is usually associated with negotiation dialogues [13, 50, 38], where the main objective of the mediator agent is to help reconcile the competing agents' positions, for which the mediator agent usually relies on mental models of the participants. More relevant here is the conceptual framework for negotiation dialogues SANA [38], that proposes a number of so called artifacts, such as a social Dialogue Artifact, that acts as a mediator which regulates the agents dialogue interaction and a social Argumentation Artifact that can be regarded as a sophisticated commitment store, where the agents' submitted arguments, as well as other arguments that may be publicly available, are organised and their *social* acceptability status can be evaluated following different algorithms. Similar to our approach, the SANA framework defines, as part of the social Dialogue Artifact, an Argumentation Store (AS) that stores a collection of *socially acceptable* arguments. The main difference being, that while a central part of *ProCLAIM* is the definition of the structure and relation of the schemes in ASR tailored for the specific purpose of deliberating over safety critical actions, the SANA's AS is presented as a placeholder for any argument scheme, that is, developers are given little guidance on which argument schemes the AS should encode. In a broader view, the SANA approach is similar to that proposed in the FP6-European project ASPIC, where a set of generic components were developed: (Inference Engine, Dialogue Manager, Learning Component, Decision-Making component) that can be plugged into an agent in order to add argumentation capabilities. This is the approach undertaken in [53] to implement *ProCLAIM* in the transplant scenario.

Due to the critical nature of the intended scenarios, *ProCLAIM* assumes a rather regulated environment. In particular, *ProCLAIM* does not address any of the normative aspects that would naturally be associated with a safety critical environment. It also assumes that issues such as information privacy, or foreign attacks from malicious agents are also resolved. A good example of the context in which *ProCLAIM* can be used is the transplant scenario we now introduce, where the model is used to extend an existing agent-based organisation [58] so that agents can deliberate over the viability of an available human organ.

3   The Transplant Scenario

The shortage of human organs for transplantation is a serious problem, and is exacerbated by the fact that current organ selection processes discard a significant number of organs deemed non-viable (not suitable) for transplantation. The organ viability assessment illustrates the ubiquity of disagreement and conflict of opinion in the medical domain. What may be a sufficient reason for discarding an organ for some qualified professionals may not be for others. Different policies in different hospitals and regions exist, and a consensus among medical professionals is not always feasible. Hence, contradictory conclusions may be derived from the same set of facts. For example, consider a donor with a

smoking history but no *chronic obstructive pulmonary disease* (COPD). The medical guidelines indicate that a donor's smoking history is a sufficient reason for deeming a donor's lung as non-viable. However, there are qualified physicians that reason that the donor's lung is viable given that there is no history of COPD [27]. Similarly, the guidelines suggest discarding the kidney of a donor whose cause of death was *streptococcus viridans endocarditis* (*sve*). However, by administrating *penicillin* to the recipient this means that the kidney can safely be transplanted.

Currently, the decision to offer or discard an organ available for transplantation, is based solely on the assessment of doctors at the donor site (Donor Agent, DA). This organ selection process does not account for the fact that: *1)* medical doctors may disagree on whether an organ is viable or non-viable; *2)* different policies in different hospitals and regions exist, and; *3)* viability is not an intrinsic property of the donor's organ, but rather, an integral concept that involves the donor and recipient characteristics as well as the courses of action to be undertaken in the transplantation process [27]. In particular, current organ selection processes allow for a DA to discard an organ that doctors at the recipient site (Recipient Agents, RA) may claim to be viable and, given the chance, could provide strong arguments to support this claim.

In [54] a novel organ selection process is proposed in which *ProCLAIM* is used to coordinate joint deliberation between donor and recipient agents in order to prevent the discard of organs due to the application of inappropriate organ acceptability criteria, and so help to reduce the disparity between the demand for and supply of organs. This proposal is framed within an agent-based organisation called *CARREL* [58], intended to efficiently manage the data to be processed in carrying out recipient selection, organ and tissue allocation, ensuring adherence to legislation, following approved protocols and preparing delivery plans.

The *ProCLAIM* model is thus used to extend *CARREL* so that DA and RA can effectively deliberate over the viability of an available organ. In short, *ProCLAIM* is instantiated as follows: the participant agents are the DA and RA, the Guideline Knowledge encodes the donor and organ acceptability criteria consented by the transplant organisations, *i.e.* the criteria the medical doctors should refer to when deciding the organs' viability.[6] The AEM relates to the agents' reputation. Namely, the MA may deem as stronger the arguments endorsed by agents with good reputations (*e.g.* a RA representing a prestigious transplant unit). Finally, the CBRc allows the MA to evaluate the submitted arguments on the basis of past recorded transplantation cases.

In [53] a prototype of this model is presented and in [57] we focus on the CBRc component. These two works assumed our first attempt to formalise an ASR [56] in which the argument schemes were formalised and constructed in a somewhat *ad-hoc* fashion, so hindering the application of *ProCLAIM* in new

---

[6]Transplant organisations periodically publish the consented organ acceptability criteria. However, these criteria rapidly evolve because of the researchers' effort in extending them to reduce organ discards. Hence, the more advanced transplant units deviate from consented criteria.

scenarios (such as in the environmental scenario presented in [52]). For this reason, in [51] we proposed a domain independent approach for the definition of the argument schemes and that we now (in §5) describe in much more detail and with some corrections and extensions. Furthermore, the prototype presented in [53] implements an interaction protocol for persuasion dialogues [41], which, as discussed in §1.2, is not always appropriate for collaborative decision making. In the following section we present *ProCLAIM*'s deliberation dialogue.

4   *ProCLAIM*'s Deliberation Dialogue

In this section we describe the interaction protocol that governs *ProCLAIM*'s dialogue. In each exchanged locution we distinguish three interaction levels: *1)* On the deepest level there is the content of the message, *e.g.* the submitted arguments. *2)* Each of these messages is wrapped in the appropriate deliberation locution defined by the dialogue game (*e.g.* an argument is wrapped in an *argue* locution); *3)* in turn, each of these deliberation locutions is wrapped in either an *inform* or *request* locution. This is because the Participant Agents (PAs) always interact with the MA, never with other PAs. They submit a request to, for example, enter the dialogue, submit a new argument or add new information. The MA then decides whether to accept or reject their request. Thus, the MA acts as a proxy for the PAs (see figure 3a.)

In the following subsection we describe the inform-request interaction which we call the *proxy* dialogue game. And in §4.2, we introduce *ProCLAIM*'s deliberation dialogue game where we define the locutions that can be submitted at each stage and layer of the deliberation. While the provided description of the dialogue game is quiet detailed, to further facilitate its implementation, we intend for future work define the dialogue game's axiomatic semantics defining the *pre* and *post* conditions for each dialogue move.

4.1   *Proxy* Dialogue Game

Agents participate in the deliberation via the MA, which decides whether an incoming message should be accepted or rejected. Messages are obviously rejected if syntactically ill-formed, but also if the content is not appropriate. For example, a submitted argument may be rejected if the MA deems it non relevant for the deliberation. For that reason, each participant message is wrapped in a `request` locution to which the MA replies with an `inform` locution, either to inform of its rejection (and *why* it is rejected) or to act upon the request. For example, if the request is to enter the dialogue, MA will inform of the participant's acceptance, along with the extra information required for the appropriate participation. The MA may also send an `inform` locution without prior requests, *e.g.* to inform of a *time-out* constraint which forces the deliberation to conclude.

`request(pa_id, ma, conv_id,msg_id, target_id, R):` where `pa_id` is the sender's id (a PA), `ma` is the receiver agent (the MA), `conv_id` is the

conversation id, `msg_id` is the message identifier, `target_id` is the message to which this locution is directed (when the message is not directed to any specific message, `target_id` should be set to `-1`). `R` is a variable denoting the content being communicated in the request locution. The possible values of `R` are discussed in the following subsection.

`inform(ma,PA,conv_id,msg_id,target_id, I)`: Here, the locution may be addressed to a single receiver, in which case `PA` is `pa_id`, or it may be broadcast to all the participants, in which case `PA` is `all`, or to a subgroup of PAs, *e.g.* to all but the sender of the request `all - {pa_id}`. `I` is a variable denoting the content being communicated in the inform locution, which may be in reply to a request of a PA's request.

In the following subsection we define the messages' content, *i.e.*, the `R` and the `I`.
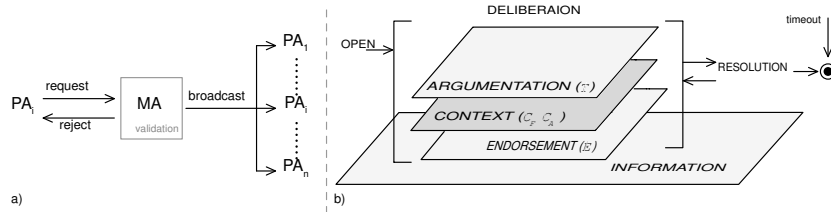
Figure 3: *a)* Illustrating the proxy dialogue game; *b)* Depiction of *ProCLAIM*'s deliberation dialogue game with its stages and interaction layers.

4.2 The Deliberation Dialogue Game

In this subsection we introduce *ProCLAIM*'s deliberation dialogue game. That is, we introduce the legal locutions, together with the rules which govern their use as well as the commencement and termination of the deliberation dialogue. As illustrated in 3a the deliberation dialogue game can be subdivided in three stages: *Open*, *Deliberation* and *Resolutions*. While the Deliberation Stage can in turn be subdivided in three layers: *Argumentation*, *Context*, *Endorsement* layers. The moves in these three layers may happen in parallel and the moves at one layer may have an effect on another layer. As depicted in 3b we define yet another interaction layer, called *Information layer* in which PAs can request the MA for updates on ongoing deliberation. It is worth noting that these distinct layers and stages are conceptual and are used as metaphors to better organise the dialogue game, *i.e.* the PAs need not know of these distraction in order to effectively participate.
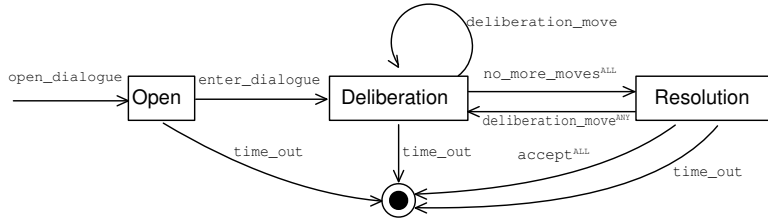
Figure 4: Deliberation dialogue, stage transition. We use `deliberation_move` to indicate any dialogue move defined in the *Deliberation Stage* (*e.g.* `assert`, `argue`, `endorse`...). Note that to move from the *Deliberation Stage* into the *Resolution Stage* all PA must have submitted the `no_more_moves` move. Similarly, to conclude the deliberation all PA must have submitted the `accept` move, unless the `time_out` has been triggered. If any PA submits a deliberation move, while being a the *Resolution Stage*, the dialogue moves back to the *Deliberation Stage*.

4.2.1 Open Stage:

The first stage is *Open* in which the proposal is made, the participants are introduced and basic available information is provided to all participants:

`open_dialogue(proposal)`: `proposal` is an argument proposing the main action (*e.g.* transplant an available organ). `proposal` also contains the preconditions for the action's performance (*e.g.* an available organ and a potential recipient). As we see in §5, `proposal` is an instantiation of an argument scheme.

If the proposal is made by a PA (and not by the MA), this message is wrapped in a `request` locution that the MA would have to validate. If the request is validated, the MA will submit the `open_dialogue` locution and contact the potential participants in order to enter the dialogue.

`enter_dialogue(proposal,role, basic_info)`: Each agent willing to participate in a deliberation over `proposal` will enter her `role` in the deliberation (*e.g.* *donor* or *recipient* agent) and the information (`basic_info`) she deems potentially relevant for the decision making (*e.g.* the patient's clinical record). This message is wrapped in a `request` locution.

If the `enter_dialogue` is accepted, the introduced facts, via `basic_info`, will be stored in a set of facts, which we denote $\mathbb{C}_F$.[7] Similarly, we define a second set denoted as $\mathbb{C}_A$ which contains the agents' proposed actions. Initially $\mathbb{C}_F$ contains only the facts introduced in the `enter_dialogue` locution and the preconditions introduced in the argument proposing the main action.

---

[7]We will assume that all facts introduced by the PAs are consistent, we discuss this in more detail in §9.

When the deliberation starts $\mathbb{C}_A$ contains only the initially proposed action (*e.g.* transplant the available organ to the potential recipient).

For simplicity, let us denote $\mathbb{C}_F$ and $\mathbb{C}_A$ together as $\mathbb{C}_{F\wedge A}$. During the deliberation $\mathbb{C}_{F\wedge A}$ may be updated, this happens at the *Context Layer* (§4.2.2).

The proposal `proposal` made in `open_dialogue(proposal)` is an argument for the main action. Thus, this is the first argument added to the tree of arguments $\mathbb{T}$. Further submitted arguments at the *Argumentation Layer* (§4.2.3) will update $\mathbb{T}$.

A PA may request to enter the dialogue at the beginning of the deliberation, or later, when both $\mathbb{C}_{F\wedge A}$ and $\mathbb{T}$ may have more information than the minimal information available at the beginning. Thus if at any stage an agent's request to participate is accepted, the MA will reply by broadcasting the following message.

`entered_dialogue(proposal, role, basic_info, pas, `$\mathbb{C}_{F\wedge A}$`,`$\mathbb{T}$`,`
    `legal_replies)`: The MA informs all the participants that an agent enacting the role `role` just entered in the deliberation and has introduced the information `basic_info`. The MA also informs, of the PAs already in the deliberation `pas`, as well as of the updated values of $\mathbb{C}_{F\wedge A}$ and $\mathbb{T}$. Within this message the MA attaches the legal replies (`legal_replies`) to the arguments in $\mathbb{T}$. This set of legal replies (argument schemes and critical questions) will guide the PA on which argument moves they can submit as a reply to those in $\mathbb{T}$. (This is further discussed in §6.2).

Thus, for example, if an agent with id `ag_id` enacting the role `role_id` wishes to enter the deliberation over the proposal `proposal` she will send a request locution:

`request(ag_id,ma,conv_id,0,-1,enter_dialogue(proposal,role_id,`
`basic_info))`

If the MA accepts the request it will broadcast to all but `ag_id` that an agent playing the role `role_id` has entered the dialogue and reply to agent `ag_id` that her request was accepted:

`inform(ma, all-{ag_id},conv_id,1,-1, enter_dialogue(proposal,`
`role_id, basic_info, pas, `$\mathbb{C}_{F\wedge A}$`, `$\mathbb{T}$`, legal_replies))`

`inform(ma, ag_id, conv_id,1,0,entered_dialogue(proposal, role_id,`
`basic_info, pas, `$\mathbb{C}_{F\wedge A}$`, `$\mathbb{T}$`, legal_replies))`

If the request is rejected the MA informs the PA with id `ag_id` why her request was rejected.

`inform(ma, ag_id, conv_id, 1, 0, rejected(reason)).`

Once a PA enters the dialogue it moves into the deliberation stage ant it can interact in its three layers:

4.2.2 Context Layer:

Once an agent enters the dialogue it can inform of facts it deems potentially relevant for the decision making, as well as propose complementary courses of actions that may prevent undesirable side effects that may be caused by the main action.

assert(fact): a PA asserts that the fact `fact` is the case. If accepted, `fact` is added to $\mathbb{C}_F$.

propose(action): a PA proposes to perform the action `action`. If accepted, `action` is added to $\mathbb{C}_A$.

retract(fact): a PA retracts an assertion that a fact `fact` is the case. If accepted, `fact` is removed from $\mathbb{C}_F$.

retract(action): a PA retracts the proposal to perform the action `action`. If accepted, `action` is removed from $\mathbb{C}_A$.

Each of the above messages, when sent by a PA, is wrapped in a request locution. If they are accepted, they will be broadcast to all participants by the MA.

Participants may assert and retract facts as well as propose and retract actions, at any time, as long as the deliberation is open. The only restriction is that facts and actions asserted or proposed cannot be inconsistent[8]. Hence, given a consequence relation $\vdash$ and a background theory Γ, then $\mathbb{C}_F$ and $\mathbb{C}_A$ must be such that $\mathbb{C}_F \nvdash_\Gamma \bot$ and $\mathbb{C}_A \nvdash_\Gamma \bot$. For instance, $\mathbb{C}_F$ cannot contain both: *a)* the donor does not have cancer and *b)* the donor has a malignant tumour.[9]. In other words, the state of affairs defined in $\mathbb{C}_{F \wedge A}$, though may be uncertain and may evolve throughout the deliberation, cannot be inconsistent.

At the current state of development, *ProCLAIM* does not support a conflict resolution among PAs that disagree over the described contexts of facts. From our explored scenarios (transplant and environmental) we have learned to be odd for one PA to dispute another PA's state of affairs description. This is because, each PA provides information on that she has a privileged access to. Hence, it is odd for a DA to dispute the information about a potential recipient given by a RA; similarly for an agent representing an industry to dispute information regarding the status of the wastewater treatment plant. For this reason, and in order to keep the deliberation focused, conflicts regarding whether or not a fact x is the case is either resolved outside *ProCLAIM* (*e.g.* by facilitating a persuasion dialogue or via a phone call) or should take x as uncertain. Nonetheless, as we will see in §5.2.3, PAs can still challenge an argument requesting evidence in support of some fact and may highlight the weakness of that evidence, which may motivate the retraction of the disagreed upon fact (*e.g.* the retraction of

---

[8]Where by inconsistent actions we mean actions that cannot be performed simultaneously (*e.g.* heat and cool, stay and go, etc...).

[9]Note however that $\mathbb{C}_F$ may contain *a)* clinical records indicate the donor does not have cancer and *b)* the donor has cancer

x, which leaves room to the submission of $\neg x$). In future work we intend to further develop this intuition, which may lead extending *ProCLAIM* to support such conflict resolution.

Note that the facts and actions introduced at this layer of the deliberation (*i.e.* $\mathbb{C}_{F \wedge A}$) do not themselves indicate whether or not the main proposed action is safe. $\mathbb{C}_{F \wedge A}$ is the context in which the main action is intended to be performed. Participants should thus decide whether the proposed action is safe given this context, where $\mathbb{C}_{F \wedge A}$ may change during the course of the deliberation. This decision making occurs at the *Argumentation Layer*. Although clearly, if the main proposed action or the preconditions for such an action are retracted, the deliberation concludes.

4.2.3 Argumentation Layer:

At the argumentation layer there are only two locutions: `argue` and `challenge`. A PA uses these locutions to *request* submitting an argument or a challenge. A challenge made on an argument questions the validity of the argument. From the perspective of an argumentation framework challenges can be represented as regular arguments that attack the argument under challenge. If the PA's request for submitting an argument or a challenge is accepted, the MA broadcasts this move to all participants using its version of the `argue` and `challenge` locutions. When this request is rejected, the MA's reply occurs at the proxy layer. Let us mark the locutions made by PAs with an **R** for request, and the MA's broadcasting message with an **I** for inform:

R: `argue(argument, target):` an argument `argument` is submitted by a PA in reply (as an attack) to the *argument* or *challenge* in $\mathbb{T}$, whose id is `target`. If the argument is accepted it will be broadcasted to all participants.

I: `argue(id, argument, target, legal_replies):` an argument `argument` submitted in reply (as an attack) to the *argument* or *challenge* whose id is `target`, has been accepted by the MA who broadcasts it to all participants, indicating that the argument's id is `id`. Within the same message, the MA attaches the legal replies (`legal_replies`) to `argument`. This set of legal replies (argument schemes and critical questions) will guide the PA on which argument moves they can submit at each point of the deliberation (this is further discussed in §6.2). `argument` is also added to $\mathbb{T}$, attacking the argument or challenge with id `target`.

R: `challenge(challenge, target):` a challenge `challenge` is made by a PA on an argument in $\mathbb{T}$ with id `target`. In reply to a challenge participants can submit an argument that meets the challenge (see §5.2.3).

I: `challenge(id,challenge, target, legal_replies):` a challenge `challenge` made on an argument with id `target` has been accepted by the MA who

broadcasts it to all participants, indicating that the challenge's id is `id`. Within the same message, the MA attaches the legal replies (`legal_replies`) to `challenge`. The challenge is added to $\mathbb{T}$ as an argument attacking the argument with id `target`.

All participants, including the MA, can submit arguments and challenges at any time as long as the deliberation is open and the target argument or challenge is in $\mathbb{T}$. However, the MA can reject a submitted argument or challenge because it is not a relevant move. That is, the MA's validation task introduced in §2 is performed at the proxy layer.

The fact that a participant submits an argument does not imply she endorses it. A participant may attack her own submitted arguments with other moves. This is because it is a collaborative setting, as opposed to a competitive one. Participants introduce the knowledge they have of the problem in the form of arguments. Thus, for example, the same agent can highlight a contraindication for performing the main action (attacking the initial argument) but then propose a complementary action that will mitigate its undesirable side effects and thus reinstate the main action proposal. In the same spirit, once a challenge or argument is added to $\mathbb{T}$ participants cannot retract it, *i.e.* delete it from $\mathbb{T}$. As discussed in §2, if an argument is added to $\mathbb{T}$ it is because the MA deemed the argument to be relevant for the deliberation. An argument may of course be defeated, but it should remain in the tree of arguments.

4.2.4   Endorsement Layer:

As arguments are added to the tree of arguments, participants can decide which arguments they endorse. This endorsement will affect MA's argument evaluation. For example, arguments endorsed by participants with a good reputation will be deemed stronger. Nonetheless, this argument may still be weak because, for instance, there is strong empirical evidence against it. The locutions at the *Endorsement Layer* are:

`endorse(pa_id,arg_id)`: The participant `pa_id` endorses argument or challenge with id `arg_id`.

`retract_endorsement(pa_id,arg_id)`: The participant `pa_id` retracts her endorsement of argument or challenge with id `arg_id`.

These moves can be submitted at any time while the dialogue is open and on any argument or challenge on $\mathbb{T}$. If an agent endorses two conflicting arguments, the later endorsement prevails and the earlier is automatically retracted.

When an endorsement (resp. its retraction) of an argument or challenge in $\mathbb{T}$ is made by a PA (via a request locution), the MA adds (resp. subtracts) this endorsement (represented as the predicate `endorse(pa_id,arg_id)`) from the **endorsement set**, which we denote as $\mathbb{E}$.

4.2.5 Resolution Stage:

Once participants have constructed the context of facts and actions $\mathbb{C}_{F \wedge A}$, the tree of arguments $\mathbb{T}$, and have informed of their endorsements, the MA proceeds to evaluate $\mathbb{T}$. The deliberation moves into the *Resolution Stage* either because all the participants have informed that they have no further moves to submit that may change either $\mathbb{C}_{F \wedge A}$, $\mathbb{T}$, or $\mathbb{E}$; or because a *timeout* was triggered. In either case, the MA proposes a solution for the deliberation, based on the evaluation of $\mathbb{T}$. If a *timeout* has been triggered, PAs will not have the chance to revise the proposed solution. In §7 we briefly discuss the nature of the argument evaluation and how a recommended solution is not merely a *yes/no* answer.

`no_more_moves():` The participant informs that she has no further moves to submit (moves that may change either $\mathbb{C}_{F \wedge A}$, $\mathbb{T}$, or $\mathbb{E}$), for consistency she does so via a request move. Once all participants submitted this move, the MA proceeds to evaluate $\mathbb{T}$. This move, however, does not prevent participants from submitting further moves, overriding her own move of `no_more_moves`. This is important to allow because new relevant information may be available at any time and should be included in the deliberation. If the move `no_more_moves` is overridden, the deliberation moves again the deliberation stage.

`leave_dialogue(reason):` The participant request that to leave the deliberation and may provide a reason `reason` for that. If this move is accepted by the MA all PAs will be informed that the participant has left the deliberation. Of course, if all participants leave the deliberation the deliberation concludes and the MA will propose a solution (via the `close_deliberation` locution) on the basis of the available knowledge $\mathbb{C}_{F \wedge A}$, $\mathbb{T}$, and $\mathbb{E}$.

`time_out(reason):` The MA informs that a timeout has been triggered. In general terms this means that too much time has been spent in the deliberation and so a new resolution policy should be applied. For instance, picking-up the telephone. How to proceed with a timeout is application dependent. Provisionally we formalise it as a trigger for the MA to evaluate $\mathbb{T}$ with the available knowledge ($\mathbb{C}_{F \wedge A}$, $\mathbb{T}$, and $\mathbb{E}$) and propose a solution while disabling any further moves from the participants. The MA may provide a reason `reason` for the timeout.

`solution(solution,sol_id):` Once all participants have submitted the `no_more_moves` (and did not override it with any other move) the MA proposes a solution `solution` whose id is `sol_id`. The proposed solution may motivate participants to submit further moves or to simply accept the solution. If a participant submits a move of the *Deliberation Stage*, she should again submit the `no_more_moves` locution for the MA to propose the new solution. However, if the timeout is triggered, the deliberation will

conclude with the given solution providing no chance for the participants to submit further moves[10].

accept(sol_id): Once a solution with id `sol_id` is given, if all agents accept it, the deliberation concludes.

close_deliberation(solution,sol_id): The deliberation is closed with the proposed solution `solution`. This locution is submitted either after all participants have submitted the `accept(sol_id)` move or the timeout has been triggered and the MA has proposed a solution.

We are working under the assumption that the CBRc (case based reasoning component) is *time consuming* and requires the full $\mathbb{T}$ for argument evaluation. However, if we manage to develop a CBRc whose performance can be adjusted to real-time deliberation, a proposal for resolution of the $\mathbb{T}$ will always be visible for the participants and the cycle `solution(solution,id_sol)`, `accept(id_sol)` will not be necessary. It would be enough to submit the `no_more_moves` locution.

4.2.6 Inform Layer:

Throughout the deliberation dialogue, participants can request from the MA an update of the argument tree, in which facts have been introduced, or request for the legal replies to a given argument or challenge in $\mathbb{T}$. Thus, if for whatever reason a participant misses a piece of information she can recover it upon request.

R: `get_arg_tree()`: A PA requests the MA for the updated $\mathbb{T}$.

I: `arg_tree(`$\mathbb{T}$`)`: The MA informs a PA of the updated $\mathbb{T}$.

R: `get_context()`: A PA requests the MA for the updated $\mathbb{C}_{F \wedge A}$.

I: `context(`$\mathbb{C}_F$`, `$\mathbb{C}_A$`)`: The MA informs a PA of the updated $\mathbb{C}_{F \wedge A}$.

R: `get_endorsement()`: A PA requests the MA for the updated $\mathbb{E}$.

I: `endorsement(`$\mathbb{E}$`)`: The MA informs a PA of the updated $\mathbb{E}$.

R: `get_legal_replies(arg_id)`: A PA requests the MA for the legal replies to an argument or challenge in $\mathbb{T}$ with id `arg_id`.

I: `legal_replies(arg_id, legal_replies)`: The MA informs a PA of the legal replies to an argument or challenge in $\mathbb{T}$ with id `arg_id`.

The above described dialogue game is rather liberal. PAs can submit almost any locutions at any time during the deliberation. There are of curse a few restrictions such as at the proxy level the MA has the obligation to reply to the

---

[10]In that case, the decision making process may indeed continue, but following a different policy.

PA's requests. Also, the deliberation dialogue can only be opened once, PAs can only request to enter the dialogue if they are not already in it and they cannot participate once the deliberation is either closed or they have left it (via the `leave_dialogue` locutions). As long as the timeout has not been triggered, PAs can submit any move of the *Deliberation Stage* with no turn restrictions. That is PAs can submit any fact (resp. complementary action) at any time of the deliberation, as long as this fact (resp. action) is not already asserted (resp. proposed) or it is inconsistent with $\mathbb{C}_F$ (resp. $\mathbb{C}_A$). Similarly, PAs can retract any facts and actions in $\mathbb{C}_{F \wedge A}$. PAs can submit the `argue` or `challenge` locution at any time of the deliberation, as long as it is open and the timeout has not been triggered. The target of their argument or challenge must be an element of $\mathbb{T}$ and, in particular, they can attack their own arguments and they do not have any obligation to defend their arguments from other arguments or challenges. What is at stake is not who is right or wrong, but whether or not the main action can safely be performed.

It is at the *Argumentation Layer* that the deliberation is kept highly focused on the subject matter, through definition of the arguments and challenges the PAs can submit throughout the deliberation. That is, the set of legal replies (argument schemes and CQs) made available to the participants. In the following section we describe in detail the *Argumentation Layer*.

5 *ProCLAIM* Argumentation Layer

One of the pillars of *ProCLAIM* is the definition of the deliberation dialogue's *Argumentation Layer*, namely, what types of arguments participants can exchange and following what rules. As a way to keep deliberations focused as well as reducing the participants' overhead in terms of argument construction, *ProCLAIM* is quite specific in what can be argued about and how. To this end, the model defines a *protocol-based exchange of arguments* that can be regarded as an argumentative process for eliciting knowledge from the participants, as opposed to defining a strategic dialogue in which a better choice of arguments may better serve the agents' individual goals. This argumentation-protocol is defined in terms of a structured set (a *circuit*) of schemes and their associated CQs (to a scheme are associated a set of CQs which are themselves defined in terms of schemes that have associated CQs, and so on...). *ProCLAIM* defines an application-independent protocol-based exchange of arguments specialised for arguing over safety critical actions. Then, for each target application (*e.g.* transplant or environmental scenario) this application-independent protocol has to be further specialised in order to construct the scenario-specific ASR[11]. This is discussed in §6.1.

In this section we present the application-independent circuit of AS and CQs. We start by introducing in the following subsection the internal structure of *ProCLAIM*'s arguments and in §5.2, we present the protocol-based exchange of arguments.

---

[11] Argument Scheme Repository

5.1 The Structure of an Argument

Action proposals are typically motivated by the goals agents wish to realise. Many formal accounts ([48, 19, 60, 59]) of action proposal assume, though sometimes implicitly, the following three dimensions:

R: Domain of facts in circumstances where the action is proposed.

A: Domain of actions.

G: Domain of goals.

Based on these domains the following argument can be constructed '*an action A is proposed in circumstances R (a set of facts) because it is expected to realise a desirable goal G*'. The problem with such an argument structure is that the notion of a goal is ambiguous, potentially referring indifferently to any direct result of the action, the consequence of those results and the reasons why those consequences are desired [4]. To account for these distinctions, Atkinson *et al.* considered two additional domains:

S: Domain of facts arrived after performing the action.

V: Domain of values where the values represent the social interests promoted through achieving the goal.

in order to propose the following argument scheme for action proposals:

> **AtkSch:**
> In the circumstances R
> we should perform action A
> to achieve new circumstances S
> which will realise some goal G[12]
> which will promote some value V

This argument scheme is presented along with sixteen associated CQs which can be classified into three categories: *What is true (* e.g. *–Questioning the description of the current circumstances–), what is best (*e.g. *–Questioning whether the consequences can be realised by some alternative action–)* and *representational inconsistencies (*e.g. *–Questioning whether the desired features can be realised–*). In [4] *AtkSch* along with its sixteen CQs are used to define a persuasion dialogue game for reasoning about action proposal.

Atkinson's persuasion dialogue is primarily addressed at resolving a choice amongst competing action proposals, choosing which action is the best, *i.e.* which action will bring about the best situation, where 'best' is relative to an agent and consideration is given to subjective value-based judgements, as well as more objective ones. In arguing about action proposals, participants

---

[12]Where a goal is some particular subset of S that the action is intended to realised in order to promote the desired value.

may undermine an action proposal by questioning whether the action will bring about any undesirable effects[13]. This is just one possibility in the persuasion dialogue; one can also argue as to which goals are desirable or not. In short, participants can argue about whatever is reasonable when deciding *what to do* in general terms. This generality is indeed a desirable feature of Atkinson's persuasion dialogue and for that reason this work is taken as a starting point for the definition of *ProCLAIM*'s Argumentation Layer. However, precisely because of this openness, it is inoperable for our intended applications.

In *ProCLAIM*, the desirable and undesirable goals are assumed to be shared by all participants. Furthermore, the main proposed action itself (*e.g.* transplant an organ or spill the industrial wastewater) is, in default circumstances, taken to be the right thing to do, requiring no further motivation in its proposal. Moreover, decisions in *ProCLAIM* are taken with respect to a single social value *safety* (or patient's quality of life, in the transplant scenario). Therefore, the value dimension can be ignored[14]. A particular consequence of this defined context is that PA's individual goals and values, while may affect which arguments they submit and endorse, in themselves do not constitute a reason for or against a proposed action. What becomes a matter of debate then, is whether the current circumstances are such that the proposed action can safely be performed. Namely, whether or not the context of facts $\mathbb{C}_F$, constructed at the *Context Layer*, is such that the main action will bring about severe undesirable side effects. The deliberation can thus be regarded as an argumentative process for eliciting from the participants (experts) what are the *relevant* facts ($\mathtt{f}_0,..,\mathtt{f}_n \in \mathbb{C}_F$) for assessing the action's safety, accounting for the complementary courses of actions (those actions added to $\mathbb{C}_A$). A formal definition of the relevance of a set of facts is given later in this section (Definition 5.1).

To illustrate the relevance of facts in the medical scenario, let us suppose a donor of a lung is infected with the Hepatitis C virus ($\mathtt{hcv}$). Now, it can be argued that the transplant is unsafe (argument $A2$ in fig. 5) because the recipient of the transplanted lung will result in having $\mathtt{hcv}$, which is a severe infection. Thus, the donor being infected with $\mathtt{hcv}$ is a relevant fact, given that, because of this fact the transplant will cause an undesirable side effect. Suppose now that the potential recipient also has $\mathtt{hcv}$. And so, it cannot be claimed that, for this recipient, having $\mathtt{hcv}$ is an undesirable side effect of the lung transplant (argument $A3$ in fig. 5). Therefore, the potential recipient's $\mathtt{hcv}$ is a relevant fact. It is because that fact holds that the action does not cause an undesirable side effect. Note however, that if the donor would not have had $\mathtt{hcv}$, whether the recipient has $\mathtt{hcv}$ or not, is irrelevant. That is, relevance is *context dependent*. An attack on argument $A3$ will assume a *context* in which the donor and recipient both have $\mathtt{hcv}$. Let us suppose that there are other

---

[13]We cannot assume that because the effect is undesirable it must be a *side effect* of the action. It may actually be a state of affairs that, from the perspective of one participant, is a desirable outcome of the action, but not for all participants.

[14]It may be interesting to bring into the deliberation the *cost* value. Some proposed actions although deemed safe, cannot be taken because the system cannot afford the expenses incurred by the actions. We leave such an extension for future work.

contraindications for the transplantation that, at least *a priori*, are independent of the donor and recipient's `hcv`. For example, that the available lung is too big for the recipient's thoracic cavity. Such an argument will directly attack argument $A1$, where the context, or to be more precise, the *local context*, is that an available organ is proposed for transplantation into a given patient. To capture this notion, we explicitly associate to each argument a *local context* of facts and of actions.
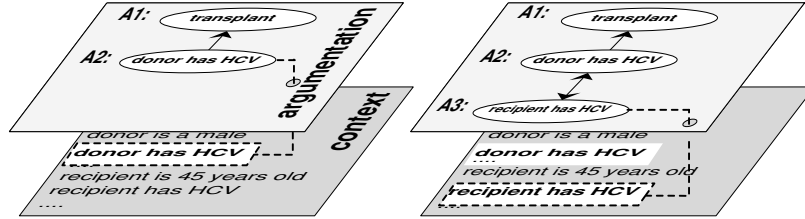

Figure 5: As arguments are submitted facts are highlighted as relevant. Note that for example, it has not been deemed relevant that the recipient is 45 years old or the donor is a male. Moreover, if the donor would not have had HCV (Hepatitis C virus), the recipient's HCV may have not been highlighted either.

We denote the **local context of actions** of an argument as $\mathcal{A}$, and the **local context of facts** as $\mathcal{C}$. Upon submission of the first argument, $\mathcal{A}$ and $\mathcal{C}$ are empty. These are updated so that for any subsequent submitted argument $\mathcal{A}$ contains the proposed action itself (*e.g.* the transplant proposal) and $\mathcal{C}$ the minimum set of facts where the proposed action can be performed (*e.g.* an available organ and a potential recipient). In general, each submitted argument updates its $\mathcal{C}$ and $\mathcal{A}$ to account for the particularities of each case (*e.g.* the donor's and recipient's particularities). In the previous example we saw how argument $A2$ extended $\mathcal{C}$ to include the donor's `hcv`. Argument $A3$ then extended $\mathcal{C}$ by adding the recipient's `hcv`. Note that while these facts were already in the (global) context $\mathbb{C}_F$, it's through their use in the argumentation that they are highlighted as relevant. Thus, for a set of facts (resp. actions) to be added to $\mathcal{C}$ (resp. to $\mathcal{A}$) it must be *relevant*. Meaning that, within their *local* context these facts or complementary actions make the main action safe or unsafe.

To continue with the identification of the elements and relations of *Pro-CLAIM*'s arguments, let us recall that a *ProCLAIM* argument expresses a relation among the four domains: current state ($\mathbf{R}$), actions ($\mathbf{A}$), arrived states ($\mathbf{S}$) and goals ($\mathbf{G}$). We can further constrain $\mathbf{S}$ and $\mathbf{G}$ so that $\mathbf{S}$ contains only *side effects* of the actions in $\mathbf{A}$, and $\mathbf{G}$ contains only *undesirable goals* which such side effects may realise.

Let us also recall, that the intended effects of the main proposed action are assume to be desirable and so, they are beyond question.

Let us formalise these domains in terms of finite sets of grounded predicates which will be written in `teletype`, *e.g.* `av_org(d,o)` $\in \mathbf{R}$ meaning that an organ `o` of a donor `d` is available.

ProCLAIM arguments express relations between elements of the above domains. Specifically, the following elements:

$\mathcal{C}$: The local context of facts assumed to be the case, where $\mathcal{C} \subseteq \mathbf{R}$.

$\mathcal{A}$: The local contexts of proposed actions, where $\mathcal{A} \subseteq \mathbf{A}$.

R: A set of facts, where $R \subseteq \mathbf{R}$. For more than one set of facts we write $R1$, $R2$,... We denote by R_p the set of facts introduced as preconditions for the performance of a proposed action.

A: A set of actions, where $A \subseteq \mathbf{A}$. We write A_m to denote the *main set of actions* and A_c the complementary courses of actions argued to prevent the achievement of an undesirable side effect. For more than one set of complementary actions we write A_c1, A_c2,...

S: The set of side effects caused by the proposed action, where $S \subseteq \mathbf{S}$. For more than one set of side effects we write $S1$, $S2$,...

g: The undesirable goal realised by S, where $g \in \mathbf{G}$. For more than one goal we write $g1$, $g2$,...

Different argument schemes defined by *ProCLAIM* correspond to different relations amongst these elements, where these relations are expressed in terms of special predicates, and a defeasible consequence relation $\vdash\!\sim$ from which conclusions follow defeasibly or non-monotonically from the set of premises. We thus assume:

- A defeasible consequence relation $\vdash\!\sim$;

- A background theory Γ;

- The special predicate `side_effect` on subsets of $\mathbf{S}$ where `side_effect`(S), with $S \subseteq \mathbf{S}$, denotes that S are side effects given a background contexts of facts $\mathcal{C}$ and actions $\mathcal{A}$.

- The special predicate `intended`[15] on subsets of $\mathbf{A}$ where `intended`(A_c), with $A_c \subseteq \mathbf{A}$, denotes that the set of actions A_c is intended.

Given a set of facts or actions, we assume its conjunction to be the case, respectively proposed. And, if A and B are two sets of either facts or actions, to say that all the elements in A and of B hold, are respectively intended, we write $A \wedge B$.

---

[15]In the deliberation presented in this work we do not distinguish between *intending* and only *proposing* to perform an action. This is discussed in §5.2.5.

Thus, for example, we can write: $R \wedge \mathcal{C} \wedge \mathtt{intended}(\mathcal{A}) \wedge \Gamma \mathrel{|\!\sim} \mathtt{side\_effect}(S)$. Meaning that if R and $\mathcal{C}$ are the case, the proposed actions $\mathcal{A}$ will result in the set of side effects S. The rationale as to why $\mathcal{A}$ will cause S is in the background theory Γ. Each agent and knowledge resource defines its own version of Γ, which may contain different rules and reasoning techniques. For example, a basic artificial agent may contain a fixed table with precodified 4-tuples relating the four dimensions $\mathbf{R} \times \mathbf{A} \times \mathbf{S} \times \mathbf{G}$. A slightly more sophisticated artificial agent will define an internal structure to each of the four dimensions with a number of transition rules. A human agent, on the other hand, will use her own reasoning (her own version of Γ and $\mathrel{|\!\sim}$;) to reason about the exchanged arguments. However, all these *heterogeneous* agents will have to agree on the syntax and semantics of the exchanged *ProCLAIM* arguments.

Typically the background theory is written as a subscript on the consequence relations: $\mathrel{|\!\sim}_{\Gamma}$. To emphasise that $\mathcal{C}$ and $\mathcal{A}$ are assumed to be the case, *i.e.* that they are contextual information, they are also written as subscripts on the consequence relations: $\mathrel{|\!\sim}_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma}$. This notation has no effect on the consequence relation, but it allow us to single out particular elements in the relation: *e.g.* to highlight the set of facts R in the consequence relation $R \mathrel{|\!\sim}_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma} \mathtt{side\_effect}(S)$. We also take the liberty of omitting the $\mathtt{intended}$ predicate wrapping the actions.

With these elements and relations the relevance of a set of facts and actions (w.r.t. realising an undesirable goal) can be defined as follows:

Definition 5.1 *Within the context of facts $\mathcal{C}$ and of proposed actions $\mathcal{A}$ a set of facts $R \subseteq \mathbf{R}$ is said to be **relevant** if one of the following two situations holds:*

- *In circumstances $\mathcal{C}$, if R holds the actions $\mathcal{A}$ will cause an undesirable side effect. Otherwise, if R does not hold, the undesirable side effect is no longer expected to be caused by $\mathcal{A}$ (in circumstances $\mathcal{C}$):*

  - $R \mathrel{|\!\sim}_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma} \textit{side\_effect}(S)$;
  - $\textit{side\_effect}(S) \mathrel{|\!\sim}_{R \wedge \mathcal{C} \wedge \Gamma} g$ *and*
  - $\mathrel{|\!\not\sim}_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma} \textit{side\_effect}(S)$.

  Or

- *In circumstances $\mathcal{C}$ actions $\mathcal{A}$ will cause an undesirable side effect. But if R holds, then either the side effect is not expected to be caused by $\mathcal{A}$, or the side effect cannot be deemed as undesirable (i.e. the degree to which the side effect realises the undesirable goal is too weak):*

  - $\mathrel{|\!\sim}_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma} \textit{side\_effect}(S)$ *and*
  - $\textit{side\_effect}(S) \mathrel{|\!\sim}_{\mathcal{C} \wedge \Gamma} g$

  but either:

29

$-\ R |\!\!\!\!\sim_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma}$ `side_effect`(S) *or;*

$-\ R \wedge$ `side_effect`$(S) |\!\!\!\!\sim_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma} g$

Note that when it is said that an undesirable side effect is not expected it is strictly in the local context defined by $\mathcal{C}$ and $\mathcal{A}$. This undesirable side effect may well occur for other reasons, *e.g.*, due to other facts not in $\mathcal{C}$ but in $\mathbb{C}_F$.

The definition of a relevant complementary course of actions is as follows:

Definition 5.2 *Within the context of facts $\mathcal{C}$ and of proposed actions $\mathcal{A}$ a set of actions $A_c \subseteq \mathbf{A}$ is said to be **relevant** if the preconditions R_p for its performance hold ($R_p \subseteq \mathbb{C}_F$) and A_c either prevents an undesirable side effect or it causes one.*
That is:

- $|\!\!\sim_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma}$ `side_effect`(S) *and*

- `side_effect`$(S) |\!\!\sim_{\mathcal{C} \wedge \Gamma} g$ *and*

- $R_p \wedge$ `intended`$(A_c) |\!\!\!\!\sim_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma}$ `side_effect`(S)

Or;

- $R_p \wedge$ `intended`$(A_c) |\!\!\sim_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma}$ `side_effect`(S) *and*

- `side_effect`$(S) |\!\!\sim_{\mathcal{C} \wedge \Gamma} g$ *and*

- $|\!\!\!\!\sim_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma}$ `side_effect`(S).

In what follows we will use the above introduced concepts to define the arguments schemes and their associated critical questions to be used in the *Argumentation Layer*. Using these schemes and critical questions participants will submit their arguments, highlighting with each argument the relevant facts and complementary courses of actions. These relevant factors (facts or actions) are the ones that can be added to the arguments' local contexts. Once PAs have submitted all their arguments, and so all the relevant facts and actions have been introduced, the tree of arguments $\mathbb{T}$ is evaluated to resolve whether the main action can safely be performed.

5.2   Protocol-Based Exchange of Arguments

In this section we introduce the argument schemes and their associated critical questions tailored for deliberating over safety critical actions. Each of these argument schemes encodes a particular relation among elements in $\mathbf{R}$, $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{G}$. Arguments instantiating these schemes represent instances of these relations, while their associated CQs question them. Thus, with the exchange of arguments, participants build a subspace of $\mathbf{R} \times \mathbf{A} \times \mathbf{S} \times \mathbf{G}$ tailored to the particular problem at hand. Hence, the deliberation process can be regarded as a mechanism for exploring the relevant facts in $\mathbf{R}$, accounting for the complementary courses of actions in $\mathbf{A}$, guided by the (undesirable) side effects which

are highlighted in **S** and **G** . The relevant elements in **R** and **A** are those that have an impact in **S** and **G** .

The schemes and critical questions will be introduced in a **modular** fashion. We start by introducing a set of assumptions that will help in constructing a basic circuit of six schemes and their associated CQs:

- **Assum_1: R**, **A**, **S** and **G** have no internal structure (*e.g.* no taxonomy). These are Assum_1a, Assum_1b, Assum_1c and Assum_1d respectively.

- **Assum_2:** All introduced facts R are in $\mathbb{C}_F$. Arguments must use facts that are in the context of facts $\mathbb{C}_F$.

- **Assum_3:** *a)* All proposed actions A are in $\mathbb{C}_A$, *b)* they can be performed ($R_p \subseteq \mathbb{C}_F$), *c)* and they do not conflict with other proposed actions (*i.e.* no two or more action are such that if jointly performed they cause an undesirable side effect).

- **Assum_4:** Each $g \in \mathbf{G}$ is such that if the main action will realise g the action is deemed unsafe.

As we relax some of these assumptions we will be extending this circuit of AS and CQs. In §5.2.2 we enrich **R** with a taxonomy by introducing a specificity relation. In §5.2.3 we add a defeasible entailment to **R** to allow for uncertainty. In §5.2.4 we permit the use of facts not in $\mathbb{C}_F$, in order to account for incomplete information. Finally, in §5.2.5 we discuss other extensions that we are formalising.

5.2.1 Basic Argument Schemes

In this subsection we present the basic protocol-based exchange of arguments consisting of six argument schemes and their associated critical questions by which players participate in the deliberation, introducing new relevant facts and complementary courses of actions.

Each scheme is presented as a four part composite: A set of *preconditions*, the scheme's *body*, its associated *critical questions* and the scheme's *context updating rule* by which the arguments' local contexts ($\mathcal{C}$ and $\mathcal{A}$) are updated. The body of the scheme is itself presented in three complementary representations: a *narrative* version written in natural language; a *formal* version; and the deliberation's dialogue locutions, *i.e.* the content of the `argue` and `challenge` locutions introduced in §4.2. Let us start by introducing the first argument scheme, AS1, that sets the deliberation's topic. In fact, this scheme is instantiated at the deliberation's *Open Stage* (§4.2) as the *proposal*. This first argument is the root of $\mathbb{T}$.

Let us just introduce some notation, scheme $AS1$ proposes the main action under the assumption that Am will cause no undesirable side effect: $\sim$ `undSideEffect`(A_m), where $\sim$ denotes the weak negation. Subsequent arguments will attack this assumption by highlighting an undesirable goal or defend this assumption arguing against the realisation of the highlighted an undesirable

goal.

AS1	
Preconditions: $R_p \subseteq \mathbb{C}_F$, $A_m \subseteq \mathbb{C}_A$, $\mathcal{C} = \{\}$ and $\mathcal{A} = \{\}$	
Body:	In circumstances R_p The proposed course of action A_m can safely be performed.
	$R_p \wedge \sim \texttt{undSideEffect}(A_m) \mid\!\!\sim_\Gamma \texttt{propose}(\text{Am})$
	$\texttt{argue}(\mathcal{C},\mathcal{A},\texttt{propose}(R_p,A_m));$
Critical Questions:	
CQ1: Are circumstances such that an undesirable side effect will be achieved?	
Context Updating Rule: $\mathcal{C} := R_p$; $\mathcal{A} := A_m$.	

To illustrate the use of this scheme, let us introduce an example from the transplant scenario, which we will develop throughout this paper. Let us suppose a `lung` of a donor `d` is available (`av_org(d,lung)`) for a potential recipient `r` (`p_recip(r,lung)`). And so the intention is to transplant the lung to this recipient (`transp(r,lung)`). Hence, `av_org(d,lung)`,`p_recip(r,lung)` $\in \mathbb{C}_F$ and `transp(r,lung)` $\in \mathbb{C}_A$. Therefore the initial argument, say A, can be submitted instantiating $AS1$ as follows:

A: `argue({},{},propose({av_org(d,lung),p_recip(r,lung)},`
         `{transp(r,lung)}))`[16]

An argument instantiating $AS1$ proposes the main action making the assumption that no undesirable goal will be realised. Any attack to such argument involves arguing that this assumption is false, that an undesirable goal will be realised. Typically, critical questions associated with a scheme enable agents to attack the validity of the various elements of the scheme and the connections between them. Also, there may be alternative possible actions and side effects of the proposed action [4]. In the particular case of arguments instantiating $AS1$ what can be questioned is whether there is a fact, or set of facts R, in the current circumstances ($R \subseteq \mathbb{C}_F$) that makes the proposed action unsafe. Hence, what can be questioned is the assumption that there are no contraindications for performing the proposed action. That is, critical question $CQ1$, which we denote as $AS1\_CQ1$, can be used.

---

[16]It is worth noting that an artificial agent may represent internally this argument in many forms, for instance in a more standard support-claim argument structure like $<$ {`av_org(d,lung)` $\wedge$ `p_recip(r,lung)` $\wedge \sim$ `undSideEffect(transp(r,lung))` $\Rightarrow$ `propose(transp(r,lung))`,`av_org(d,lung)`,`p_recip(r,lung)`},`propose(transp(r,lung))` $>$

An answer 'no' to this question, implicitly encoded in the assumption of the initial argument, would imply little progress in the deliberation. An answer 'yes' to this question constitutes an attack on the argument. Thus, for the deliberation to effectively progress, $AS1\_CQ1$ can only be addressed by introducing a contraindication, *i.e.* a set of facts R that will result in the action causing an undesirable side effect. This use of $AS1\_CQ1$ is effected by an argument instantiating the scheme $AS2$, and that attacks the argument instantiating $AS1$.

Finally, to illustrate the scheme's context updating rule, note that any reply to argument A will assume as contextual information that there is an available `lung` of a donor `d`, a potential recipient `r` for that organ and that the transplant is intended. That is, $\mathcal{C} = \{\texttt{av\_org(d,lung)},\texttt{p\_recip(r,lung)}\}$ and $\mathcal{A} = \{\texttt{transp(r,lung)}\}$. Needless to say, if the assertion of any of these facts or actions is retracted at the *Context Layer*, the deliberation concludes.

AS2
Preconditions: $R \subseteq \mathbb{C}_F$, $S \subseteq \mathbf{S}$, $S \neq \emptyset$, $g \in \mathbf{G}$, and $\mathcal{C}$ and $\mathcal{A}$ the context of facts and actions of the target argument.

Body:	In circumstances $\mathcal{C}$ Because R holds, actions $\mathcal{A}$ will cause a side effect S which will realise some undesirable goal g.
	○ $R \mathrel{\mid\!\sim}_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma} \texttt{side\_effect}(S)$; and ○ $\texttt{side\_effect}(S) \mathrel{\mid\!\sim}_{\mathcal{C} \wedge R \wedge \Gamma} g$;
	$\texttt{argue}(\mathcal{C},\mathcal{A},\texttt{contra(R,S,g)})$;

Critical Questions:
CQ1: Are circumstances such that the stated side effect will not occur?
CQ2: Are circumstances such that the side effect will not realise the stated goal?
CQ3: Is there a complementary course of action that prevents the achievement of the stated effect?

Context Updating Rule: $\mathcal{C} := \mathcal{C} \cup R$; $\mathcal{A} := \mathcal{A}$.

An argument instantiating **AS2** identifies contraindications R for performing the proposed actions $\mathcal{A}$, in circumstances $\mathcal{C}$.

Continuing with the above example, let us suppose that the donor of the offered lung has *smoking history* (`d_p(d,s_h)`: donor `d` has property `s_h`). Let us suppose, as well, that the donor agent, DA, that offers the lung for transplantation, believes `s_h` to be a contraindication because the lung may be rejected

by the recipient, thus realising the undesirable goal `grft_fail(r)`. Hence, DA believes $AS1\_CQ1$ to be the case, and so may want to attack argument A. This can be done by submitting an argument $B1$ (see fig. 6), that instantiates $AS2$ as follows:

$B1$: `argue(`$\mathcal{C}$`,`$\mathcal{A}$`,contra({d_p(d,s_h)},{reject(r,lung)},grft_fail(r)));`

Let us now identify $AS2$'s critical questions. That is, which lines of attack can be pursued in order to, for example, attack argument $B1$. For that purpose, let us highlight what is being asserted by an argument instantiating $AS2$, taking into account that $\mathcal{C}$ has been updated (*e.g.* in argument $B1$, $\mathcal{C}$ = `{av_org(d,lung),p_recip(r,lung),` `d_p(d,s_h)}`) while $\mathcal{A}$ remains the same and that *ProCLAIM* arguments only assert a relation among the sets $\mathbf{R}$, $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{G}$:

1. $\mathcal{C} \wedge \mathcal{A} \mathrel{\vdash\!\!\!\sim}_{\Gamma}$ `side_effect`(S); and

2. `side_effect`$(S) \wedge \mathcal{C} \mathrel{\vdash\!\!\!\sim}_{\Gamma} g$;

Firstly, whether these two relations hold is evaluated first at the *Proxy Level* (§4.1) where the MA validates the incoming arguments and latter at the *Resolution Stage* (§4.2) where a relative strength of the accepted arguments is assigned. Secondly, under the assumptions presented at the beginning of this subsection, the local contexts are such that $\mathcal{C} \subseteq \mathbb{C}_F$ and $\mathcal{A} \subseteq \mathbb{C}_A$ (Assum_2 and Assum_3a resp) and thus they are taken to be the case (*e.g.* in argument $B1$ `d_p(d,s_h)` holds). And with Assum_4 we have that if g holds as consequence of the action this should be deemed unsafe. What can be done to attack an argument instantiating scheme $AS2$ is an update to either $\mathcal{C}$ or $\mathcal{A}$ so that either of the two relations does not hold ($\mathrel{\not\vdash\!\!\!\sim}$ `side_effect`(S) or $\mathrel{\not\vdash\!\!\!\sim} g$). Since each fact in $\mathcal{C}$ and each action in $\mathcal{A}$ has to be in $\mathbb{C}_F$ and $\mathbb{C}_A$ respectively, and $\mathbb{C}_F$ and $\mathbb{C}_A$ do not allow for inconsistencies, any update on the local contexts has to be truth preserving. Retracting or negating an element of $\mathbb{C}_F$ or $\mathbb{C}_A$ is done at the *Context Layer* and the effect of such moves is discussed in §5.2.4. Since neither $\mathbf{R}$ or $\mathbf{A}$ have an internal structure (we discuss relaxation of Assim_1 in §5.2.2), truth preserving updates on $\mathcal{C}$ or $\mathcal{A}$ can only be done by adding a new set of (relevant) facts R to $\mathcal{C}$ or complementary courses of actions A_c to $\mathcal{A}$. Therefore, what can be questioned on arguments instantiating scheme $AS2$ is whether there exists a set $R \subseteq \mathbb{C}_F$ such that in the new context $\mathcal{C} \cup R$ the side effect S is no longer expected ($AS2\_CQ1$); or in which the undesirable goal g would not be realised ($AS2\_CQ2$)[17]. Note that $\mathcal{A}$ only appears in the first assertion. Thus, changes in $\mathcal{A}$ ($\mathcal{A} \cup A_c$) can only be proposed in order to argue that the complementary course of action A_c can prevent the side effect S ($AS2\_CQ3$). These three critical questions have only practical use if the appropriate relevant R or A_c are provided. The critical questions $AS2\_CQ1$, $AS2\_CQ2$ and $AS2\_CQ3$ are therefore addressed as attacking arguments respectively instan-

---

[17]That is, given the new context $\mathcal{C} \cup R$ the degree by which S realises g is too weak.

tiating schemes $AS3$, $AS4$ and $AS5$, and so introducing the relevant Rs and A_cs.

AS3	
Preconditions: $R \subseteq \mathbb{C}_F$, S the side effect of the the target argument $\mathcal{C}$ and $\mathcal{A}$ the updated context of facts and actions of the target argument.	
Body:	In circumstances $\mathcal{C}$ Because R holds, the side effect S is not expected as caused by $\mathcal{A}$. $R \hspace{0.1em}\not\hspace{-0.3em}\mid\sim_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma} \texttt{side\_effect}(S)$ $\texttt{argue}(\mathcal{C},\mathcal{A},\texttt{no\_side\_effect(R,S)});$
Critical Questions:	
CQ1: Are circumstances such that an undesirable side effect will occur?	
Context Updating Rule: $\mathcal{C} := \mathcal{C} \cup R$; $\mathcal{A} := \mathcal{A}$.	

AS4	
Preconditions: $R \subseteq \mathbb{C}_F$, S and g of the target argument replied to $\mathcal{C}$ and $\mathcal{A}$ the updated context of facts and actions of the target argument.	
Body:	In circumstances $\mathcal{C}$ And assuming $\mathcal{A}$ will be performed It is because R holds, that S does not realises $\texttt{g}$ $\texttt{side\_effect}(S) \wedge R \hspace{0.1em}\not\hspace{-0.3em}\mid\sim_{\mathcal{C} \wedge \Gamma} g$ $\texttt{argue}(\mathcal{C},\mathcal{A},\texttt{not\_realised\_goal(R,S,g)});$
Critical Questions:	
CQ1: Are circumstances such that the side effect will realise the undesirable goal?	
Context Updating Rule: $\mathcal{C} := \mathcal{C} \cup R$; $\mathcal{A} := \mathcal{A}$.	

AS5	
Preconditions: $A_c \subseteq \mathbb{C}_A$, $R_p \subseteq \mathbb{C}_F$ preconditions to perform A_c, S of the target argument; $\mathcal{C}$ and $\mathcal{A}$ the updated context of facts and actions of the replied argument.	
Body:	In circumstances $\mathcal{C} \cup R_p$ The complementary course of action A_c Prevents actions $\mathcal{A}$ from causing the side effect S.<hr>$A_c \wedge R_p \mathrel{\vert\!\!\not\sim}_{\mathcal{C}\wedge\mathcal{A}\wedge\Gamma} \texttt{side\_effect}(S)$<hr>$\texttt{argue}(\mathcal{C},\mathcal{A},\texttt{preventive\_action}(\texttt{A}_c,\texttt{R}_p,\texttt{S}));$
Critical Questions:	
CQ1: Are circumstances such that an undesirable side effect will be achieved?	
Context Updating Rule: $\mathcal{C} := \mathcal{C} \cup R_p$; $\mathcal{A} := \mathcal{A} \cup A_c$.	

Figure 6 illustrates the use of these three argument schemes. Argument $B2$, instantiating $AS3$, attacks $B1$, indicating that because the donor does not have a Chronic Obstructive pulmonary disease ($R = \{\texttt{d\_p(d,no\_copd)}\}$) the donor's smoking history is no longer a contraindication. Argument $C2$, instantiating $AS4$, attacks argument $C1$, indicating that because the potential recipient already has HIV ($\texttt{p\_r\_p(r,hiv)}$), the infection cannot be deemed as a severe infection caused by the lung transplant[18]. Finally, argument $D2$ illustrates an instantiation of scheme $AS5$ proposing to administrate *penicillin* to the recipient ($\texttt{treat(r,penicillin)}$) of a lung of a donor whose cause of death was a *streptococcus viridans endocarditis* ($\texttt{d\_p(d,sve)}$) so as to prevent an infection of that same bacteria ($\texttt{r\_p(r,svi)}$). The set of preconditions R_p in argument $D2$ is empty.It is assumed in this scenario that there is an availability of penicillin and means to administrate the antibiotic. Otherwise such facts should be added in the set of preconditions.

Note that the attacks made on argument A by $B1$, $C1$ and $D1$ are asymmetric (one way attacks), whereas the attacks on $B1$, $C1$ and $D1$ made respectively by $B2$, $C2$ and $D2$ are symmetric. The reason for these differing attack relations is that in the former case, arguments *in favour* of the proposed action are always based on an assumption that no contraindication exists; an assumption that is undermined by the attacking arguments. (e.g., $D1$ undermines A's default assumption of no contraindication by identifying a contraindication ($d\_p(d, svi)$). In the second case, complementary courses of actions are proposed

---

[18]Given that $\texttt{p\_r\_p(r,hiv)}$, the degree by which $\texttt{side\_effect(r\_p(r,hiv))}$ realises a $\texttt{sev\_inf(r)}$ is too weak.

ID	Type	Argument
A	AS1	argue({},{}, propose({av_org(d,lung), p_recip(r,lung)}, {transp(r,lung)}))
B1	AS2	argue(C,A, contra({d_p(d,s_h)}, {reject(r,lung)}, graft_failure))
B2	AS3	argue(C,A, no_side_effect({¬d_p(d,copd)}, {reject(r,lung)}))
C1	AS2	argue(C,A, contra({d_p(d,hiv)}, {r_p(r,hiv)}, sever_infect))
C2	AS4	argue(C,A, not_realised_goal({p_r_p(r,hiv)}, {r_p(r,hiv)}, sever_infect))
C3	AS6	argue(C,A, contra({r_p(r,superinf)}, sever_infect))
D1	AS2	argue(C,A, contra({d_p(d,sve)}, {r_p(r,svi)}, sever_infect))
D2	AS5	argue(C,A, preventive_action({treat(r,penicillin)}, {r_p(r,svi)}))
D3	AS2	argue(C,A, contra({p_r_p(r,pen_allergy)}, {r_p(r,anaphylaxis)}, sever_infect))

a)   b)
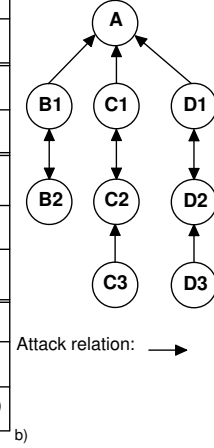
Figure 6: Example of three lines of argumentation structured reasoning: the B arguments which address the donor's smoking history (`d_p(d,s_h)`), the C arguments addressing the donor's HIV (`d_p(d,hiv)`); and the D arguments which address the fact that the donor's cause of death was *streptococcus viridans endocrditis* (`d_p(d,sve)`) which may result in the recipient of the lung contracting a *streptococcus viridans infection* (`r_p(r,svi)`). Each argument's $\mathcal{C}$ and $\mathcal{A}$ is updated according to the schemes' context updating rules.

to prevent undesirable side effects, where whether or not such prevention will be realised may still be a matter of debate. Hence, $D2$ attacks $D1$ by proposing $treat(r, penicillin)$ to prevent $reject(r, lung)$, where the efficacy of this preventative measure may still be debatable (implicitly then, $D2$ and $D1$ disagree on whether $d\_p(d, svi)$ is or not a contraindication). This disagreement is made explicit with a symmetric attack. To resolve whether the transplant is safe or not will require a decision as to whether or not $d\_p(d, svi)$ is a contraindication, that is, whether $D2$ is preferred to $D1$ or vice versa (this is further discussed in §7). Note however, that if a fourth argument $D3$ is submitted attacking argument $D2$, by indicating for instance that the potential recipient is allergic to penicillin, such an attack will again be asymmetrically directed on an assumption of argument $D2$ that no other contraindication exists. And so argument $D2$ does not defend itself against (i.e attack) $D3$ as would be the case with a symmetric attack.

Let us return to schemes $AS3$, $AS4$ and $AS5$ in order to identify their CQs. An argument instantiating schemes $AS3$ or $AS4$ introduces a new set of relevant facts R. An argument instantiating $AS5$ introduces a complementary course of actions A_c with a possibly empty set of preconditions R_p. At this stage R, R_p and A_c are taken to be the case (resp. intended), under assumptions Assum_2 and Assum_3a. As with arguments instantiating $AS2$, whether R and A_c are *relevant* is decided first at the Proxy Level (*e.g.* should argument $B2$ be accepted,
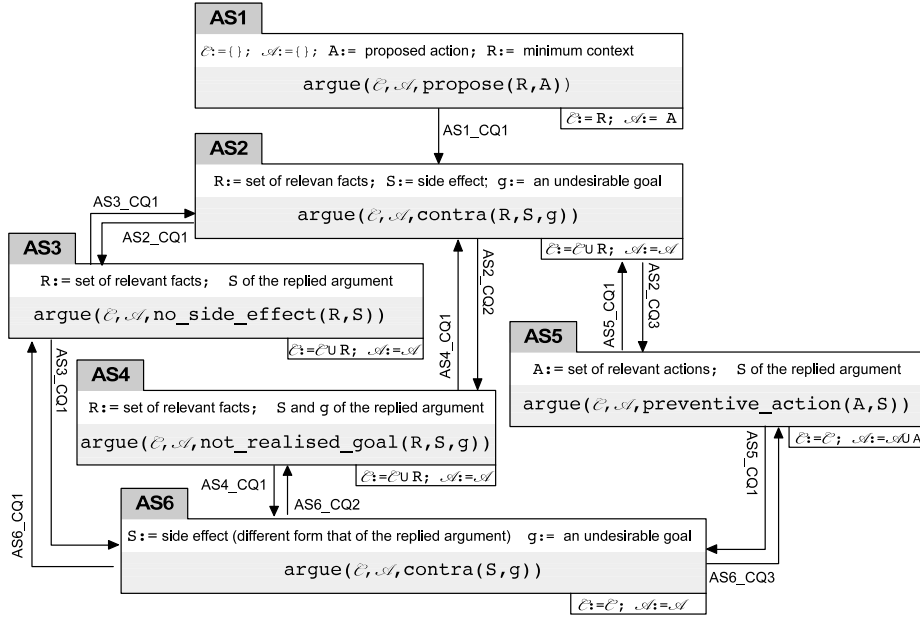
Figure 7: Argument Schemes connected via their associated Critical Questions

i.e., does $\{$`d_p(d,no_copd)`$\}\hspace{0pt}|\hspace{-3pt}\sim_{\mathcal{C}\wedge\mathcal{A}\wedge\Gamma}$ `side_effect`$(\{reject(r,lung)\})$ make sense) and latter at the *Resolution Stage* (*e.g.* does argument $B2$ defeats argument $B1$, *i.e.*, would $\{reject(r,lung)\}$ be prevented).

An argument, say Arg, that instantiates scheme $AS3$, $AS4$ or $AS5$, assumes (as in the case of the first submitted argument) that no (other) contraindication exists for performing the main action. This assumption is questioned by $AS3\_CQ1$, $AS4\_CQ1$ and $AS5\_CQ1$. As in $AS1$, such critical questions can only be addressed as attacks identifying the contraindications and the associated undesirable side effects. Such attacks can thus be embodied by arguments instantiating scheme scheme $AS2$, analogous to attacks on the first submitted argument by arguments instantiating $AS2$. However, this time, as a way to defend the main action's safety, Arg introduces a new set of factors (facts or actions) which themselves may warrant, respectively cause, some undesirable side effect. That is, this time, an attack can be made via $AS3\_CQ1$, $AS4\_CQ1$ and $AS5\_CQ1$ without having to introduce a new set of facts. Such attacks are embodied by argument scheme $AS6$ which differs from $AS2$ in that it does not require introducing an additional set of relevant facts R:

AS6	
Preconditions: $S \subseteq \mathbf{S}$, non-empty and **different** from the replied argument's stated effect, $g \in \mathbf{G}$; $\mathcal{C}$ and $\mathcal{A}$ the updated context of facts and actions of the replied argument.	
Body:	In circumstances $\mathcal{C}$ The actions $\mathcal{A}$ will cause a side effect S which will realise some undesirable goal g. ○ $\vdash_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma}$ side_effect(S); and ○ side_effect(S) $\vdash_{\mathcal{C} \wedge \mathcal{A} \wedge \Gamma}$ g; argue($\mathcal{C}$,$\mathcal{A}$,contra(S,g));
Critical Questions:	
CQ1: Are circumstances such that the stated side effect will not be achieved?	
CQ2: Are circumstances such that the achieved side effect will not realise the stated goal?	
CQ3: Is there a complementary course of action that prevents the achievement of the stated effect?	
Context Updating Rule: $\mathcal{C} := \mathcal{C}$; $\mathcal{A} := \mathcal{A}$.	

We can continue with our medical example to illustrate the use of schemes $AS2$ and $AS6$ in order to attack arguments instantiating schemes $AS3$, $AS4$ or $AS5$ (see fig. 6). Suppose, for instance, that the recipient to whom the lung is intended is *allergic to penicillin*. Thus, if as a way to prevent the recipient's bacterial infection penicillin is administered ($D2$), the allergic reaction may be quite severe, (*anaphylaxis*). Such an argument against the action's safety is embodied by $D3$ which instantiates scheme $AS2$. To illustrate the use of scheme $AS6$, let us continue with the argumentation line A, $C1$ and $C2$, where it has been argued that the lung may safely be transplanted despite the donor having *HIV* because the potential recipient already has the same viral infection. It is currently believed that in most cases such transplants will cause a *superinfection*[63], which is an uncontrolled, severe infection. Note that no new factors were introduced in order to attack argument $C2$. Thus, such an attack can be embodied by an argument $C3$ that instantiates $AS6$. In this basic circuit of schemes and critical questions, $AS6$'s critical questions are the same as those for $AS2$.

Figure 7 depicts the circuit of argument schemes connected via their associated critical questions presented in this section. In the following subsections

we relax some of the assumptions introduced in this subsection so as to address required extensions to this basic circuit.

5.2.2 Accounting for Specificity

Let us suppose now that a DA offers for transplantation the lung of a donor with a history of cancer (d_p(d,h_cancer)). The DA herself may argue that in such history the recipient will result having as a side effect cancer. As depicted in figure 8 this argument ($E1$) can be instantiated using scheme $AS2$. Let us suppose as well that the DA have added to $\mathbb{C}_F$ the fact d_p(d,h_nonmel_skin_c) meaning that the donor had a nonmelanoma skin cancer. A history of cancer is in general an excluding criteria for being a donor. However, for some kind of past malignancies, such as nonmelanoma skin cancer, the risk of transmitting the malignancy to the recipient is believed to be marginal [25]. Let us suppose the RA believes that to be the case and would wish to argue that for this particular type of cancer the transplant is safe. At first sight it may seem that this argument could be constructed by instantiating scheme $AS3$ with $R = $ {d_p(h_nonmel_skin_c)} being the new relevant set of facts. And so updating the local context of facts to be:

$\mathcal{C}$ = {av_org(d,lung),p_recip(r,lung),d_p(d,h_cancer),
d_p(d,h_nonmel_skin_c)}

Although clearly $\mathcal{C}$ holds ($\mathcal{C} \subseteq \mathbb{C}_F$), there is a bit of information that despite being important is not captured if $AS3$ is to be used. That is, d_p(d,h_cancer) and d_p(d,h_nonmel_skin_c) are not independent facts, the latter is a subclass of the former. Furthermore, there is an implicit assumption that donor had a history nonmelanoma skin cancer and no other type of cancer.
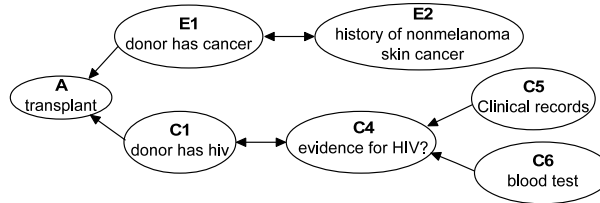
In order to account for this we need first to relax Assum_1a by associating to $\mathbf{R}$ a relation of specificity $\prec$ so as to account for the fact that, for instance, {d_p(d,h_nonmel_skin_c)} $\prec$ {d_p(d,h_cancer)}.

Having defined a taxonomy in $\mathbf{R}$ the circuit of schemes and CQs is extended. The CQs of the kind – *Are the current circumstances such that...?* – (*i.e.* $AS2\_CQ1$, $AS2\_CQ2$, $AS3\_CQ1$, $AS4\_CQ1$, $AS5\_CQ1$, $AS6\_CQ1$ and $AS6\_CQ2$) can now be embodied as an attack not only by schemes $AS2$, $AS3$ and $AS4$ but also by their *specific* versions $AS2s$, $AS3s$ and $AS4s$. Below we introduce only scheme $AS3s$, schemes $AS2s$ and $AS4s$ are defined analogously:

AS3s	
Preconditions: $R_g \subseteq \mathcal{C}$, $R_s \subseteq \mathbb{C}_F$, S of the replied argument $\mathcal{C}$ and $\mathcal{A}$ the context of facts and actions of the replied argument.	
Body:	Because R_s, a particular case of R_g, holds in circumstances $(\mathcal{C} - R_g)$ the side effect S is not expected as caused by $\mathcal{A}$.
	○ $R_s \prec R_g$ ○ $R_s \not\hspace{-2pt}\sim_{(\mathcal{C}-R_g)\wedge\mathcal{A}\wedge\Gamma} \texttt{side\_effect}(S)$
	`argue(`$\mathcal{C}$`,`$\mathcal{A}$`,no_side_effect(replace_s(`R_g, R_s`),`S`));`
Critical Questions: Same as $AS3$	
Context Updating Rule: $\mathcal{C} := (\mathcal{C} - R_g) \cup R_s$; $\mathcal{A} := \mathcal{A}$.	

The main change in these new schemes is the way the local context of facts $\mathcal{C}$ is updated. Instead of introducing an additional set of facts R (as it is the case with $AS2$, $AS3$ and $AS4$) a subset $R_g \subseteq \mathcal{C}$ is replaced by a more specific set of facts R_s ($R_s \prec R_g$). In this way, it is made explicit that R_g does not holds by itself, independent of R_s. Rather, R_g is the case only because R_s holds, since R_s entails R_g. Thus, for example, `d_p(d,h_cancer)` would hold only because `d_p(d,h_nonmel_skin_c)` is the case.

ID	Type	Argument
A	AS1	argue({},{}, propose({av_org(d,lung), p_recip(r,lung)}, {transp(r,lung)}))
E1	AS2	argue(C,A, contra({d_p(d,cancer)}, {r_p(r,cancer)}, cancer))
E2	AS3s	argue(C,A, no_side_effect(replace_s({d_p(d,cancer)},{d_p(d,h_nonmel_skin_c)}), {reject(r,lung)}))
C1	AS2	argue(C,A, contra({d_p(d,hiv)}, {r_p(r,hiv)}, sever_infect))
C4	AS2_CQ1	challenge(evidence(d_p(d,hiv)))
C5	AS2ev	argue(C,A, contra(replace_ev(d_p(d,hiv),{clin_rec(d,hiv)}), {r_p(r,hiv)}, sever_infect))
C6	AS2ev	argue(C,A, contra(replace_ev(d_p(d,hiv),{test(d,blood,hiv)}), {r_p(r,hiv)}, sever_infect))

Figure 8: Example illustrating the use of argument scheme AS3s and of a challenge.

To continue with our example, argument $E1$ can now be attacked by an argument $E2$ instantiating scheme $AS3s$ as follows:

$E2$: `argue(`
`{av_org(d,lung),p_recip(r,lung),d_p(d,h_cancer)},{transp(r,lung)},`
`no_side_effect(`
`replace_s({d_p(d,h_cancer)}, {d_p(d,h_nonmel_skin_c)}),`
`{r_p(r,cancer)}));`

With its updated local context of facts being:

$\mathcal{C}$ `= {av_org(d,lung),p_recip(r,lung),d_p(d,h_nonmel_skin_c)}`

5.2.3 Accounting for Uncertainty

As argued in §4.2, while *ProCLAIM* does not support a dispute regarding whether a fact in $\mathbb{C}_F$ holds or not, it is still important for the decision making to account for the evidence that supports facts asserted during the *Argumentation Layer*. Thus participants should be able to request for and provide such evidence. For example, the *RA* may want to know the evidence that supports the fact that the donor has HIV.

To enable this Assum_1a has to be relaxed by associating to **R** a defeasible consequence relation $\mathord{\sim}_{ev}$ where $Ev\mathord{\sim}_{ev}Fact$ indicates that a set of facts $Ev \subseteq \mathbf{R}$ is evidence in support of the fact $Fact \in \mathbf{R}$. For example `{clin_rec(d,hiv)}` $\mathord{\sim}_{ev}$`d_p(d,hiv)` indicating that donor's clinical records support the fact that the donor has HIV.

Secondly, the circuit of schemes and CQs has to be extended so that argument schemes that introduce relevant set of facts R[19], for each asserted fact $r_i \in R$, there is an associated CQ of the form – *Is there evidence to believe r_i is the case?* –. Now, this CQ is indeed intended to question r_i so that participants have to provide evidence in its support. However, it is not intended for participants to argue that r_i is false, for this should be resolved outside *ProCLAIM*. Thus, such CQs are formalised only as challenge locutions:

`challenge(evidence(`r_i`))`

Where in their replay is expected an argument that provides the evidence, a set of facts ($Ev \subseteq \mathbb{C}_F$), in support of r_i ($Ev\mathord{\sim}_{ev}r_i$). So, a challenge directed on argument $C1$ `challenge(evidence(d_p(d,hiv)))` may be replied providing the set of facts `{clin_rec(d,hiv)}`.

The purpose of these CQs is to allow bringing in the evidence on which the introduced facts are based. In so doing the inherent uncertainty of the facts conforming to the circumstances in which the decision making takes place is made explicit. In this way, decisions are made accounting for this uncertainty, which may, of course, motivate further enquiries in order to make more informed decisions. For example, doctors may proceed to perform a serological (blood)

---

[19]These are schemes $AS2$, $AS2s$, $AS3$, $AS3s$, $AS4$ and $AS4s$.

test on the donor in order to have more conclusive evidence on whether the donor does actually have `HIV`. However, while the results of any such enquiry can be fed into *ProCLAIM*'s deliberation by updating $\mathbb{C}_F$, the actual enquiry is not formalised by *ProCLAIM*.

As stated above these CQs are associated to any argument scheme that defines the introduction of a new set of facts, *i.e.* to schemes $AS2$, $AS2s$, $AS3$, $AS3s$, $AS4$ and $AS4s$. Here we present only scheme $AS2ev$ which should be instantiated to construct an argument in reply to a challenge made on an argument instantiating $AS2$ or $AS2s$. The other schemes ($AS3ev$ linked to $AS3$ and $AS3s$ and scheme $AS4ev$ linked to $AS4$ and $AS4s$) are defined analogously:

AS2ev

Preconditions: r_i the questioned fact, $R_{ev} \subseteq \mathbb{C}_F$, S and g of the argument being challenged, and $\mathcal{C}$ and $\mathcal{A}$ its updated context of facts and actions.

Body:

R_{ev} is evidence for r_i being the case, and such that
in circumstances $(\mathcal{C} - \{r_i\}) \cup R_{ev}$
actions $\mathcal{A}$ will cause a side effect S
which will realise some undesirable goal g.

$\circ \quad R_{ev} \mathrel{\mid\!\sim}_{ev} r_i$

$\circ \quad R_{ev} \mathrel{\mid\!\sim}_{(\mathcal{C} -\{r_i\}) \wedge \mathcal{A} \wedge \Gamma} \texttt{side\_effect}(S)$; and

$\circ \quad \texttt{side\_effect}(S) \mathrel{\mid\!\sim}_{R_{ev} \wedge \mathcal{C}_i \wedge \Gamma} g$

$\texttt{argue}(\mathcal{C},\mathcal{A},\texttt{contra}(\texttt{replace\_ev}(r_i, R_{ev}),S,g));$

Critical Questions: Same as $AS2$ and $AS2s$ to which we now add the CQs

CQ4$_i$: Is there evidence to believe r_i is the case? ($r_i \in R$,
R the new introduced set of facts)

Context Updating Rule: $\quad \mathcal{C} := (\mathcal{C} - \{r_i\}) \cup R_{ev}$; $\mathcal{A} := \mathcal{A}$.

Note that an argument instantiating scheme $AS2ev$ not only provides the evidence (R_{ev}) supporting the challenged fact, but its claim is that if the asserted fact is replaced by the evidence on which it is based on the same undesirable side effects will be caused (see figure 8.). Analogously arguments instantiating scheme $AS3ev$ will claim that the side effect is not expected and; arguments instantiating scheme $AS4ev$ will claim that the side effect are not undesirable in this updated circumstances.

The lack of evidence to support a challenged fact may motivate participants to get that evidence within the deliberation (*e.g.* perform a serological test

on the donor: `test(d,blood,hiv)`). However, it may well be the case that such evidence cannot be acquired, so leaving a challenge weakly replied, or even unreplied. This may lead PAs to retract the challenged fact and so subtract it from $\mathbb{C}_F$, which brings us to the next extension to the circuit, accounting for incomplete information.

Whether $\mathbb{T}$ is left with uncertain or unknown facts, decision makers will still have to decide what to do. Having resolved which the preferred arguments are in $\mathbb{T}$, if the safety of the action amounts to deciding whether some uncertain and/or unknown facts are the case or not, such resolution would plausibly aim to assess the likelihood of these facts being the case, accounting for the risk involved in them being or not the case. While *ProCLAIM* aims to identify the relevant facts and the risk involved in them being or not the case, it is not intended for addressing the resolution process of weighting likelihood versus risk. This is further discussed in §7.

5.2.4 Accounting for Incomplete Information

Players may start the deliberation with a set of facts believed to be the case, $\mathbb{C}_F$, and during the argumentation process realise that some potentially relevant information, say r, is missing. That is, $\neg r, r \notin \mathbb{C}_F$. But still, even if some facts are unknown, a decision needs to be made on whether or not to perform the proposed action. Decision makers should be made aware that potentially relevant information is missing. To account for this situation, the argumentation circuit is extended so that participants can submit arguments that introduce a set of fact R as relevant, despite $R \nsubseteq \mathbb{C}_F$.[20] That is, while it is argued that R is relevant, it is unknown whether it holds or not. In that way, participants can make explicit that some data, presumed to be relevant, is missing. And so, they can submit *hypothetical* arguments. Arguments of the form *–If R were the case, then...–*.

These hypothetical arguments are formalised in exactly the same manner as those presented above, the only difference is that we now have relaxed the precondition that facts used in an argument must be in $\mathbb{C}_F$. That is, we relax the assumption Assum_2. In general, updates at the *Argumentation Layer* can be made independently from to those at the *Context Layer*, and vice versa. This independence results in the definition of three types of arguments:

Definition 5.3 *Suppose $\mathcal{C}$ is the updated local context of facts of an argument Arg, then:*

- *If $\mathcal{C} \subseteq \mathbb{C}_F$, Arg is a **factual argument**.*

- *If $\exists r \in \mathcal{C}$ s.t. $\neg r \in \mathbb{C}_F$, Arg is a **overruled argument**.*

- *Otherwise, Arg is a **hypothetical argument**.*

---

[20]Assuming, as we will later see, that $R \cup \mathcal{C} \nvdash_\Gamma \bot$

To illustrate a practical use of hypothetical arguments, let us introduce a new organ acceptability criterion from [27]: "*For pancreas transplantation, guidelines suggest that donor age should be less than 45 yr; nonetheless, using pancreas with good appearance on inspection after retrieval from donors aged 45-62 yr; can achieve the same graft survival as pancreas from donors aged under 45 ys.*". Hence, if a donor is elderly (over 45 years, for the pancreas case) and her pancreas is transplanted it is likely that it will be rejected, and so realising a graft failure. Unless, the pancreas has *good appearance*. However, in order to check the pancreas' appearance, the organ must first be retrieved. Hence, the transplant should have been deemed safe, at least provisionally.

Let us suppose that a pancreas of a 60 year old donor is available with the donor having `hcv`. Suppose the DA offers the pancreas (argument A, see figure 9) and argues that: *1)* because the donor is elderly, the recipient will reject the organ (argument $G1$, instantiating scheme $AS2$), and *2)* that the donor's `hcv` is a contraindication (argument $H1$, instantiating $AS2$), unless the recipient already has this same infection (hypothetical argument $H2$, instantiating $AS4$). Suppose that, in response to DA's submitted arguments the RA adds to $\mathbb{C}_F$ the fact `p_r_p(r,hcv)` (the recipient has `hcv`) and so making argument $H2$ factual. Also, let us suppose the RA submits the hypothetical argument $G2$ that instantiates $AS3$ as follows:

$G2 =$ `argue(`$\mathcal{C}, \mathcal{A},$ `no_side_effect({o_p(d,pancreas, good_app)},`
    `{reject(r,pancreas)}) )`

with `o_p(d,pancreas, good_app)` indicating that the donor's pancreas has good appearance. Argument $G2$ can only become factual once the organ is retrieved. Taking this into account, and supposing arguments $G2$ and $H2$ are deemed preferred to $G1$ and $H1$ respectively (fig. 9 *c.*), the pancreas will be deemed suitable for this recipient, subject to the organ's appearance on retrieval. That is, if after retrieval `o_p(d,pancreas, good_app)` holds, the organ can be transplanted, otherwise the transplant should not be performed, argument $G2$ would become overruled.

Note that, if the potential recipient does not have `hcv` ($\neg$`p_r_p(r,hcv)` $\in \mathbb{C}_F$), the transplant should be deemed unsafe, irrespective of the pancreas' appearance (fig. 9 *d.*). Or similarly, if $H1$ would have been submitted as a hypothetical (it is unknown whether the donor has `hcv` or not) and $H2$ as factual, what becomes irrelevant, for deciding the action's safety, is whether the donor has or not `hcv`. Namely, hypothetical and factual arguments together indicate which of the unknown facts are worth checking to see whether they hold.

The independence between the elements of $\mathbb{C}_F$ and $\mathcal{C}$ makes each argument potentially factual, overruled or hypothetical. To allow for such independence we have relaxed the precondition that each additional set of facts R must be in $\mathbb{C}_F$. Because we have defined $\mathbb{C}_F$ such that it has to be a consistent set of facts ($\mathbb{C}_F \nvdash_\Gamma \bot$) this precondition enforced that each argument's context has to be, in turn, consistent. To preserve such a property with the hypothetical arguments, we must ensure that each additional set of facts R is consistent with the

ID	Type	Argument
A	AS1	argue({},{}, propose({av_org(d,pancreas), p_recip(r,pancreas)}, {transp(r,pancreas)}))
G1	AS2	argue(C,A, contra({d_p(d,elderly)}, {reject(r,pancreas)}, graft_failure))
G2	AS3	argue(C,A, no_side_effect({**o_p(d,pancreas,good_app)**}, {reject(r,lung)}))
H1	AS2	argue(C,A, contra({d_p(d,hcv)},{r_p(r,hcv)}, sever_infect))
H2	AS4	argue(C,A, not_realised_goal({p_r_p(r,hcv)}, {r_p(r,hcv)}, sever_infect))
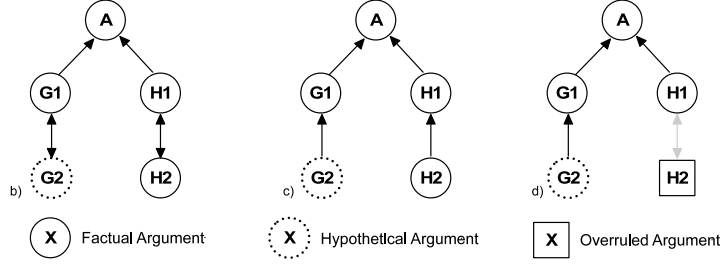
Figure 9: Example illustrating the use of hypothetical arguments.

elements of the argument's local context $\mathcal{C}$. To do so, to schemes $AS2$, $AS2s$, $AS2ev$, $AS3$, $AS3s$, $AS3ev$, $AS4$, $AS4s$, $AS4ev$, we add the precondition:

The introduced set of relevant facts R must be such that $R \cup \mathcal{C} \nvdash_\Gamma \bot$

5.2.5   Discussing Further Extension

There are a number of extensions that can be proposed to this circuit of schemes and CQs. Any extension involves *1)* identify the motivating set of examples that needs to be addressed; *2)* relax the appropriate assumptions and finally; *3)* define the procedure through schemes and CQs for capturing the right relation among the sets **R**, **A**, **S** and **G** while appropriately updating the sets $\mathcal{C}$ and $\mathcal{A}$. Each such procedure, argument scheme, must be motivated by a change in the assessment on the main action's safety (within the local contexts of facts an actions). In this subsection we describe a few extensions we are currently formalising.

The first required extension is intended to allow PAs to point at actions that are incompatible across different local contexts of actions. Take for example two complementary actions A_{c1} and A_{c2} that are proposed each to mitigate or prevent different side effects highlighted in a different branch of $\mathbb{T}$. Each action corresponds to a different local contexts: say $< \mathcal{C}_1, \mathcal{A}_1 >$ and $< \mathcal{C}_2, \mathcal{A}_2 >$. Suppose that A_{c1} and A_{c2} are such that when performed together they cause an undesirable side effect. Firstly to address this example assumption Assum_3c has to be relaxed, so that complementary actions can be deemed in conflict. Secondly a procedure must be defined by which an undesirable side effect is caused when the A_{c1} and A_{c2} are jointly performed. This suggest that the

update of the local contexts $< \mathcal{C}_1, \mathcal{A}_1 >$ and $< \mathcal{C}_2, \mathcal{A}_2 >$ is for them to be merged ($i.e.$ $< \mathcal{C}_1 \cup Lfd_2, \mathcal{A}_1 \cup \mathcal{A}_2 >$), capturing the fact that these local context are no longer $independent$.

Another extension related with actions involves making a distinction between $intending$ and merely $proposing$/suggesting an action. For example, it may seam reasonable that while a RA can argue that he $intends$ to treat the recipient with antibiotics to prevent a certain infection, the DA can only $suggest$ treatments on the recipient. This can be formalised in a similar fashion as we did in §5.2.4 to address the problem of incomplete information. Relaxing Assum_3a, so that an argument instantiating scheme $AS5$ can use complementary actions that are not in $\mathbb{C}_A$, redefine the $AS5$'s preconditions and identifying which are the factual, hypothetical and overruled arguments.

The last extension we discuss here is intended to allow addressing the fact that in some circumstances any alternative to performing the main proposed action will derive in more undesirable consequences than the side effects caused by the proposed action. Thus, PAs should be able to question the degree of undesirability of goals. Questioning, for example, whether `cancer` is undesirable enough as a side effect of a organ transplant when any alternative to the organ transplant will result in the death of the potential recipient. To address this example Assum_1d must be relaxed by associating to **G** a relation of undesirability, next Assum_4 needs to be relaxed so that not any realised $g \in \mathbf{G}$ is reason enough so as to abort the proposed action. Finally the appropriate procedure has to be defined.

6 Using the ASR to deliberate

Once the circuit of schemes and CQs is defined, and tailored to encode stereotypical reasoning patterns for deliberating over safety-critical actions, we can further specialise this circuit to a particular application, $e.g.$ the transplant or environmental scenario.

To illustrate, let us consider the argument scheme $AS1$ in which, given the preconditions R_p, an action A_m is proposed. In the transplant scenario the proposed action is always the same: $transplant\ an\ organ$, and the preconditions are: to have an $available\ organ$ for the $potential\ recipient$. Of course, in each instance the donor, the recipient and the organ are different. Thus, tailoring $AS1$ to the transplant scenario involves capturing this recurrent pattern while allowing for different donor, organ and recipient instantiation. This can be done by ungrounding the predicates `av_org(donor,organ)`,`p_recip(recipient,organ)` and `transp(recipient,organ)`. So denoting variables with upper-case letters we can define the tailored version of $AS1$ as:

$AS1_T :$ `argue({},{},propose({av_org(D,O),p_recip(R,O)},{transp(R,O)}))`

The scenario-specific schemes and CQs are encoded in the ASR. The MA references this repository in order to provide the legal replies to an argument.

In so doing, *ProCLAIM* facilitates a highly focused deliberation, paramount for its intended applications. This is not only because participants are directed in their argument submission to a degree where they only need to fill in some blanks (as in $AS1_T$), but also, in referencing the ASR the MA can easily identify the arguments that though logically valid, make little sense in the application scenario. Furthermore, the specialisation of the ASR plays an important role in the CBRc retrieval process, helping identify potentially similar cases (broadly speaking, cases in which the same reasoning patterns – specialised schemes – were used), as discussed in §8.

6.1   Constructing an ASR

Given a new scenario application, one of the most important tasks when instantiating the *ProCLAIM* framework, is the construction of the ASR, which defines what is to be argued about and how. In this section we describe a procedure for its construction, by way of illustrating with the transplant scenario.

Firstly, the ASR developers[21] must identify the type of information to be used in the deliberation. This information is encoded in the sets $\overline{\mathbf{R}}$, $\overline{\mathbf{A}}$, $\overline{\mathbf{S}}$ and $\overline{\mathbf{G}}$, which respectively denote the ungrounded versions of $\mathbf{R}$, $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{G}$. That is, if for example, `av_org(d,lung)` $\in \mathbf{R}$ then, `av_org(D,O)` $\in \overline{\mathbf{R}}$. The next step is to choose the (type of) safety-critical action to be argued about (*e.g.* {`transp(R,O)`}$\subseteq \overline{\mathbf{A}}$) and identify a set of preconditions required for the action's performance (*e.g.* {`av_org(D,O)`,`p_recip(R,O)`}$\subseteq \overline{\mathbf{R}}$). For each chosen set of actions $A_i \subseteq \overline{\mathbf{A}}$ and their set of preconditions $R_{p\_i} \subseteq \overline{\mathbf{R}}$ developers can define the specialised versions of $AS1$:

$AS1_i$: `argue(`{}`,`{}`,propose(`R_{p_i}, A_i`))`

To each such $AS1_i$ there is associated the CQ $AS1_i\_CQ1$: *–Is there any contraindication for performing action A_i?–*, which can be embodied as an attack by a specialised version of $AS2$. Thus, given a specialisation of $AS1$ developers must produce specialised versions of $AS2$. Any specialised version of $AS2$ that replies to an argument instantiating $AS1_T$ is of the form:

$AS2_T$: `argue(`{`av_org(D,O)`,`p_recip(R,O)`}`,`{`transp(R,O)`}`,contra(R,S,g))`[22]

Now, for each undesirable goal that the action can bring about, (*e.g.* `sev_inf`, `cancer`, `grft_fail`, `death`,...) there is a partially specialised version of $AS2$, e.g.:

$AS2_{T\_gf}$: `argue(`{`av_org(D,O)`,`p_recip(R,O)`}`,`{`transp(R,O)`}`,`
`contra(`R, S`,grft_fail(R)))`

---

[21]Most naturally, the construction of the ASR will be carried out mainly by computer science developers under the supervision of domain experts.

[22]Note that R is a set of facts and `R` is a variable bounded by `p_recip(R,O)` and `transp(R,O)`.

Developers must now identify the type of side effects S that realise each of these undesirable goals, and in turn, identify which are the type of contraindications R that may lead the main action to cause these side effects. Thus, for example, a graft failure occurs when a recipient rejects the organ (`{reject(R,O)}`) which may be because of a donor property (`{d_p(D,P)}`, *e.g.* `d_p(d,s_h)`), due to a blood mismatch (*e.g.* `{blood(D,BtypeD)`, blood(R,BtypeR)`}`) or because of a combination of the organ property and recipient property (`{o_p(O,Po)`, p_r_p(R,Pr)`}` *e.g.* the lung is too big for the recipient's thoracic cavity), *etc.* Each of these combinations constitutes a specialised version of $AS2$:

$AS2_{T\_gf1}$: `argue({av_org(D,O),p_recip(R,O)},{transp(R,O)},`
`contra({d_p(D,P)},{reject(R,O)},grft_fail(R)))`

$AS2_{T\_gf2}$: `argue({av_org(D,O),p_recip(R,O)},{transp(R,O)},`
`contra({blood(D,BtypeD), blood(R,BtypeR)},{reject(R,O)},grft_fail(R)))`

$AS2_{T\_gf3}$: `argue({av_org(D,O),p_recip(R,O)},{transp(R,O)},`
`contra({o_p(O,Po),p_r_p(R,Pr)},{reject(R,O)},grft_fail(R)))`

Figure 10: ASR builder.

Now, to each such specialised schemes there are associated the CQs of the scheme $AS2$, which should direct developers in further constructing the ASR. For example, respectively embodying the Critical Questions $AS2_{T\_gf1}\_CQ1$, $AS2_{T\_gf1}\_CQ2$ and $AS2_{T\_gf1}\_CQ3$ are the specialised schemes:

$AS3_{T\_gf1\_1}$: `argue({av_org(D,O),p_recip(R,O), d_p(D,P)},{transp(R,O)},`
`no_side_effect({d_p(D,P2)},{reject(R,O)}))`

$AS4_{T\_gf1\_1}$: `argue({av_org(D,O),p_recip(R,O), d_p(D,P)},{transp(R,O)},`
`not_realised_goal({p_r_p(R,Pr)},{reject(R,O)},grft_fail(R)))`

$AS5_{T\_gf1\_1}$: `argue({av_org(D,O),p_recip(R,O), d_p(D,P)},{transp(R,O)},`
`preventive_action({treat(R,T)},{},{reject(R,O)})))`

The process continues in a similar way with each of these specialised schemes. Developers are thus directed in the construction of the ASR by the circuit of schemes and CQs described in §5. Furthermore, as described in §5, schemes are also represented in natural language form, so that each specialised scheme will have an associated natural language version *e.g.*:

$AS5_{T\_gf1\_1}$: The organ rejection can be prevented by treating the recipient with T.

Figure 11: ASR Browser.

As noted earlier, the ASR construction is the key activity when instantiating *ProCLAIM* for use in a given domain. It requires effort from both computer science developers and domain experts, none of whom may be familiar with argumentation. To facilitate their task we have developed two online tools developed in PHP and MySQL: the first one intended to assist developers in the the step by step ASR construction[23] (see figure 10) and another tool[24] which allows domain experts to navigate the natural language forms of the ASR's schemes and CQ (see figure 11). These tools are currently in a prototype phase of development and provides a useful proof of concept illustrating the potential value of our approach.

We are currently also using these tools in application of *ProCLAIM* to the environmental domain [52]. In this alternative scenario, decision makers must

---

[23]http://www.lsi.upc.edu/∼tolchinsky/newASR
[24]http://www.lsi.upc.edu/∼tolchinsky/ASR

decide whether an industrial wastewater can safely be discharged into a wastewater treatment plant. Hence, for this case, the potentially relevant facts for deciding the action safety include the industrial spill's content, the treatment plant's conditions and characteristics as well as external factors such as weather conditions. The undesirable side effects a spill may bring about relate to problems that can occur in the treatment plant, which in turn may cause ecological imbalances in the fluvial ecosystem. The actions that may prevent or mitigate such undesirable side effects include the use of different organic and chemical products on the wastewater. These factors are combined following the procedures introduced here, to build the ASR for the environmental scenario (see [52]).

In this subsection we have illustrated how the full space of argumentation can be codified in the ASR in a form useful for artificial and human agents. In the following subsection we show how this effort enables a highly focused deliberation process among heterogeneous agents.

6.2 MA's guiding task

The deliberation begins with an argument proposing the main action, through instantiation of a specialised version of $AS1$ in the ASR. The basic idea is that an action (*e.g.* {transp(R,O)}) can only be proposed if the precondition (*e.g.* {av_org(D,O),
p_recip(R,O)}) are met. In the transplant scenario, as soon as there is an available organ (*e.g.* `kidney`) of a donor (*e.g.* `d`) for a potential recipient (*e.g.* `r`) $AS1_T$ can be instantiated automatically and the deliberation is triggered:

```
inform(ma,all,conv_id,0,-1
    open_dialogue(
        propose({av_org(d,kidney),p_recip(r,kidney)},{transp(r,kidney)})
```

The DA that offers the organ and the RA responsible for the potential recipient may then enter the dialogue, first submitting a request:

```
request(da_id,ma,conv_id,1,0,enter_dialogue(proposal,DA, d_basic_info))
```

```
request(ra_id,ma,conv_id,2,0,enter_dialogue(proposal,RA, d_basic_info))
```

which, if accepted by the MA are replied to with an inform message broadcasted to all participants:

```
inform(ma,all, conv_id,3,1,
    entered_dialogue(proposal,da,d_basic_info,{ma},ℂ_{F∧A},𝕋, legal_replies))
```

```
inform(ma,all, conv_id,4,2,
    entered_dialogue(proposal,ra,r_basic_info,{ma,da},ℂ_{F∧A},𝕋,legal_replies))
```

Where, for example: `d_basic_info = {d_p(d,sve), d_p(d,young), loctn(d,hosp1),`
`blood(d,ab+)...}`; and `r_basic_info={loctn(r,hosp2),blood(r,ab+)...}`.

With these two messages $\mathbb{C}_F$ is updated to contain `d_basic_info` and `r_basic_info`. $\mathbb{C}_A$ contains only action `transp(r,kidney)` and $\mathbb{T}$ contains only the initial proposal, say $A1$.

Note that in these broadcasted messages the MA already informs the participants of the possible lines of attack on each argument in $\mathbb{T}$. In this example these are the replies to argument $A1$. Among these legal replies are the specialised schemes:

$AS2_{T\_inf1}$: `argue(`$\mathcal{C}$`,`$\mathcal{A}$`, contra({d_p(d,Pd)},{r_p(r,Pr)},sev_inf(r)))`

$AS2_{T\_gf3}$: `argue(`$\mathcal{C}$`,`$\mathcal{A}$`, contra({o_p(lung,Po),p_r_p(r,Pr)},`
`{reject(r,lung)},grft_fail(r)))`

$AS2_{T\_cncr3}$: `argue(`$\mathcal{C}$`,`$\mathcal{A}$`, contra({o_p(lung,Po)},{r_p(r,cancer)},`
`cancer(r)))`

where $\mathcal{C} = \{$`av_org(d,kidney),p_recip(r,kidney)`$\}$ and $\mathcal{A} = \{$`transp(r,kidney)`$\}$.

Hence, if RA were to argue that the donor's `sve`[25] may cause a streptococcus viridans infection (`svi`) in the recipient, RA would select scheme $AS2_{T\_inf1}$ among the legal replies and would require only to replace `Pd` with `sve` and `Pr` with `svi` in order to construct the desired argument.
`request(ra_id,ma,conv_id,5,4,`
`    argue(argue(`$\mathcal{C}$`,`$\mathcal{A}$`, contra({d_p(d,sve)},{r_p(r,svi)},sev_inf(r)))).`

If the argument is accepted the MA will broadcast the submitted argument providing the participants a list of legal replies such as:

$AS3_{T\_inf1}$: `argue(`$\mathcal{C}$`,`$\mathcal{A}$`, no_side_effect({d_p(d,Pd)},{r_p(r,svi)}))`

$AS4_{T\_inf1}$: `argue(`$\mathcal{C}$`,`$\mathcal{A}$`, not_realised_goal({p_r_p(d,Pd)},{r_p(r,svi)},`
`    sev_inf(r)))`

$AS5_{T\_inf1}$: `argue(`$\mathcal{C}$`,`$\mathcal{A}$`, preventive_action({treat(r,T)},{},{r_p(r,svi)})`

$AS2\_inf1\_CQ4_1$: `challenge(d_p(d,sve))`

Where $\mathcal{C} = \{$`av_org(d,kidney),p_recip(r,kidney),d_p(d,sve)`$\}$

Thus, with each new argument or challenge added to $\mathbb{T}$, MA provides the participants with challenges ready to be submitted and tailored schemes that are partially instantiated and require only to be filled in. For example, scheme $AS5_{T\_inf1}$ requires only to identify an action `T` (*e.g.* penicillin or teicoplanine) that would prevent a recipient infection. While scheme $AS3_{T\_inf1}$ guides participants to consider whether the donor has any property `Pd` which would make the transplant safe in spite of having `d_p(d,sve)`[26].

_____

[25]Note that `d_p(d,sve)`$\in\mathbb{C}_F$.
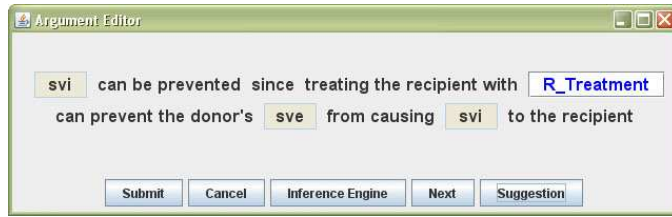[26]To our knowledge, no such property is known. Nonetheless, there may be in the future.

Figure 12: The argument editor of the application presented in [53]. This panel provides a human user with a legal reply ($AS5_{T\_inf1}$) to an argument. The *Inference Engine* button will validate the argument according to the knowledge base of an artificial agent that aids the user in the deliberation. The *Next* button provides the user with another legal reply (another scheme, *e.g.* $AS4_{T\_inf1}$) and button *Suggestion* proposes a scheme instantiation suggested by the artificial agent's knowledge base.

As noted above in §6.1, the ASR encodes each scheme in a '*code*' format useful for artificial agents and a natural language representation form for human users. In general an application instantiating *ProCLAIM* can feature a number of visualisation modes. For example, the content of the deliberation can be displayed as a tree of interacting nodes, where nodes are labelled with the relevant facts or actions the arguments highlight. Upon clicking on a node, its full natural language representation can be displayed, as well as legal replies presented as natural language templates for instantiation by the user. In [53], an artificial agent guided users in the argument submission and possible scheme instantiations (see fig.12), and validated alternative instantiations proposed by users. The user ultimately decided the argument to submit, and engaged in deliberation with an artificial agent that interacted solely with the schemes' formatted in Prolog code. This Application was implemented using two software components developed in the FP6 European project ASPIC[27]: an Argumentation Engine used for the agents' reasoning and a Dialogue Manager which with some extensions acted as the MA. The application was mainly implemented in Java, using Jade[28] for the agents' implementation and the Argumentation Engine was partially implemented in Java and prolog (see [53]) .

Having submitted the arguments, PAs may submit the `no_more_moves()` move. If all PAs do so, or the MA submits the `time_out(reason)` locution, then the *Resolution Stage* is initiated and the MA may proceed to evaluate $\mathbb{T}$.

7   Argument Evaluation

In the *Resolution Stage*, the MA has to evaluate the status of arguments in $\mathbb{T}$. This involves applying Dung's theory in order to identify which are the

---

[27]http://www.argumentation.org/
[28]http://jade.tilab.com/

53

defensible, defeated and justified arguments in $\mathbb{T}$. If the argument at the root of $\mathbb{T}$ is evaluated to be justified, the action is deemed safe, whereas if defeated, then the lines of arguments that lead to the rejection, identify the contraindications that warrant deeming the action as unsafe.

Prior to computing the arguments' status MA has to: *1)* reference the DCK and the CBRc as these component may submit additional arguments; *2)* assign a preference relation between arguments that mutually attack each other, since symmetric attacks may prevent a definitive status evaluation of the main proposed argument; and finally *3)* appropriately deal with arguments that are hypothetical or that are not well defended from a challenge made on them.

As discussed in §2, MA may submit additional arguments by referencing both the DCK and the CBRc. The first knowledge resource will help identify any fact in $\mathbb{C}_F$ or alternative complementary course of action in **A** that though *relevant* for the decision making, according to *domain consented knowledge*, was not taken into account by the PAs. These relevant factors are added to $\mathbb{T}$ by means of submitting an argument that instantiates the appropriate legal reply. The second knowledge resource will propose for submission arguments that have been submitted in previous similar cases but are not in the current $\mathbb{T}$(see §8). Thus, in this task, the MA plays the role of two additional PAs: an expert or specialist in domain consented knowledge, and another specialist in reusing evidence collected from past deliberations. Consider for instance the argumentation line A, $D1$, $D2$ and $D3$ in §5.2.1 and illustrated in Figure 13c, where, in order to prevent a streptococcus infection on the recipient (argument $D1$) PA proposed administrating the patient *penicillin* ($D2$). However, the patient was allergic to penicillin ($D3$). The MA can use DCK or CBRc to suggest an alternative antibiotic (*e.g.* teicoplanine) for preventing the infection on the recipient. This can be posed as an argument ($D4$ in fig. 13) instantiating scheme $AS5_{T\_inf1}$.

In the same way as regular PAs can endorse which arguments they support at the *Endorsement Layer*, the DCK and CBRc can assign a preference relation between mutually attacking arguments. This is the second task of the MA. Let us suppose, for now, that $\mathbb{T}$ contains no hypothetical arguments and all challenges are successfully replied to. The problem with symmetric attacks is that they may prevent evaluating the root argument's main action proposal as either justified or defeated (that is, it may be defensible). To solve this impasse a preference relation is assigned between mutually attacking arguments so as to decide which asymmetrically attacks the other. Consider the argumentation framework with arguments A, $B1$ and $B2$ in §5.2.1, where the debate is whether an available lung is viable when the donor has a smoking history but no COPD (see fig. 13a.). In this example deciding whether the transplant is safe or not amounts to deciding whether argument $B1$ is preferred to $B2$ (and so asymmetrically attacks $B2$) or $B2$ is preferred to $B1$ (and so asymmetrically attacks $B1$). In the first case the root argument would be rejected under the grounded semantics, and so the transplant would be deemed unsafe (a graft failure is expected because of the donor's smoking history, even though the donor has no COPD). In the later case the root argument would be justified under the

grounded semantics, and so the transplant deemed safe because it is believed that if {d_p(d,no_copd)} holds {d_p(d,s_h)} is not a contraindication.

The preference relation between mutually attacking arguments is determined by *ProCLAIM*'s three knowledge resources DCK, CBRc and AEM. Each knowledge resource will provide its own perspective on the arguments' relative strength. The DCK will derive the preference assignment from the standard guidelines and regulations of the domain (*e.g.* donor and organ acceptability criteria). For instance, if a lung of a donor with a smoking history but no COPD is deemed viable for transplantation according to current agreed-upon transplant guidelines the DCK will deem $B2$ preferred to $B1$. The CBRc will derive its assignment from previous similar cases. Broadly speaking, if arguments similar to $B2$ have been *successfully* deemed preferred to arguments similar to $B1$ in previous similar cases the CBRc will also deem $B2$ as preferred to $B1$ (as discussed in §8). Finally the AEM takes the PAs' endorsement moves (those in $\mathbb{E}$) and assigns weights based on a measurement of trust specific to the application scenario (*e.g.* based on the role of the PAs or the prestige of the transplant unit the PAs represent). Thus, if more trusted agents have endorsed $B2$ rather than $B1$, the AEM will deem $B2$ as preferred to $B1$.

In this simple example, $B2$ is deemed preferred to $B1$ and the MA may broadcast a solution of the kind: *–the transplant is safe–*. However, not only may different knowledge resources yield conflicting preferences (DCK may deem $B1$ as preferred to $B2$ while the CBRc may deem the opposite), but their preference assignments may vary in degrees of confidence. Furthermore: novel proposals from PAs may lead to situations about which the DCK has little knowledge; depending on the case the CBRc may have more or less evidence to prefer one argument over another; and finally, equally trustful PAs may each endorse a competing argument preventing the AEM from deeming one argument as stronger than the other.

To address these issues, we maintain the independence of the preference assignments so that the final decision makers have an account of the different perspectives' recommendations. Let us then define the preference assignment as a mapping:

$$pref : \mathcal{A} \times \mathcal{A} \mapsto [-1,1] \times [-1,1] \times [-1,1]$$

Thus, $\texttt{pref}(A1,A2) = (a,b,c)$, where a is the preference assignment of the DCK, b of the CBRc and c of the AEM, and where positive values express a preference for the first argument over the later ($A1$ preferred to $A2$) and negative values the opposite. Zero means there is no preference at all. The bigger the absolute value of the number, the more the confidence in the preference assignment. Thus if $\texttt{pref}(A1,A2) = (-1,-1,-1)$ then $A2$ is deemed preferred to $A1$ with full confidence. When the preference assignments are not all in agreement, say for instance $\texttt{pref}(A1,A2) = (0.2, -0.6, -0.5)$, then decision makers must decide whether or not to override guidelines ($A1$ preferred to $A2$ with confidence 0.2), and trust the PA's assessment knowing that he is a reliable expert ($A2$ preferred to $A1$ with confidence 0.5) and his opinion is backed by evidence ($A2$ preferred

to $A1$ with confidence 0.6). Of course symmetric attacks are only important to resolve when they preclude definitive evaluation of the status of the root argument proposing the main action. For example, in figure 13c. determining the direction of an asymmetric attack (based on a preference) between $D1$ and $D2$ is not relevant, as irrespective of such a determination, $D2$ is defeated by argument $D3$.

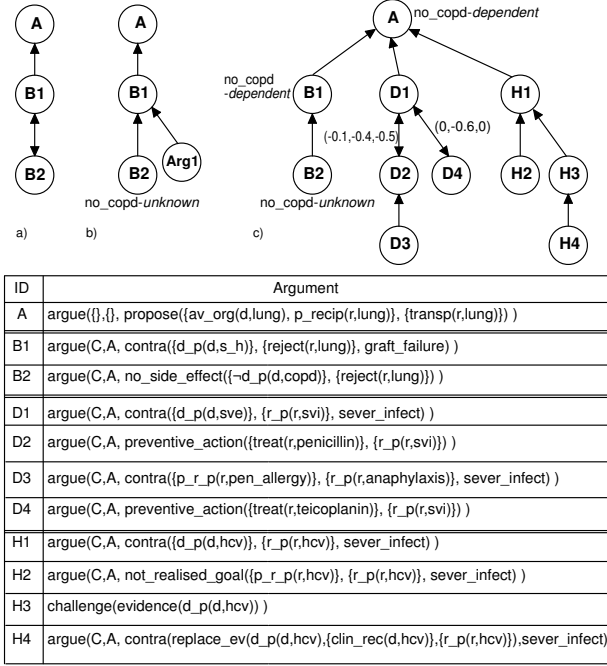| ID	Argument
A	argue({},{}, propose({av_org(d,lung), p_recip(r,lung)}, {transp(r,lung)}))
B1	argue(C,A, contra({d_p(d,s_h)}, {reject(r,lung)}, graft_failure))
B2	argue(C,A, no_side_effect({¬d_p(d,copd)}, {reject(r,lung)}))
D1	argue(C,A, contra({d_p(d,sve)}, {r_p(r,svi)}, sever_infect))
D2	argue(C,A, preventive_action({treat(r,penicillin)}, {r_p(r,svi)}))
D3	argue(C,A, contra({p_r_p(r,pen_allergy)}, {r_p(r,anaphylaxis)}, sever_infect))
D4	argue(C,A, preventive_action({treat(r,teicoplanin)}, {r_p(r,svi)}))
H1	argue(C,A, contra({d_p(d,hcv)}, {r_p(r,hcv)}, sever_infect))
H2	argue(C,A, not_realised_goal({p_r_p(r,hcv)}, {r_p(r,hcv)}, sever_infect))
H3	challenge(evidence(d_p(d,hcv)))
H4	argue(C,A, contra(replace_ev(d_p(d,hcv),{clin_rec(d,hcv)},{r_p(r,hcv)}),sever_infect))

Figure 13: Argument tree evaluation

We now address the MA's third evaluative task which involves accounting for hypothetical arguments and challenges that are either weakly replied or even unreplied. As discussed both in §5.2.3 and §5.2.4, the purpose of *ProCLAIM*'s deliberation is not to decide whether or not uncertain or unknown facts are the case, but whether these are relevant for the actions' safety, and if so, what is the risk involved in these facts being or not the case. The risk involved in a fact being or not being the case is highlighted by the arguments when indicating what undesirable side effects may or may not be expected, so let us discus now how *ProCLAIM* identify the uncertain or unknown facts that decision makers should be ware of.

Once the preference assignment process has taken place, where hypothetical arguments and weakly replied challenges are taken as regular elements of $\mathbb{T}$, the following labelling process takes place:

- Arguments whose updated local context of facts contain an *unknown* fact,

f (*i.e.* $f, \neg f \notin \mathbb{C}_F$) are labelled as f-*unknown*;

- Arguments whose updated local context of facts contain an *uncertain* fact f, *i.e.* while $f \in \mathbb{C}_F$, f has been challenged but not well defended. These arguments are labelled f-*uncertain*;

- Arguments and challenges which acceptability status (defeated or justified) depends on arguments labelled either as f-*unknown* or f-*uncertain* are labelled as f-*dependent*.

Let us continue with arguments A, $B1$ and $B2$, supposing argument $B2$ has been deemed preferred to $B1$ and that fact d_p(d,no_copd) is unknown, and so $B2$ is d_p(d,no_copd)-*unknown* and both arguments A and $B1$ are d_p(d,no_copd)-*dependent*. This is because, if d_p(d,no_copd) is taken to be the case, $B1$ becomes defeated and A justified, whereas if it is taken to be false, $B2$ would be overruled, and so $B1$ would be justified and A defeated. Namely, both A and $B1$'s acceptability status depends on d_p(d,no_copd).

Note that in this example, a decision on whether or not to transplant the lung must first address the problem of not knowing whether d_p(d,no_copd) is or is not the case. This is because argument A is d_p(d,no_copd)-*dependent*. In figure 13b argument $B1$ is attacked and defeated by an argument $Arg1$. In this case, whether d_p(d,no_copd) holds or not, argument A will be justified. In general, only when the argument proposing the main action is labelled as f-*dependent* for some fact f, that the uncertainty of the fact f has to be addressed by decision makers, since only in this case that the final decision actually depends on whether or not f is believed to be the case.

Figure 13c. illustrates what could be a proposed solution to a deliberation. Rather than providing a *safe/unsafe* solution MA returns a new version of $\mathbb{T}$ with *1)* possibly new arguments (*e.g.* $D4$), *2)* where mutual attack are resolved into one way attacks when there is enough confidence to do so (*e.g.* $B2$ preferred to $B1$, $H1$ preferred to $H1$) and when there is not enough confidence, decision makers are given the actual values of the preference assignment so they can ultimately decide which arguments to prefer (*e.g.* pref($D1, D2$) = (-0.1,-0.4,-0.5) and pref($D1, D4$) = (0,-0.6,0)). And finally *3)* the main proposed argument is labelled with the facts that need to be resolved in order to decide upon the main action's safety (*e.g.* argument A is d_p(d,no_copd)-*dependent*). In the example depicted in figure 13c the transplant will be deemed safe if decision makers believe d_p(d,no_copd) to be the case and rely on teicoplanin to prevent the recipient's infection. If the timeout locution has not been triggered yet, before a decision is taken, this new version of $\mathbb{T}$ is returned to the PAs who may accept the solution submitting the locution accept(sol_id) or they may continue adding information, *e.g.* endorsing argument $D4$ or submitting a further argument $D5$ indicating that the recipient is also allergic to teicoplanin. In the latter case, PAs will have to inform when they have no additional moves or the timeout is triggered for the MA to compute the new solution. When participants submitted the accept(sol_id) locution or the timeout has

been triggered, the deliberation concludes with the submission of the locution `close_deliberation(solution,sol_id)` with the last given solution.

8   Using the Tree of Arguments as Evidence

Once a deliberation has concluded, the tree of arguments $\mathbb{T}$ contains **all** the facts and actions deemed *relevant* for assessing the main proposed action's safety, from the view point of domain experts, guidelines, regulations and past collected evidence. If the main action is deemed safe and eventually performed, $\mathbb{T}$ can then be updated by the appropriate PAs so as to record the actual outcome of the action's performance. For instance, if the recipient of a lung of a donor with smoking history and no COPD rejects the transplanted organ, the RA updates $\mathbb{T}$ so that $B1$ is preferred to $B2$ (*e.g.* change in Figure 13 the attack relation between $B1$ and $B2$ so that $B1$ asymmetrically attacks $B2$).Note that after this update the arguments in $\mathbb{T}$ are no longer *presumptive* but *explanatory* in nature. They describe the actual outcome of the performed action. And so, the updated $\mathbb{T}$ can be reused as evidence for resolving future similar deliberations, which is the CBRc's role.

In this section we only outline how the circuit of argument schemes defined here facilitate the reuse of the evidence encoded in previous deliberations. We refer the reader to [57] for a more comprehensive description of how the four reasoning cycles – Retrieve, Revise, Reuse and Retain [1]– are implemented.

There are two aspects of the schemes defined here that further facilitate the CBRc task: *1)* the specificity of the schemes in the ASR (as described in §6) and *2)* that *relevant* facts and complementary courses of actions are introduced in a structured fashion, each singled out and introduced step by step. The schemes' specificity allows identifying potentially similar cases with little computational cost. The idea is that cases in which the same specialised schemes (reasoning patterns) were used, may be similar. Thus, by organising the case-base in terms of the argument schemes, a set of broadly similar cases can effectively be retrieved. The latter aspect of the schemes facilitates a more detailed comparison between cases on the basis of the similarity between the cases' introduced *relevant* facts and actions. We illustrate with a simple example from the medical scenario.

Suppose the deliberation consisted only of the arguments A, $D1$ and $D2$, where a `lung` of a donor whose cause of death was *streptococcus viridans endocarditis* (`d_p(d,sve)`) is offered for transplantation, and the donor's `sve` is believed to be a contraindication ($D1$) because the recipient may be infected by this bacteria. Argument $D2$ indicates that the infection can be prevented by administrating *penicillin* to the recipient. Arguments A, $D1$ and $D2$ respectively instantiate schemes $AS1_T$, $AS2_{T\_inf1}$ and $AS5_{T\_inf1}$ (see §6.2) encoding the following reasoning pattern:

  –An organ $\boldsymbol{O}$ was intended for transplantation. The donor had some condition $\boldsymbol{P}$ which would bring about a severe infection in the recipient. Treatment

T for the recipient was proposed to prevent this infection–

Thus by retrieving from the case-base all the deliberations which consisted of these three schemes we obtain cases that are already quite similar to our target case. So now, if we take from these past cases those where the organ `O` is a `lung`, the condition `P` is *similar* to `sve` (*e.g. streptococcus bovis endocarditis*) and where the treatment `T` is *similar* to `penicillin`, we obtain the desired set of cases from which to evaluate the target case on an evidential basis. Thus, while the argument schemes are used as a heuristics for a first, broad case retrieval, the similarity between cases is ultimately derived from a similarity between the facts and actions highlighted as relevant for the decision making. The similarity between facts and between actions can be derived from a distance measure between terms in an ontology. So, for instance, if the distance in a medical ontology between the terms `penicillin` and `teicoplanin` is below a given threshold, it can be derived that treatments with these two antibiotics are *similar*. And thus, if two arguments instantiate the same scheme of the ASR, and the used terms for their instantiation are *similar*, we can then say that these two arguments are similar (see [57] for more detail).[29]

Having retrieved the set of similar cases, represented by argument trees, the CBRc can derive its preference assignment on mutually attacking arguments. The retrieved $\mathbb{T}$s represent cases where the action was already performed, and thus it only contains asymmetric attacks. In our example this results in two types of retrieved argument trees: $\mathbb{T}+$, where the arguments *similar* to $D2$ asymmetrically attacks and so defeat those *similar* to $D1$, *i.e.* the action was successful; and $\mathbb{T}$-, where the arguments *similar* to $D1$ defeat those *similar* to $D2$, *i.e.* the treatment did not prevent the recipient's infection. If the incidence of $\mathbb{T}+$ cases *significantly* outnumber the $\mathbb{T}$- cases then argument $D2$ would be deemed preferred to $D1$, otherwise either argument $D1$ would be deemed preferred to $D2$ or, if there is not enough evidence so as to prefer one argument over the other, their conflict will remain unresolved.

Once the target $\mathbb{T}$ has been accordingly edited by all the *ProCLAIM*'s knowledge resources, if the final evaluation indicates that the action is safe, the target case will be retained in the case-base to be reused as evidence in future similar cases. This is described in more detail in [57].

9   Conclusions, Future and Related Work

In this paper we have presented an argumentation-based model –*ProCLAIM*– for deliberating over safety-critical actions. The model aims to provide a setting for an effective and efficient deliberation: by *1)* facilitating participation and

---

[29]Other works (*e.g* [37]) address issues such as the equivalence between argumentation frameworks (where frameworks are equivalent if they show same results under different semantics). The CBRc uses the $\mathbb{T}$ associated to each case only as a heuristic to identify the similar cases, where their similarity is determined by the similarity in the facts highlighted as relevant. In other words, the retrieved $\mathbb{T}$'s may not be equivalent as argumentation frameworks.

exchange of arguments among heterogeneous agents and *2)* focusing the deliberation on the *relevant* matters to be discussed. Central to the realisation of these key features is the use of Argument Schemes and Critical Questions. In §5 we defined a circuit of schemes and CQs that defines a protocol-based exchange of arguments, specialised for deliberating over safety critical actions. In §6 we have illustrated how this circuit can be further specialised for a particular scenario and how these scenario-specific schemes facilitate the argument construction for both human and artificial agents. In particular we have illustrated how, directed by a mediator agent, and with the use of the dialogue game introduced in §4, participants are guided at each stage of the deliberation on what can be argued about and how. In this way, the deliberation can effectively be modelled as an argumentative process for eliciting knowledge from the participants.

The primary contribution of *ProCLAIM* is that it enables the automation of deliberation dialogues between agents (human or software) over organ transplant decisions (or environmental decisions see [52]), in a manner which is structured and orderly, and which elicits all the information needed to make such decisions jointly and rationally, even when this information is possessed only by some of the participating agents. A secondary contribution is that these dialogues do not require the participants to have specialised knowledge of argumentation theory, because the framework embeds medical domain expertise in a natural way using scenario-specific argumentation schemes.

Over the last years a growing number of proposals appeal to the use of argument schemes for argumentation-based dialogues [42, 38, 47, 21, 5, 10]. These works generally assume the schemes proposed by Walton [60] or that proposed by Atkinson *et al.* [4]. While these schemes are undoubtedly of great value, we believe they are too abstract for many real life applications. The possibility to cover any possible line of reasoning is of course appealing, and may be required in some circumstances (*e.g.* in legal applications [22, 6, 62] or in e-democracy [21, 12]). However, other decision-making application can benefit from narrowing down the lines of reasoning to only what is essential to the problem at hand, thus making a better use of the decision making context. The specialised schemes and CQ not only reduce the computational cost for the reasoners but they also focus the dialogue on what is essential, increasing the chances for a successful deliberation process. To the best of our knowledge we know of no other work that have proposed and explored the added value of scenario-specific schemes and CQs.

One of the main contributions of our work is in showing that the provision of the scenario specific schemes and CQs can facilitate relatively sophisticated deliberations in sensitive domains such as human organ transplantation or industrial wastewater management [52], while reducing the complexity of argument construction to filling in simple templates (as shown in §6.2). While the main focus of our work has been on facilitating the agents' exchange of arguments, another contribution is *ProCLAIM*'s approach to argument validation and evaluation. The former is required to flexibly prevent spurious arguments from disrupting the deliberation. The latter is required to provide decision support as to whether the proposed action is safe or not, and is achieved by

incorporating the relevant facts and actions into a tree of arguments that is evaluated on the basis of guidelines and regulations, expert opinion and past collected evidence. When all knowledge resources are in agreement, and the action's safety does not depend on the uncertainty of any fact, the proposed solution provides the reasons to deem the action safe or unsafe. Otherwise, the proposed solution highlights the relevant issues that must be resolved.

Taken together, the above contributions provide foundations for the *practical* realisation of deliberations involving artificial and human agents[30]. We thus believe *ProCLAIM* helps bridge the gap between theoretical models of argumentation for agent systems (as embodied by works such as [4]) and their practical realisation.

Other contributions of this paper include *ProCLAIM*'s decoupling of the resolution of *what is the case* and the deliberation over the actions' safety. Firstly, this gives priority to the main question:–*Is the action safe in current circumstances?*-so that, for example, questioning the current circumstances (*i.e.* the facts in $\mathbb{C}_F$) is licensed only if this challenges the action's safety (at least in a local context). Secondly it allows one to address, in a relatively simple fashion, problems such as incomplete or uncertain information, at the time of constructing the arguments, when updating the new available information and when evaluating the arguments.Another contribution is the way in which the relevant facts and actions are explicitly singled out in the argument construction. This, together with the Argument Scheme Repository, is of great importance for the CBRc task of reusing past deliberations as evidence.

It is worth recalling at this point that *ProCLAIM* is intended for regulated environments where PAs (human or artificial) are expected to be fully cooperative domain experts. We also assume a shared agreement on the rules and purpose of the deliberation. In particular we assume the proposed action to be desirable in default circumstances and while there may be disagreement on which circumstances the action can safely be performed, there is an agreement on which side effects, if believed to be caused, will be sufficiently undesirable so as to prevent performing the proposed action as deemed unsafe (*e.g.* cancer, graft failure or death of the recipient of an organ). In particular, PAs individual goals and values are not part of the argumentation.[31]

While we believe to have made important progress in the development of the reasoning patterns for deliberating over safety critical actions, we do make a number of assumptions that should be addressed in future works. Most notably are the assumptions that there is no preference between undesirable goals and that complementary proposed actions are compatible. The former disables the possibility to argue in favour of an action that although may cause an undesirable side effect (*e.g.* cancer to the recipient) any alternative will result in a worse outcome (*e.g.* death of the recipient). The later assumption disregards

---

[30] as illustrated by [53]'s implementation of a prototype that uses schemes and CQs similar to those described in this paper to facilitate the deliberation between a human and an artificial agent on the viability of a human organ for transplantation

[31] They may influence which arguments they submit and endorse, but they do not constitute a reason within *ProCLAIM*'s deliberation to deem the action as safe or not.

the possibility of two courses of actions causing an undesirable side effect if jointly performed. We briefly discus these limitations in §5.2.5. In future work we should also add more expressivity to the action dimension $\mathbf{A}$ so as to at least incorporate a notion of order in which actions should be performed.

In *ProCLAIM*'s definition we make an assumption that PAs would not dispute each others' description of the state of affairs. The rationale for this assumptions lies on the intuition that each PA provides information on that she has a privileged access. So, while a DA has a privileged access to the information about the donor, the RA does so of the recipient. Similarly in the environmental scenario, an agent representing an infrastructure would not be disputed over the information she gives about that infrastructure. This assumption then motivates limiting the possibility of arguing strictly about *what is the case* to allowing to request evidence in support of an asserted $\mathtt{f} \in \mathbb{C}_F$. This intuition should further be developed and explicitly integrated into *ProCLAIM*'s definition. Note however, that because we decouple the deliberation into $\mathbb{T}$ and $\mathbb{C}_F$, a conflict resolution procedure to decide what is the case can easily be plugged into *ProCLAIM* which outcome will update the set of facts $\mathbb{C}_F$ which in turn may affect $\mathbb{T}$'s arguments shifting their type from factual, hypothetical or overruled. Nonetheless, we will also investigate generalising $\mathbb{C}_F$ to include, for example, rules as well as facts. More involved techniques for belief revision and contraction will then be required (recall that currently we assume only the negation and retraction of facts in $\mathbb{C}_F$.

Future work will also address reformulating the Case-Based reasoning cycle presented in [57] in order to accommodate the formalisation of the Argument Schemes and Critical Questions in this paper. Another important requirement for the implementation of *ProCLAIM* in realistic situations is to expand the corpus of schemes and CQs for the transplant and the environmental scenario ASRs. Also important is to perform more rigorous evaluation of the scope and limitations of the elaborated schemes. Another requirement for future work is to better formalise the dialogue game presented in §4.2 in order to further facilitate its implementation, in particular we intend to describe its axiomatic semantics, that is defining the *pre* and *post* conditions for each dialogue move.

With regard to related work, there are a number of works (*e.g.* [26], [11]) proposing deliberation, persuasion or negotiation models of argumentation for agent systems[32]. However, to the best of our knowledge, none of these works address the more practical aspects that enable actual implementation of the proposed models in scenarios more elaborated than simple illustrative examples. There are also a number of works applying multi-agent systems to safety critical domains; particularly the medical (see [24]) and the environmental domains (see [15]). The most relevant that we are aware of is [35], in which a specific logic for argumentation is proposed for aiding medical doctors in their decision making in a multi-agent setting. However this work is primarily conceptual and does not address the agents' dialogical interaction or the roles of schemes and criti-

---

[32]See proceedings of the Argumentation in Multi-Agent Systems (ArgMAS) Workshop Series (http://www.mit.edu/∼irahwan/argmas/).

cal questions in guiding argument construction. Other works, such as [45] and [20] are related in the sense that a repository of Argument Schemes and Critical Questions play a central role. The former is intended to assist users in argument diagramming, and the latter is intended to help (human) users construct a wide variety of arguments, improving their ability to protect their interests in (potential) dialogues, especially in the legal domain. Another interesting approach is taken in the Magtalo system[46], in which a repository of fully instantiated arguments is used to help users express their position regarding a subject of public debate. The user can direct a dialogue among different artificial agent which allows them to explore the system's knowledge base following the natural flow of a dialogue. The user may then agree with the different exposed arguments, may select an argument directly from the argument store and as a last resource, type her own arguments in natural language (with no additional support). This interaction is presented as non intrusive mode for eliciting knowledge from users. This claim is based on what is termed the *maieutic function* of dialogue [61], *i.e.* because users are immerse in a dialogue they do not feel that they are being interrogated. In that sense we believe to go beyond this meiautic function by not only placing importance on the underlying structure of arguments (noted in [9] to be of value for this purpose) but by exploiting the context of application so that: *1) PAs* need not be concerned about the argument construction, but only in filling in the blanks of templates presented in their domain expertise jargon; and *2)* the elicited knowledge is readily available for computational use, as opposed to embedded in a free text paragraph.

References

[1] A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Commun.*, 7(1):39–59, 1994.

[2] L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proccedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI 1998)*, pages 1–7, 1998.

[3] Aristotle. *Topics.* Clarendon Press, Oxford, UK, 1928.

[4] K. Atkinson, T. Bench-Capon, and P. McBurney. Computational representation of practical argument. *Synthese*, 152(2):157–206.

[5] K. Atkinson, T. Bench-Capon, and S. Modgil. Argumentation for decision support. In *Database and Expert Systems Applications*, pages 822–831. Springer, 2006.

[6] T. Bench-Capon and H. Prakken. Using argument schemes for hypothetical reasoning in law. *Artificial Intelligence and Law*, pages 1–22, 2010.

[7] T.J.M. Bench-Capon. Persuasion in practical argument using value based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–48, 2003.

[8] T.J.M. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.

[9] J. Bentahar, B. Moulin, and M. Bélanger. A taxonomy of argumentation models used for knowledge representation. *Artificial Intelligence Review*, 33(3):211–259, 2010.

[10] E. Black and K. Atkinson. Dialogues that account for different perspectives in collaborative argumentation. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 867–874. International Foundation for Autonomous Agents and Multiagent Systems, 2009.

[11] E. Black and K. Atkinson. Agreeing what to do. *ArgMAS 2010*, page 1, 2010.

[12] D. Cartwright and K. Atkinson. Using computational argumentation to support e-participation. *Intelligent Systems, IEEE*, 24(5):42–52, 2009.

[13] M. Chalamish and S. Kraus. AutoMed: an automated mediator for bilateral negotiations under time constraints. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, page 248. ACM, 2007.

[14] C.I. Chesnevar, G.R. Simari, and L. Godo. Computing dialectical trees efficiently in possibilistic defeasible logic programming. *Logic Programming and Nonmonotonic Reasoning*, pages 158–171, 2005.

[15] U. Cortés and M. Poch, editors. *Advanced Agent-Based Environmental Management Systems*, Whitestein Series in Software Agent Technologies and Autonomic Computing. Birkhuser Basel book, Springer, 2009.

[16] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[17] P.M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(10-15):642–674, 2007.

[18] BR Gaines, DH Norrie, and AZ Lapsley. Mediator: an intelligent information system supporting the virtual manufacturing enterprise. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, volume 1, pages 964–969. IEEE, 2002.

[19] M. P. Georgeff and A. L. Lansky. Reactive reasoning and planning. In *AAAI*, pages 677–682, 1987.

[20] T. F. Gordon, H. Prakken, and D. Walton. The carneades model of argument and burden of proof. *Artif. Intell.*, 171(10-15):875–896, 2007.

[21] T.F. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-15):875–896, 2007.

[22] T.F. Gordon and D. Walton. Legal reasoning with argumentation schemes. In *12th International Conference on Artificial Intelligence and Law*, pages 137–146. ACM, 2009.

[23] C. L. Hamblin. *Fallacies*. Methuen and Co Ltd, London, UK, 1970.

[24] D. Isern, D. Sánchez, and A. Moreno. Agents applied in health care: A review. *International Journal of Medical Informatics*, 2010.

[25] H.M. Kauffman, M.A. McBride, and F.L. Delmonico. First Report of the United Network for Organ Sharing Transplant Tumor Registry: Donors With A History of Cancer1. *Transplantation*, 70(12):1747, 2000.

[26] E.M. Kok, J.J.C. Meyer, H. Prakken, and G.A.W. Vreeswijk. A Formal Argumentation Framework for Deliberation Dialogues. *ArgMAS 2010*, page 73, 2010.

[27] A. Lopez-Navidad and F. Caballero. Extended criteria for organ acceptance. Strategies for achieving organ safety and for increasing organ pool. *Clinical transplantation*, 17(4):308–324, 2003.

[28] P. Lorenzen and K. Lorenz. *Dialogische logik*. Wissenschaftliche Buchgesellschaft Darmstadt, Germany, 1978.

[29] D. Marelli, H. Laks, S. Bresson, A. Ardehali, J. Bresson, F. Esmailian, M. Plunkett, J. Moriguchi, and J. Kobashigawa. Results after transplantation using donor hearts with preexisting coronary artery disease. *The Journal of Thoracic and Cardiovascular Surgery*, 126(3):821–825, 2003.

[30] M. Mbarki, J. Bentahar, and B. Moulin. Specification and complexity of strategic-based reasoning using argumentation. *Argumentation in multi-agent systems*, pages 142–160, 2007.

[31] P. McBurney, D. Hitchcock, and S. Parsons. The eightfold way of deliberation dialogue. *International Journal of Intelligent Systems*, 22(1):95–132, 2007.

[32] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 13:315–343, 2002.

[33] P. McBurney and S. Parsons. Dialogue games for agent argumentation. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, chapter 13, pages 261–280. Springer, Berlin, Germany, 2009.

[34] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

[35] S. Modgil and J. Fox. A guardian agent approach to safety in medical multi-agent systems. *Safety and Security in Multiagent Systems*, pages 67–79, 2009.

[36] P. S. Munindar. Agent communication languages: Rethinking the principles. *IEEE Computer*, 31(12):40–47, 1998.

[37] E. Oikarinen and S. Woltran. Characterizing Strong Equivalence for Argumentation Frameworks. In *Proceedings of the 12th International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*. AAAI Press, 2010.

[38] E. Oliva, P. McBurney, A. Omicini, and M. Viroli. Argumentation and Artifacts for Negotiation Support. *International Journal of Artificial Intelligence*, 4(S10):90, 2010.

[39] S. Parsons, P. McBurney, E. Sklar, and M. Wooldridge. On the relevance of utterances in formal inter-agent dialogues. In *Proceedings of the 4th international conference on Argumentation in multi-agent systems*, pages 47–62. Springer-Verlag, 2007.

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

[41] H. Prakken. Formal systems for persuasion dialogue. *Knowledge Eng. Review*, 21(2):163–188, 2006.

[42] I. Rahwan, B. Banihashemi, C. Reed, D. Walton, and S. Abdallah. Representing and classifying arguments on the semantic web. *The Knowledge Engineering Review (to appear)*.

[43] I. Rahwan and G.R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 2009.

[44] C. Reed and T. J. Norman, editors. *Argumentation machines: New frontiers in argument and computation*. Kluwer Academic Publishers, 2004.

[45] C. Reed and G. Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(4):983–, 2004.

[46] C. Reed and S. Wells. Dialogical Argument as an Interface to Complex Debates. *IEEE Intelligent Systems*, pages 60–65, 2007.

[47] C. Reed, S. Wells, J. Devereux, and G. Rowe. Aif+: Dialogue in the argument interchange format. In *Proceeding of the 2008 conference on Computational Models of Argument: Proceedings of COMMA 2008*, pages 311–323. IOS Press, 2008.

[48] J.R. Searle. *Rationality in action*. The MIT Press, 2003.

[49] W. Shen, F. Maturana, and D. H. Norrie. MetaMorph II: an agent-based architecture for distributed intelligent design and manufacturing. *Journal of Intelligent Manufacturing*, 11(3):237251., 2000.

[50] S. Simoff, C. Sierra, and R.L. De Mántaras. Requirements towards automated mediation agents. In *Pre-proceedings of the KR2008-workshop on Knowledge Representation for Agents and Multi-Agent Systems, Sydney, September 2008*, page 171. Citeseer, 2008.

[51] P. Tolchinsky, K. Atkinson, P. McBurney, S. Modgil, and U. Cortés. Agents deliberating over action proposals using the *proclaim* model. In *CEEMAS*, pages 32–41, 2007.

[52] P. Tolchinsky, M. Aulines, U. Cortes, and M. Poch. Deliberation Over the Safety of Industrial Wastewater Discharges into Wastewater Treatment Plants. In *Advanced Agent-Based Environmental Management Systems*, Whitestein Series in Software Agent Technologies and Autonomic Computing, chapter 2, pages 37–60. Birkhuser Basel. Springer, 2009.

[53] P. Tolchinsky, U. Cortes, and D. Grecu. Argumentation-Based Agents to Increase Human Organ Availability for Transplant. In *Agent Technology and e-Health*, Whitestein Series in Software Agent Technologies and Autonomic Computing, chapter 3, pages 65–93. Birkhuser Basel. Springer, 2008.

[54] P. Tolchinsky, U. Cortés, S. Modgil, F. Caballero, and A. López-Navidad. Increasing human-organ transplant availability: Argumentation-based agent deliberation. *IEEE Intelligent Systems*, 21(6):30–37, 2006.

[55] P. Tolchinsky, U. Cortés, J. C. Nieves, F. Caballero, and A. López-Navidad. Using arguing agents to increase the human organ pool for transplantation. In *3rd Workshop on Agents Applied in Health Care (IJCAI-05)*, 2005.

[56] P. Tolchinsky, S. Modgil, and U. Cortés. Argument schemes and critical questions for heterogeneous agents to argue over the viability of a human organ. In *AAAI 2006 SS Series; Argumentation for Consumers of Healthcare*, pages 105-111, AAAI Press, 2006.

[57] P. Tolchinsky, S. Modgil, U. Cortés, and M. Sànchez-Marrè. CBR and Argument Schemes for Collaborative Decision Making. In *COMMA*, volume 144 of *Frontiers in Artificial Intelligence and Aplications*, pages 71–82. IOS Press, 2006.

[58] J. Vázquez-Salceda, U. Cortés, J. Padget, A. López-Navidad, and F. Caballero. The organ allocation process: a natural extension of the CARREL Agent-Mediated Electronic Institution. *AiCommunications.*, 3(16), 2003.

[59] B. Verheij. Dialectical argumentation with argumentation schemes: An approach to legal logic. *Artif. Intell. Law*, 11(2-3):167–195, 2003.

[60] D. N. Walton. *Argument Schemes for Presumptive Reasoning.* Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.

[61] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning.* SUNY Series in Logic and Language. State University of New York Press, Albany, NY, USA, 1995.

[62] A. Wyner and T. Bench-Capon. Argument schemes for legal case-based reasoning. *Legal knowledge and information systems. JURIX*, pages 139–149, 2007.

[63] S. Zink, H. Smolen, J. Catalano, V. Marwin, and S. Wertlieb. NATCO, the organization for transplant professionals public policy statement. HIV-to-HIV transplantation. *Progress in transplantation (Aliso Viejo, Calif.)*, 15(1):86, 2005.