Open Protocol in Multi-agent Systems*

Gerard A.W. Vreeswijk

University of Limburg, Department of Computer Science (FdAW),
P.O. Box 616, NL-6200 MD Maastricht, The Netherlands,
Email address: vreeswijk@cs.rulimburg.nl
Phone: +31-43 88 2021 (work) +31-1720 40953 (home)

ABSTRACT

Protocols for intelligent agents are difficult to write. This is partially because intelligent agents accommodate their interaction to the situation that arises. As a result, the interaction among intelligent agents cannot be governed by a fixed protocol. In this paper I describe an open protocol paradigm for reasoning in a multi-agent system with decentralized control. An open protocol is a collection of rules of interaction that lies open to further alteration and adjustment. I explore the notion of open protocol, and show how it can be altered by rational claims. This may lead to further insights concerning the government of interaction among intelligent agents in multi-agent systems.

1. Introduction

The trend towards the development of increasingly intelligent multi-agent systems makes new demands on theories of interaction in computer science. Protocols governing these interactions must keep up with these developments. Writing protocols for intelligent and autonomous agents is therefore an important but difficult task.

Agents in multi-agent systems perform different activities. These activities include planning, negotiation, deliberation, and decision making. Each activity requires a special protocol. A great deal of effort has been spent in formulating the right protocols for each of these activities. Multi-agent planning protocols, for example, are formulated in (Durfee and Lesser, 1987), (Georgeff, 1987), and (Lansky, 1991). Protocols for negotiation have been proposed in (Davis and Smith, 1983), and (Conry, Meyer, and Lesser, 1986). Decision making protocols have been suggested by (Tenney and Sandell, 1981), and (Stirling, 1992). Protocol is essential to multi-agent systems because it enables agents to interact in an organized way.

In this paper, I am concerned with (artificial) group reasoning, an activity in which several agents deliberate on the status of their knowledge. Group reasoning is performed not for its own sake, but serves as a basis for decision and action within the constraints of allowable resources. In a typical group reasoning situation, actions are analyzed by communicating claims, and arguments supporting that claims, related to the action considered. The study of

*This document is "Technical Report CS 95-", available from the Department of Computer Science, University of Limburg. Copyright © January 1995 by Gerard A.W. Vreeswijk. Permission to redistribute for academic purposes granted provided that this notice is not removed.

protocol regulating group reasoning is called dialectic (Rescher, 1977), (Loui, 1992), (Vreeswijk, 1993), (Gordon, 1993). Because dialectic involves high-level interaction and communication, the study of protocol springs from dialectic.

2. What is the problem?

A major problem in the design of multi-agent systems is the design of the protocols. Despite the efforts mentioned above, there is not only a lack of rigorous theory of interaction and communication, but also the notion of protocol *itself* seems to be a slippery concept. It is a matter of fact that, within any intelligent distributed system, it is impossible to settle fixed protocol standards once and for all.

Considering the nature of rational dispute and interaction more closely, it is not strange that the notion of protocol is hard to capture within a rigorous definition. Different situations require different ways of conduct. Take group reasoning. Sometimes, the importance of a certain point warrants a profound analysis; at other times, a superficial approach is more to the point. In either way, the definition of an optimal protocol depends on the agents' current situation and can thus be determined only during execution. The problem is: how do we define such a protocol?

An answer to this problem is referring to the existence of a so-called *open protocol*. When group reasoning is conducted by an open protocol, agreement over protocol is not a part of shared basis, but subject to dispute. This means that agents can advance claims about how to define or alter the protocol.

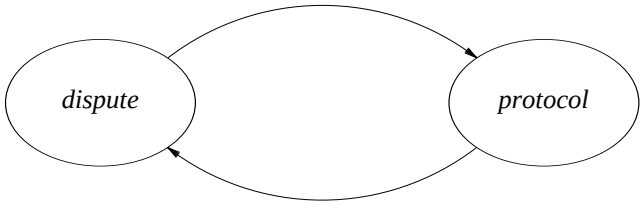


Figure 1. Scheme of interaction

If agents use arguments to exercise influence on the protocol, we get a kind of feedback control system where the conclusions that pertain to protocol are lead back to the process by which they are generated, in order to exercise corrective control (cf. Fig. 1). This situation reminds to the respectable field of *systems theory*, where engineers study physical control systems (like thermostats, electrical power supply, and servo mechanisms) with numerical means to understand what is going on in self-regulating systems (Doyle, Francis, and Tannenbaum, 1992). Social psychologists also study feedback phenomena, but then related to human behavior (Carver and Scheier, 1981). The types of theories developed in these disciplines are completely different. Social psychologists tend to rely on theories expressed in natural language; engineers on real variables; theoretical computer scientists on process calculi. Here I investigate the phenomenon in the context of rational multi-agent systems, acknowledging the effort in other disciplines.

The idea of an open protocol is not new. In fact, a number of researchers already recognized the importance of amendable protocol.

In the field of distributed artificial intelligence (DAI), for example, Carl Hewitt (1991) argues for the adoption of the legal conceptualization of fair, effective and apropos protocol, called *due process*.[1] For Hewitt, due process is the organizational activity of humans and computers for generating sound, relevant, and reliable information as a basis for decision and action within the constraints of allowable resources. Among other things, it should be self-reflective:

> *"Due process is an inherently self-reflective process in that the process by which information is gathered, organized, compared, and presented is subject to evaluation, debate, and evolution within the organization. Thus the debate is not just about whether to lower prices* [i.e. points of contention in the field of discourse], *but also about the beliefs used in the decision and the process used by the organization to decide whether to lower them."*

Hewitt is serious on this, and argues repeatedly for the importance of a self-regulating component.

In the field of computational dialectic, there is a prominent paper of Ron Loui (1992), titled 'Process and Policy', in which he shows the importance of fair, effective, and amendable protocols for non-demonstrative argument:

> *"Locutions pertain not only to the substance of dispute, but also to the protocol. Arguments can advance claims about how to define or alter the protocol. Arguments can seek to establish conditions relevant to protocol, such as defeat relations among arguments, adequacy of responses, and termination. These are rightly considered meta-arguments. The more effective the definitions of aspects of protocol, such as when an argument defeats another, when response is adequate, and when termination is appropriate, the less the need to ascend to meta-argument. Frequently, agreement over protocol is part of shared basis."*

Unlike general approaches to self-regulation and self-government in organizations, Loui is more in line with the goal of this article. I am, like Loui, concerned with self-regulation in the context of rational protocol.

There are more initiatives I consider relevant to this article. This is a game, called NOMIC, where players move by recursively changing the rules of the game. Players start off following some 'initial rule set', which dictates how the rules can be changed. Once a rule change has been made, players then follow this new rule set. The game ends regularly or in paradox. NOMIC is proposed in 1980 by Suber as an abstract form of rule making and legislation. Hofstadter (1982) picked up the idea, and presented it as a self-modifying game to the scientific community. Since then, NOMIC developed into a popular concept, with actual games played on international computer networks. The records of these network games (with a length varying from 2000 to 32.000 lines, and an average rate of about 50 proposals per month) show that the game is somewhat akin to parliament or congress, with all the verbal ingeniousity which that implies. The creator of NOMIC, Peter Suber, originally proposed the game as a study of self-amendment in law. In his recent book, he discusses much of the law of the U.S. federal amending process and convincingly argues that any mature juridical system inevitably involves self-regulating constructions (Suber, 1990).

Besides Hewitt, Loui, and perhaps Suber, there are other researchers arguing for the importance of open protocol (Miller, 1977), (Sartor, 1993). The problem, however, is that, to my

[1] This is a legal concept. It is interesting to note that other important legal concepts, like *defeasibility* of arguments and decisions, preceded the introduction of due process in artificial intelligence. Apparently, legal science is an important source of inspiration when it comes to the formalization of commonsense reasoning.

knowledge, all of the work on open protocol is pre-formal. Much of the discussion is devoted to choices on design and implementation, but fail to materialize in sharp and formal definitions. In consequence, it is impossible to produce the mathematical results desired.

The purpose of this paper is to work towards mathematical results on open protocol. To achieve this, I study the concept of an open protocol in a formal setting. Informal notions like knowledge, belief, argument, and dispute are formalized, and applied to a formal case study. In this way, I discover how protocol is altered by dispute: not directly but by means of changes in the beliefs maintained by the agent. I show that these beliefs dictate the protocol what to do, and that the protocol directs dispute to new beliefs (cf. Fig. 2).

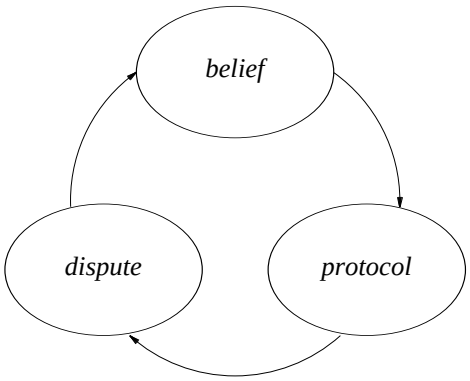


Figure 2. Extended scheme of interaction

Beliefs are thus *working beliefs*, an official term in this paper, a 'current opinion' that is used as a rationale for further actions and decisions, especially those that bear upon protocol and dispute.

In the introduction, it has already been mentioned that each type of multi-agent interactivity requires a special protocol. In addition, I mentioned that writing specialized protocols is a laborious task. The idea of an open protocol, however, is not meant to save us from this task. More to the point, *it is not the case that an open protocol is a universal concept that applies to every interaction situation.*

Rather, the situation is like depicted in Figure 3. For each activity (such as planning, negotiation. deliberation, and decision making), there is a specialized protocol. Normally, these protocols are fixed. So, normally, there is a fixed protocol for planning, a fixed protocol for negotiation, etc. But since we are studying open protocols here, we consider the situation in which every protocol is regulated by belief. And then the situation is like depicted in Fig. 3. When all protocols are supposed to be open, they are supposed to be regulated by an epistemic component, in the picture referred to as *belief*, and the protocol governing dispute is among them. The protocol governing dispute is also the most interesting protocol, because it regulates the production of new beliefs.

Figure 3 should also explain why a study on open protocol necessarily involves a study of (artificial) group reasoning. In this paper, I am concerned with (artificial) group reasoning because motives to change open protocol are always dictated by current opinion (belief) which,

on its turn, is often (but not always) formed on rational grounds. Group reasoning is a means to feed the effects of protocol, caused by opinion, back to opinion. This explains why open protocol is bound to involve (artificial) group reasoning, and why I confine my efforts in this paper to the dashed box.

Before unfolding my ideas on open protocol, I need to make a final disclaimer. Besides the fact that the idea of an open protocol does not coincide with the idea of a protocol 'scheme,' the concept *itself* is sometimes uncalled-for. Sometimes, it is not necessary to have amendable patterns of interaction. In a distributed problem solving environment (DPS), for example, communication is regulated by a fixed protocol, because knowledge is shared, and the field of discourse and the problem addressed are completely determined (Davis and Smith, 1983). In such situations, amendements are unnecessary and undesirable. The goal, as well as the path that might lead to its fulfillment, are clear right from the start. Another typical example is the multistage negotiation paradigm proposed by Conry, Meyer, and Lesser (1986). This protocol is a subtle but fixed algorithm to let agents negotiate their plans. Also here, the path that might lead to the goal desired is clear from the outset. Thus, for some systems and/or for some problems, open protocol is simply irrelevant. For many other systems, however, it is a necessary concept, suiting patterns of communication to the spur of the moment.

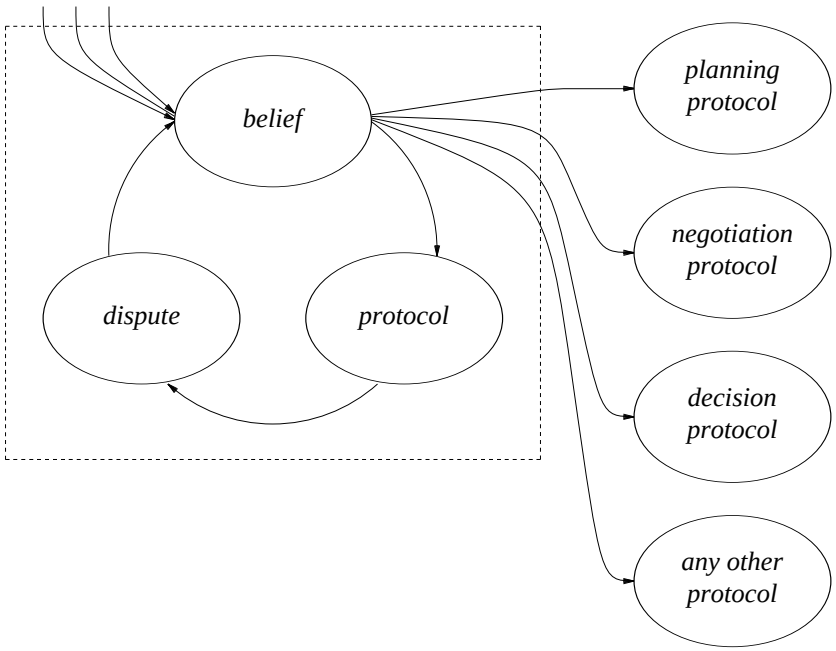


Figure 3. Context

The logical preliminaries are presented in Sect. 3. Section 4 describes a formal model of a multi-agent epistemic environment, where agents have knowledge and belief, and communicate according to a basic protocol. Section 5 describes the workings of open protocol. Section 6 deals with protocol performance, and how performance is related to complete rationality.

Some conclusions are presented in Section 7.

3. Logical preliminaries

As I have argued, an appropriate exploration of multi-agent open protocol must take into account the logical machinery to build claims. Together with the hypothesis of a multi-agent environment, I construct a simple model of (artificial) group reasoning. First, I will create a device for constructing the logical background. This will be done along the lines of (Loui, 1992), (Vreeswijk, 1993), and (Pollock, 1994).

3.1. DEFINITION. A *logical framework* $\mathcal{F}$ is a triple $\langle \mathcal{L}, R, m \rangle$, where the first component $\mathcal{L}$, the *language*, is a set of literals, i.e. a set of atomic language elements, together with their negation; the second component, R, is a set of *open rules*; the third component, m, is an ordered set of *modalities*, to create rules of different strength.

AD 3.1. Literals are letters of the alphabet (lowercase), possibly preceded by a negation symbol. Thus, a, $\neg b$, $\neg p$, and z are four literals. Examples of languages are $\mathcal{L}_1 = \{a, \neg a\}$, $\mathcal{L}_2 = \{a, \neg a, \ldots, z, \neg z\}$, and $\mathcal{L}_3 = \{\neg r, s, \neg c, \neg f, x, \neg e\}$.

A rule represents an elementary and indivisible step of reasoning. My formal notion of argument is constituted of rules. Here are some examples:

$$a,b,c \text{—}[\mathit{strong}]\!\!\rightarrow \neg d$$
$$p \text{—}[\mathit{weak}]\!\!\rightarrow q$$
$$d,f,g \text{—}[\mathit{conclusive}]\!\!\rightarrow \neg k$$

The first rule says that believing a, b, and c is a strong reason for believing $\neg d$. Similarly, believing p is a weak reason for believing q, and believing d, f, and g is a conclusive reason for believing $\neg k$.

If we remove the words in the middle we get *open rules*:

$$a,b,c \text{—}[\ ]\!\!\rightarrow \neg d$$
$$p \text{—}[\ ]\!\!\rightarrow q$$
$$d,f,g \text{—}[\ ]\!\!\rightarrow \neg k$$

An open rule is a rule in which the word in the middle has yet to be filled in.

The words in the middle, bracketed by an arrow, are called *modalities*. A modality denotes the strength of a rule. A possible set of ordered modalities is, for instance,

$$m = \{\mathit{very\ weak, weak, fair, strong, conclusive}\}$$

A set of ordered modalities is not intended to represent the full spectrum of logical forces. This is not my ambition. Rather, a set of ordered modalities is intended as a simple theoretical device to differentiate among rules in the case studies lying ahead. For our discussion on open protocol, for example, a set m consisting of no more than two or three modalities suffices. I am aware of the fact that, in reality, the spectrum of modalities is much broader than is assumed here.

3.2. Constructon of arguments

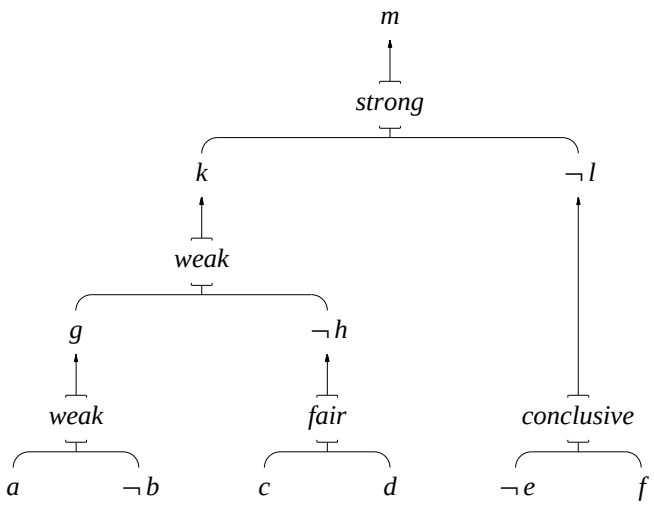
Protocol changes can be established by authority, by negotiation, or by other non-rational means. In the present context, however, we are interested only in rational amendments. Thus, if an agent wants to change the protocol, he must find reason to do so, which means that he must find arguments that justify the amendment proposed. Other types of amendments are left out of consideration here, but it is imaginable how non-rational amendments can be

incorporated in a more ambitious framework. For the present, it is assumed that protocol changes can be established by arguments only.

An agent gets arguments by chaining rules together into trees. Consider, for example, the following rules:

$a, \neg b$ —[*weak*]→ g
c, d —[*fair*]→ $\neg h$
$\neg e, f$ —[*conclusive*]→ $\neg l$
$g, \neg h$ —[*weak*]→ k
$k, \neg l$ —[*strong*]→ m

These rules assemble into the following argument for *m*:



for considerations of space, however, arguments are usually written like this[2]

m ←[*strong*]—
k ←[*weak*]—
g ←[*weak*]— $a, \neg b$
$\neg h$ ←[*fair*]— c, d
$\neg l$ ←[*conclusive*]— $\neg e, f$

The root *m* is the *conclusion* of the argument. The *premises* of an argument are the leaves of the tree, viz. a, $\neg b$, c, d, $\neg e$, and f. If all premises are facts, then the argument in question is *constructible*. Else, it is just an argument. We assume that the *strength* of an argument is equal to the strength of its weakest rule. According to this definition, the present argument for *m* is weak.[3] There may be other arguments for *m*. It is not self-evident, however, that *m* is always supported by the strongest argument. Strong arguments might be defeated by arguments that

[2] In effect, this ascii-representation frees us from writing advanced graphical user interfaces.
[3] We do not claim that this is the only way of determining the strength of arguments. Rather, it is one out of many definitions that works fine for now.

are even stronger. (Cf. Section 6.1.)

Continuing our definition of arguments, we see that the argument for m has two immediate subarguments, viz.

$$k \leftarrow [weak] \text{—}$$
$$g \leftarrow [weak] \text{—}\ a, \neg b$$
$$\neg h \leftarrow [fair] \text{—}\ c, d$$

and

$$\neg l \leftarrow [conclusive] \text{—}\ \neg e, f$$

of which the last is a so-called strict argument. An argument is *strict* if it is constructed from conclusive reasons only. Otherwise, an argument is called *defeasible*.

Thus, the argument for m has two immediate subarguments, of which the one for k is defeasible and the one for $\neg l$ is strict. The premises of the defeasible argument are a, $\neg b$, c, d; the strict argument is based on the premises $\neg e$, $\neg f$.

3.3. Claims

Arguments are used to advance claims. A *claim* is a modalized literal, i.e. a literal to which a certain modality is attached. Thus, $a[strong]$ is a claim, representing strong belief in a. Similarly, the expressions $\neg b[weak]$, $\neg p[weak]$, and $z[very\ weak]$ are three claims. There are two kinds of claims: facts and proper claims.

1. *Facts.* A fact is an improper claim. It claims knowledge without further justification. Accordingly, the modality of a fact is not related to arguments. Facts possess an absolute modal status, indicated by $[fact]$. The expressions $a[fact]$ and $\neg b[fact]$, for example, represent two facts. Facts have their own modality $[fact]$, but equivalent and more general modalities like $[absolutely\ sure]$, or $[provable]$, can be used also.
2. *Proper claims.* Proper claims receive their modal status by means of an argument. In this way, every argument may exercise influence on the modal status of a claim. A weak argument for m, for example, is a potential supporter of the claim $m[weak]$. A *potential* supporter, because other arguments might oppose the status argued.

Sometimes, claims pertain to the organization of dispute. Such claims are called *procedural claims*. For example, if t means that 'search must be terminated as soon as possible', then $t[strong]$ is a strong procedural claim. For future reference, let us denote the collection of all procedural claims by P.

Procedural claims are intended to partially define the organization of dispute. In the latter case, such a claim is called an *amendment*. An amendment is a procedural claim, intended to change the way in which the dispute is conducted. In addition, let us call an amendement *successful* if dispute terminates in favor of the amendement. Thus, a successful amendment is a proposal to alter the order of dispute, defended successfully *in* dispute. This makes amendements into a very special type of claim.

4. Multi-agent epistemic environment

A multi-agent epistemic environment is a collection of agents against the backdrop of a logical framework. For each agent, the epistemic component consists of knowledge and belief. Furthermore, each agent is supposed to run its own protocol.

4.1. DEFINITION. A *multi-agent epistemic environment* $\mathcal{E}$ is a pair $\langle \mathcal{F}, (A_i)_{i \in I} \rangle$, where the last component represents a collection of *agents*. An *agent* A is a tuple $\langle K, B, i, \boldsymbol{P} \rangle$, where K is a

knowledge base, B a set of working beliefs, i an incoming message queue, and $\boldsymbol{P}$ a protocol. A *knowledge base* consists of facts and rules. A set of *working beliefs* B is a finite set of claims and rules, such that $K \subseteq B$ and $B - K$ consists of claims only. An *incoming message queue* is a first in first out storage medium of incoming information. A *protocol* is a function mapping triples $\langle K, B, i \rangle$ onto informal action descriptions.

AD 4.1. Examples of informal action descriptions are: 'agent i sends message m to agent k,' 'agent j adopts $q\,[weak]$ as a working belief,' and 'agent k stops waiting for a response to query nr. 5'.

AD 4.1. The present definition of a multi-agent epistemic environment is not intended as a full and comprehensive model of multi-agent interaction. Instead, I want to formalize what I think is necessary for the theory, leaving other aspects like plans, goals, intentions, commitments, etc., temporarily out of consideration.

Definition 4.1 gives us the global picture of a multi-agent epistemic environment, where agents with an explicit epistemic component act according to a certain protocol. Sect. 4.2 goes further into the details of knowledge and belief (the epistemic component), Section 4.3 elaborates communication and interaction, and Section 4.4 specifies a basic protocol.

4.2. Knowledge and belief

Within AI, knowledge and belief are loaded concepts. It is generally agreed, however, that knowledge and belief are a kind of epistemic commitments that bear a degree of 'adherence', ranging on a scale from 'absolute adherence' if something is believed to be known, to 'provisional adherence' if something is just believed. This idea is followed here as well.

Within the present approach, K is considered to be fixed and unchangeable. It is the hard core of B, the working beliefs. Here are some examples of knowledge bases:

$$
\begin{aligned}
K_1 &= \{d\,[fact]\} \\
K_2 &= \{a\,[fact],\ \neg b\,[fact],\ c\,[fact];\ a,c \text{—}[weak]\!\mapsto d;\ a,\ \neg b \text{—}[strong]\!\mapsto\ \neg d\} \\
K_3 &= K_1 \cup K_2 \\
K_4 &= \{\neg x[fact];\ \neg x \text{—}[weak]\!\mapsto y;\ y \text{—}[absolute]\!\mapsto x\}
\end{aligned}
$$

I do not consider situations in which knowledge bases assimilate new facts of a changing environment, that is, I do not wish to consider updates or revisions. Not because these operations are undesirable or impossible, but because a moving epistemic ground makes the situation more complicated. For now, fixed knowledge suffices to get the ideas across.

The set of working beliefs B contains claims of which the agent 'believes' they are true, but may be disposed of or replaced in changing circumstances. So beliefs, as opposed to knowledge, *are* subject to change. An agent maintains working beliefs as provisional truths, on the basis of which to decide and to act if necessary. Examples:

$$
\begin{aligned}
B_1 &= K_1 \\
B_2 &= K_2 \cup \{\neg d\,[strong]\} \\
B_3 &= K_3 \cup \{d\,[fact]\} \\
B_4 &= K_4 \cup \{z[strong]\}
\end{aligned}
$$

Different 'extensions' are possible. Thus proper beliefs are in no way determined by the contents of the knowledge base. The first three 'extensions' merely suggest plausible belief. The last extension, B_4, is a formally correct but intuitively implausible set of beliefs. Implausible, because there is a strong claim, $z\,[strong]$, that seems to be incorporated arbitrarily into the set of working beliefs.

The beliefs of agent A_4, for example, can be displayed as follows:

Beliefs of A_4	
knowledge	$\neg x$ —[*weak*]↦ y y —[*absolute*]↦ x $\neg x$ [*fact*]
proper belief	z [*strong*]

The table says that agent considers three items as known, and maintains one item as a belief.

A knowledge base consists of different sorts of knowledge. For the present purposes, it is important to distinguish between *domain knowledge* and *procedural knowledge*. Domain knowledge consists of facts and rules about the field of discourse. Finance, for example, or financial administration. Procedural knowledge is concerned with the protocol. It consists of facts regarding the context of disputation, and rules regarding order, procedure, and amendement.

Similarly, the set of working beliefs is divided into beliefs about the field of discourse and beliefs about how to define or alter the protocol. Consider the following examples in the field of credit decisions and loan assessment:

- Domain knowledge: 'client ID-23 has a credit of 34,-', 'client ID-29 has a debt of 260,-'
- Procedural knowledge: 'if loaning history of a particular client is not known, try to consult the nearmost site'
- Belief: 'there is strong evidence to assume that client ID-29 is able to redeem within due time'
- Belief concerning procedure: 'for the time being, client capacity checking and loan assessment is board out to site S-12'

Finally, there is the theoretical concept of *collective knowledge*. Collective knowledge in an epistemic environment $\mathcal{E}$ is defined as $K = \cup \{K_i \mid i \in I\}$. Collective knowledge (not to be confused with common knowledge) is a simple theoretical concept, meant to be able to refer to the epistemic basis of the group as a whole.

4.3. Communication

In a multi-agent epistemic environment, communication is a special type of interaction, meant to pass knowledge. Exchange of knowledge is realised by means of sending (and receiving) messages. In the present context, it suffices to distinguish between two types of messages: queries and responses.

1. *Queries.* If agent *i* seeks support for *p*, it issues the query [i,j] $+p$?, where *j* is a query identification number. If agent *i* searches for information against *p*, it issues the query [i,j] $-p$?

   Depending on the protocol, an agent may *broadcast* a query to all other agents, or to a selected group of consultants. It may also be ordered to handle the query itself.

2. *Responses.* A response is either a positive or a negative answer to a query. If it is a positive answer, a modality is attached, denoting the 'grade of confirmation'. A negative answer is not accompanied by a modality.

Examples of positive responses are *'yes [weak]'*, *'yes [strong]'*, and *'yes [conclusive]'*. There is only one type of negative response, namely a plain *'no'*.

If an agent issues a query, there may be immediate response, delayed response, or there may be no response at all. Furthermore, the response may be positive or negative. This leads to a number of different situations. If there is an immediate positive response, for example, the agent who issued the query may decide to wait for a stronger confirmation in order to consolidate his claim. In other situations, an agent may decide to content himself with the modality of the first positive response. An agent may also decide to time out after receiving only negative responses to a query. If there is no response at all, this may have two reasons. The first is communication failure. The second reason is that some of the agents consulted are still dealing with the query. These agents are yet unable to return either a qualified 'yes' or a definite 'no'. In either way, it will be advantageous to have agents operating with certain time-out mechanisms. This mechanism is also directed by the current protocol.

4.4. Protocol

In a multi-agent epistemic environment, each agent runs a so-called *local protocol*. A local protocol is a function from local states to local actions (using a fact, applying a rule, preparing a message). Naturally, these actions include possible interactions with other group members. There are as many protocols as there are agents. In the literature, one often speaks of a so-called *distributed protocol*. Our conceptualization of protocol thus fits in with the more general notion of protocols as proposed by (Lynch and Fischer, 1981), (Ladner and Reif, 1986), and (Burmeister, Haddadi, and Sundermeyer, 1993).

Each agent runs an *open* local protocol. An open local protocol consists of a core, the *basic protocol*, supplemented or altered by amendments. The basic protocol is the starting point for interaction, the 'initial communication convention', that may be adapted in due course. How far these adaptations may be pursued is left open in this paper, although I believe that interaction must be framed within certain 'laws'. Delimiting the space in which protocol can be negotiated is certainly important, but must be marked as future research for considerations of space.

A run governed by the basic protocol proceeds along the following lines.

4.4.1. DEFINITION. (Basic protocol.) Each agent examines its incoming message queue for communications from other agents. If the queue is not empty, there is a top message. The top message is either a query (from another agent) or a response (to one of its own queries). If it is a query, it is a query for support, or a query for opposition.

1. *A query for support in favor of a certain claim.* If the expression $[j,n]+p?$ is on top of the incoming message queue of agent *i*, this means that agent *j* sent a message in which he requested for support in favor of *p*, and that *i* was one of the addressees. This message must now be dealt with.

   1.1. The first action for *i* is to see if p [*fact*] is in the knowledge base. If so, agent *i* sends the message '$[j,n]$: *yes* [*fact*]' to agent *j*.

   1.2. When *p* is not registered as a fact, agent *i* examines its knowledge base to see if it has a rule of the form x,y,z —[*whatever*]→ *p*. Suppose that is the case. Then $[j,n]+p?$ induces 4 subqueries a priori, of which

   $$[i,m_1]+x?,\ [i,m_2]+y?, \text{ and } [i,m_3]+z?$$

   are issued immediately to see whether x,y,z —[*whatever*]→ *p* is grounded in true evidence. The query $[i,m_4]-p?$ is generated only if the three positive queries are answered affirmatively.

1.3. Suppose that the three positive queries are answered affirmatively. Let us assume that the answers to the three subqueries are

$[i,m_1]$: *yes* [*conclusive*],
$[i,m_2]$: *yes* [*fact*], and
$[i,m_3]$: *yes* [*weak*],

respectively. It follows that the antecedent of x,y,z —[*whatever*]→ p is weakly supported. If the present rule is weaker than *'weak'*, then p is supported with the modality of the rule. Else, the literal p remains weakly supported. If we assume the latter, the query $[i,m_4]-p$ [*weak*]? is issued to investigate if there might be information to the contrary, stronger than *weak*, defeating the support furnished by x [*weak*], y [*fact*], and z [*weak*], and $x,y,z,$ —[*whatever*]→ p.

1.4. If $[i,m_1]+x$?, $[i,m_2]+y$?, and $[i,m_3]+z$? are answered positively, while $[i,m_4]-p$ [*weak*]? remains unanswered, then a message is send to agent j, with the content that query $[j,n]$ is answered in the affirmative: '$[j,n]$: *yes* [*weak*]'.

In all other cases, agent i returns '$[j,n]$: *no*'.

2. *A query for information against.* If $[j,n]-p$? is on top, this means that agent j has send out a request for information against p. Thus, agent i has to find information against p. Because the the background language is equipped with a negation operator, this query is easily dealt with by converting it into $[i,m]+(\neg p)$? The query thus obtained is directly forwarded to other consultants.

The incoming message queue is examined again and activity proceeds as described above (from step 1). Consultancy proceeds until an agent is finished with the examination of all relevant literals.

AD 4.4.1. It is important to note that the basic protocol is underspecified. It is unclear with respect to, e.g., selection of consultants, distribution of search, maximal search depth, local time-outs, and termination.

1. *Selection of consultants.* Consider step 1.2 of the basic protocol. That step involves issuing queries to other agents. The set of agents to which these queries are send to (the addressees) is unspecified.
2. *Distribution of search.* Which type of search is performed by which agent? An agent can for example be ordered to build claims on local evidence only, while opposition must be board out. There are various possibilities here.
3. *Maximal search depth.* Defeaters provoke queries requesting for opposition. From the theory of complete rationality (cf. Section 6.1), it follows that provisionally defeated arguments are able to provoke an infinite number of such queries. The depth in which opposition is iterated cannot be overseen by an individual agent. Therefore, agents have to put limits on iterated consultancy.
4. *Local time-outs.* Consider step 1.4 of the basic protocol. Three queries for support are confirmed, while one query for opposition timed out. The question here is: how long did it take to receive all three confirmations? More generally: how long is agent willing to wait for a response?
5. *Termination.* Questions regarding to local time-out limits, also hold for termination.

A specification of these aspects help to appoint the basic protocol.

AD 4.4.1. The treatment of negative queries is relatively simple here. With more a more sophisticated mechanism of defeat, however, these simple conversions are no longer possible

(Vreeswijk, 1993). A distinction between positive and negative queries is therefore still useful.

AD 4.4.1. Care should be taken that circular consultancy is detected and properly dealt with. This can be done by writing a circularity check. We do it here by hand, with the knowledge that an implementation of such a check is not the real problem here.

5. Open protocol

We now have arrived at a point were it is possible to present a precise definition of open protocol. This will be done with the help of three (formerly defined) concepts, viz. basic protocol, procedural claim, and amendment.

First, we note that the basic protocol is incomplete, and leaves room for supplementary definition. Second, we recall that a procedural claim is an element of the working beliefs, intended to define or alter the protocol. Thus, combining these two observations, we obtain an open protocol if the basic protocol specification is completed with procedural claims. Procedural claims function as 'tentative directives' of the underspecified basic protocol. Stated in precise terms, *an open protocol is a completely specified mapping from local states to informal action descriptions, lying open to further alteration and adjustment.*

In the following sections I further explain the workings of open protocol. Sect. 5.1 deals with the defining parameters of open protocol, the *order*, while Section 5.2 shows how these parameters can be altered by applying the rules of procedure.

5.1. Order

Let B_i be equal to the working beliefs of agent i. Part of B_i consists of procedural claims, intended to specify the protocol of agent i. With the help of the set of all procedural claims P, this subset can be denoted by $B_i \cap P$. The *order of procedure* (order, for short) held by agent i is equal to this subset.[4] The local protocol $\boldsymbol{P}_i$ always operates in accordance with $B_i \cap P$. Suppose for example that

$$B_i \cap P \;=\; \{\neg\, c[\mathit{strong}]\,,\, s[\mathit{weak}]\,,\, t[\mathit{weak}]\}$$

where

$\neg\, c[\mathit{strong}]$	represents a strong claim for limited consultancy,
$s[\mathit{weak}]$	represents a weak claim for extensive search, and
$t[\mathit{weak}]$	represents a weak claim for driving at a quick settlement

(The interpretation of the claims $\neg\, c\,[\mathit{strong}]$, $s\,[\mathit{weak}]$, and $t\,[\mathit{weak}]$ are informal and suggest alterations that are supposed to be intuitively obvious.) Let us consider the set of procedural working beliefs $B_i \cap P$. As long as $B_i \cap P$ does not change, $\boldsymbol{P}_i$'s operation does not change either. Suppose, however, that $\neg\, t[\mathit{strong}]$ is a successful amendment. Then $\neg\, t[\mathit{strong}]$ is adopted as a working belief and $B_i \cap P$ changes into

$$B_i \cap P \;=\; \{\neg\, c[\mathit{strong}]\,,\, s[\mathit{weak}]\,,\, \neg\, t[\mathit{strong}]\}$$

causing $\boldsymbol{P}_i$ to operate accordingly. In a next situation, perhaps, the claim $t\,[\mathit{very\ strong}]$ might successfully re-amend the order of procedure $B_i \cap P$, again causing changes in the operation of $\boldsymbol{P}_i$. In this way, the local protocol is determined continuously by the order of procedure.

Since $B_i \cap P$ is the sole determinant of $\boldsymbol{P}_i$, the procedural claims may be identified with the local protocol. Viewed in this way, agent i's protocol $\boldsymbol{P}_i$ is a program that meets the

[4] A bit like having a personal copy of Robert's rules of order in parliamentary procedure (1971).

specification $B_i \cap P$.

5.2. Rules of procedure

Development of protocol is embodied by the *rules of procedure*. A rule of procedure is a piece of knowledge suggesting amendements to the customary mode of procedure.

The following rules suggest a collection of rules of procedure:

l —[*compulsive*]↦ $\neg c$	If communications channels are overloaded, this is a compulsive reason to reduce consultation activities;
$\neg t, l$ —[*conclusive*]↦ s	If a prolonged analysis is desired but communication channels are overloaded, there is no other way than to perform extensive local search;
r —[*strong*]↦ $\neg s$	Scarcity of resources is a strong reason to curtail search;
t —[*strong*]↦ $\neg u$	Quick settlement is a strong reason for shortening time-out periods;
t, c —[*compulsive*]↦ $\neg u$	Quick settlement in the presence of a large number of consultants is a compulsive reason for shortening time-out periods;
t —[*strong*]↦ $\neg c$	Quick settlement is a strong reason for limited consultancy;
$\neg t, r$ —[*very strong*]↦ u	A deep analysis with few resources is a very strong reason to wait for the results of other agents.

Where
l means *overloaded communication channels; c* means *generous consultancy, in which there is no limit on the number of consultants; $\neg c$* means *limited consultancy; r* means *computational resources are scarce; s* means *extensive search; $\neg s$* means *limited search; t* means *driving at a quick settlement; $\neg t$* means *a prolonged analysis; u* means *lengthening time-out periods.*

In the present approach, a rule of procedure is considered to be knowledge. It is an improper working belief that it is never subject to dispute. In a more realistic situation, however, rules of procedure may be advanced as claims. The next argument, for example, supports the rule l —[*weak*]↦ $\neg c$.

$$(l \text{ —[}weak\text{]}\mapsto \neg c) \leftarrow [strong]\text{—} \begin{array}{l} \neg b \leftarrow [strong]\text{—} \; a \\ \neg e \leftarrow [conclusive]\text{—} \; \neg c, d \end{array}$$

The supporting arguments consist of three rules, of which the top rule is equal to

$$\neg b, \neg e \text{ —[}strong\text{]}\mapsto (l \text{ —[}weak\text{]}\mapsto \neg c)$$

Since an argument is as strong as its weakest rule, this argument builds a strong claim for l —[*weak*]↦ $\neg c$. We thus obtain the procedural claim $(l$ —[*weak*]↦ $\neg c)$ [*strong*]. This (strong) procedural claim suggests a (weak) rule of procedure, to be used in regulating dispute.

Introducing rules as claims is one step further in the theory of open protocol. My current approach does not go that far. [But cf. (Hage, 1993) for efforts in these directions.]

5.3. An example

At this point, the workings of open protocol can best be illustrated with the help of an example. In this example, the open protocol will be altered along the dimension of distributed search. Other dimensions (like selection of consultants, maximal search depth, and termination) are supposed to be fixed.

Distributed search (almost naturally) involves the concept of a distributed argument. A *distributed argument* is an argument of which at least two rules are distributed among different agents. Thus if

$$a_1 \text{—}[weak]\!\!\rightarrow b_1 \in K_1,$$
$$a_2 \text{—}[weak]\!\!\rightarrow b_2 \in K_1, \text{ and}$$
$$b_1, b_2 \text{—}[weak]\!\!\rightarrow c \in K_2,$$

then

$$c \leftarrow\!\![weak]\text{—}$$
$$b_1 \leftarrow\!\![weak]\text{—}\, a_1$$
$$b_2 \leftarrow\!\![weak]\text{—}\, a_2$$

is a distributed argument. If an argument is distributed among a collection of agents $I' \subseteq I$, then all members of I' are involved in a possible construction of this argument.

There are different ways to distribute search, leading to different outcomes. In the present article, I distinguish three different possibilities.

1. *The independent reasoner protocol.* This is the protocol in which agents deliberate individually on the tenability of a claim.

   Remark (i): with this protocol, agents are forced to engage in a so-called *unilateral dispute*. With unilateral dispute, group members are 'self-critical thinkers' producing their own counterarguments.

   Remark (ii): since independent reasoners do not communicate, they usually come to different conclusions. This may put to use in a decision procedure, where independent reasoners cast their vote in accordance with their individual findings. Decision procedures, however, are 'extra-rational' and thus lie beyond the scope of the topic addressed in this paper.

2. *The disputer protocol; the communicated argument protocol.* With this protocol, an agent tries to support relevant claims independently, without communication. After an argument has been constructed, it is communicated to (a possibly selected collection of) consultants. The task of these consultants is to probe the argument submitted.

   Remark (i): once an agent has constructed an argument in favor of a certain claim, this argument is open to criticism of other members of the group.

   Remark (ii): with this type of protocol, less knowledge need to be shared because opposition is board out.

   Remark (iii): a deficiency of the disputer protocol is that a distributed argument (of which the reasons are spread throughout the group) cannot be constructed, and cannot have influence on the final outcome.

   Remark (iv): also in this setting it makes sense to vote because different conclusions may still arise.

3. *The communicated reason protocol.* With this protocol, arguments are constructed groupwise. Arguments that are implicit in the group play a role in the reasoning process and may put a stamp on the final outcome. Support as well as opposition is constructed in cooperation.

Remark (i): there is no need for shared knowledge, because all relevant knowledge can, in principle, be communicated. For efficiency purposes, however, it is desirable to have most of the knowledge shared. In that way, less knowledge needs to be communicated.

The independent reasoner protocol and the communicated reason protocol exemplify two entirely different modes of reasoning. With the communicated reason protocol, it is possible to work towards a thoroughly underpinned conclusion which, from a logical point of view, is favorable. From a computational point of view, however, the communicated reason protocol is *un*favorable. It involves a large communication overhead and, therefore, a severe claim on valuable resources. On the other side of the spectrum lies the independent reasoner protocol. Independent reasoners do not communicate on a rational level, and therefore arrive at locally underpinned conclusions only. This makes the independent reasoner protocol computationally attractive, but logically inferior to the communicated reason protocol.

There is no favored state of reasoning. Therefore, an open protocol must be able to move between different modes of distributed search. If, for example, in a state of 'isolated reasoning' there are strong arguments for a more communicative mode of reasoning, the open protocol should allow these arguments, thus undermining the state of isolated reasoning, and thus moving itself into a more communicative mode of reasoning. An open protocol is thus able to direct itself into the appropriate state of reasoning if there are valid arguments for doing so.

The goal of the example below is to demonstrate precisely these self-directive properties of open protocol. Let me start with an informal description of the situation we want to formalize:

> There is a protocol alternating between two states. The first state allows the construction of distributed arguments. This is not allowed in the second state. But when the protocol is in the second state, the dispute is carried to an argument supporting the conclusion that distributed arguments *are* allowed. The protocol changes accordingly, so that it is back in the first state. In this state, however, the dispute is lead back to the conclusion that distributed arguments are *not* allowed. This brings the protocol again in the second state, and the circle is closed.

This situation of negative feed-back can be formalized as follows. First, we introduce the following point of order:

d = 'search for support may be board out'

If search for support may be board out, there is the possibility for distributed reasons to assemble into one argument. Further, we introduce three agents A_1, A_2, and A_3 with the following beliefs.

For some reason, the procedural claim $d[weak]$ is by all agents adopted as a working belief. Assuming that knowledge is distributed as depicted in Figure 4, an examination of $\neg d$ by A_1 would result in the following run.

A_1: I want to know whether $\neg d$ is a claim warranted by this group.

[1,1]: is there anyone able to construct a valid argument for $\neg d$*?*

A_3: Reacting on [1,1]: I have a rule for $\neg d$, namely $a,b,c \multimap[strong]\rightarrowtail \neg d$. To know whether $\neg d$ is warranted as a conclusion of our group, I'll have to find support for the antecedent first:

[3,1]: is there anyone able to construct a valid argument for a?,
[3,2]: for b?, and
[3,3]: for c?

Beliefs of A_1	
knowledge	a,b,c —[*strong*]↦ $\neg d$ a [*fact*]
proper belief	d [*weak*]

Beliefs of A_2	
knowledge	b —[*weak*]↦ d b [*fact*]
proper belief	d [*weak*]

Beliefs of A_3	
knowledge	c [*fact*]
proper belief	d [*weak*]

Figure 4. Initial configuration

Since I have *c* in my knowledge base K_3, I can answer [3,3] myself: the answer to query [3,3] is *'yes [fact]'*. Search for support for *a* and *b* is board out, since d[*weak*], considered as part of my current protocol, allows me to do so.

A_1: Reacting on [3,1]: I have *a* stored as a fact in my K_1. So my answer to your query [3,1] is *'yes [fact]'*.

A_2: Reacting on [3,2]: I have *b* in my K_1. So my answer to [3,2] is *'yes [fact]'*.

A_3: Just received messages from A_1 and A_2: both *a* and *b* are supported by valid arguments, with the strength of absolute evidence. Since I just draw the analogous conclusion for *c* myself, I conclude with the help of a,b,c —[*strong*]↦ $\neg d$ that there is strong reason to believe $\neg d$.

Having mobilized all support for $\neg d$[*strong*], it's time to search for contradictory information:

[3,4]: is there anyone able to construct a valid argument against $\neg d$*, stronger than 'strong'?*

A_1: Reacting on [3,4]: I have a rule against $\neg d$[*strong*], namely a —[*weak*]↦ d. To know whether *d* is warranted by our group, I'll have to find support for the antecedent:

[1,2]: is there anyone able to construct a valid argument for a?

I can produce the answer to [1,2] myself, since *a* is in K_1: *'yes [fact]'*. With the help of a —[*weak*]↦ d I conclude that I am able to construct a weak argument against $\neg d$[*strong*]. This is not enough, however, since [3,4] asks for an argument stronger than [*strong*] I can't produce such an argument.

Before returning a definite 'no' to [3,4], however, let me think…

A_3: Uptil now, nobody has been responding to my [3,4]. I conclude that there is no information against $\neg d[strong]$ which, on its turn, together with the support for *a*, *b*, and *c*, means that $\neg d[strong]$ is warranted.

To A_1: the answer to [1,1] is *'yes [strong]'*.

A_1: Receiving a message from A_3 in which the existence of a valid argument for $\neg d[strong]$ is indicated. Therefore, I adopt $\neg d[strong]$ as a working belief. From now on, my protocol will operate accordingly.

Eventually, every agent analizing the epistemic credentials of $\neg d$ will adopt the belief $\neg d[strong]$:

Beliefs of A_1	
knowledge	$a,b,c \text{—}[strong]\!\!\rightarrow \neg d$ $a\,[fact]$
proper belief	$\neg d\,[strong]$

Beliefs of A_2	
knowledge	$b \text{—}[weak]\!\!\rightarrow d$ $b\,[fact]$
proper belief	$\neg d\,[strong]$

Beliefs of A_3	
knowledge	$c\,[fact]$
proper belief	$\neg d\,[strong]$

Figure 5. Configuration after transactions under *d*

Now suppose A_1 issues the first query again. This time, with $\neg d[strong]$ held as a working belief, we get the following run:

A_1: I want to know whether $\neg d$ is a claim warranted by this group.

[1,1]: am I able to construct a valid argument for $\neg d$*?*

(Search for [1,1] is kept local this time, because $\neg d[strong]$, considered as part of $B_i \cap P$, does not allow distributed search.) Since $\neg d$ doesn't match one of the consequents of the rules in my K_1, I am unable to support $\neg d$.

Thus, due to a change in the current protocol, search is scheduled differently, so that conclusions that were previously warranted, now lose their epistemic status. Figure 7 shows even more clearly that the same query now initiates a different run.

Does the invalidity of $\neg d$ imply that *d* emerges victorious? Not necessarily so. If an agent is unable to find support for a claim, he isn't supporting the negation of that claim automatically.

In Figure 8 we see what happens if $+d?$ is issued by A_1: A_1's argument for *d* cannot be

Agent A_1	Agent A_2	Agent A_3
[1,1] broadcast $+(\neg d)$?		
	receiving [1,1]	receiving [1,1]; working on [1,1]: a,b,c —[*strong*]↦ $\neg d \in K_3$ leads to [3,1] broadcast $+a$?, [3,2] broadcast $+b$?, [3,3] broadcast $+c$?
receiving [3,1], [3,2], [3,3]; working on [3,1]: $a \in K_1$ leads to send [3,1]: 'yes [fact]'	receiving [3,1], [3,2], [3,3]; working on [3,2]: $b \in K_2$ leads to send [3,2]: 'yes [fact]'	working on [3,3]: $c \in K_3$ leads to conc [3,3]: 'yes [fact]'
		receiving A_1 [3,1]: 'yes [fact]' receiving A_1 [3,2]: 'yes [fact]' knows: $\neg d$ has strong support; this leads to [4] broadcast $-(\neg d)$ [strong]?
receiving [3,4]; working on [3,4]: a —[*weak*]↦ $d \in K_1$ leads to [1,2] broadcast $+a$?	receiving [3,4]	
  working on [1,2]: $a \in K_1$ leads to conc [1,2]: 'yes [fact]' knows: d has weak support; this does not suffice to respond to [3,4] affirmatively	receiving [1,2]	receiving [1,2]
		no response to [3,4]; this leads to send [1,1]: 'yes [strong]'
receiving A_3 [1,1]: 'yes [strong]'		

Figure 6. Summary of transactions provoked by $+(\neg d)$? under d

defeated by joint opposition of A_2 and A_3, because the construction of distributed arguments, in particular the argument a,b,c —[*strong*]↦ $\neg d$, is now forbidden.

Agent A_1	Agent A_2	Agent A_3
[1,1] $+(\neg d)$? can't find support for $\neg d$		

Figure 7. Summary of transactions provoked by $+(\neg d)$? under $\neg d$

Agent A_1	Agent A_2	Agent A_3
[1,1] $+d$? working on [1,1]: a —[*weak*]→ $d \in K_1$ leads to [1,2] $+a$? working on [1,2]: $a \in K_1$ leads to conc [1,2]: 'yes [fact]' knows: d has weak support; this leads to [1,3] broadcast $-d$[weak]?		
	receiving [1,3]	receiving [1,3]; working on [1,3]: a,b,c —[*strong*]→ $\neg d \in K_3$ leads to [3,1] $+a$?, [3,2] $+b$?, [3,3] $+c$? working on [3,3]: $c \in K_3$ leads to conc [3,3]: 'yes [fact]' can't find support for a and b; this makes it impossible to respond to [1,3] affirmatively
no response to [1,3]; this leads to conc [1,1]: 'yes [weak]'		

Figure 8. Summary of transactions provoked by $+d$? under $\neg d$

6. Performance

In the literature of computer networks (Tanenbaum, 1989) and distributed knowledge based consultancy (Mohan, 1984), a (fixed) protocol is considered to perform well if the flow of the messages between agents is mediated in a fair, effective, and economic way, appropriate to the task considered.

When a protocol lies open to amendment, its performance is defined less clear. First, it varies along with the procedural beliefs of the agents. If all agents, for instance, firmly believe in common silence (a state worse than unilateral dispute), each individual agent might bear insufficient rationality to break the deadlock on its own, thus helping to maintain the state of common silence. This is obviously bad performance. On the other hand, there is the prospect of unlimited rationality, where agents believe they have all the time in the world to deal with their queries. This should also be avoided. Furthermore, agents might believe in quick reasoning for the sake of speed, which should be avoided as well. In summary, agents must agree on a protocol that, on the one hand, maintains discipline, but, on the other hand, allows searching for the right conclusions.

The second reason why the notion of performance is yet unclear, is that I do not have objective means to measure performance:

1. *Correctness.* How can we say that an agent has arrived at the right conclusion if we do not have an objective criterion of 'rightness'? Formally: given $\mathcal{E} = \langle \mathcal{F}, (A_i)_{i \in I} \rangle$, where $A_i = \langle K_i, B_i, i, \mathbf{P}_i \rangle$, and a claim p, how can we say that p rationally follows from $K = \cup \{K_i \mid i \in I\}$?
2. *Promptness.* How do we measure the promptness by which an agent finds and returns its conclusions?

In order to be able to discuss open protocol performance, I elaborate on dialectical issues to arrive at a notion of complete rationality (cf. Section 6.1). After performance is precisely defined (Sect. 6.2), I show how it may change along with the protocol. I then show how performance is related to correctness. Finally, it is argued that the quality of performance is determined essentially by the quality of the arsenal of rules on protocol.

6.1. Complete rationality

Dispute must be effective. It must end with conclusions that are right or approximately right. But what is considered to be right? When is it right to say that a conclusion rationally follows from the facts? To answer this question, this section is concerned with defining a notion of complete rationality, namely *warrant*.

Warrant is the product of complete rationality. More to the point, a conclusion is warranted if and only if it follows from the facts by exhaustive argumentation. I will now show how to compute warrant.

Consider a set of facts, rules, and modalities. Given the facts, there is a set of constructible arguments. Some of these arguments, however, are contradicted and defeated by other arguments. In addition it may happen that an argument is defeated by a second argument that is ultimately undefeated. In that case, the first argument is ultimately defeated, and thus unable to support its conclusion. Other arguments, like the second, remain ultimately undefeated. The conclusion of precisely *these* arguments are rationally entailed by the facts.

6.1.1. DEFINITION. (Defeat, counterargument, and interference.) A *counterargument* is an argument of which the conclusion contradicts the conclusion of a subargument of the argument in question; this subargument is said to be the *target* of the counterargument; if a

counterargument is as strong as its target we say that the counterargument *interferes* with the argument in question; if the counterargument is stronger, we say that the counterargument *defeats* the argument in question. A defeating counterargument is called a *defeater*.

AD 6.1.1. Let us as an example consider the argument for *m* which we have studied in Section 3.2:

$$
\begin{array}{l}
m \leftarrow[\mathit{strong}]— \\
\qquad k \leftarrow[\mathit{weak}]— \\
\qquad\qquad\qquad g \leftarrow[\mathit{weak}]— a, \neg b \\
\qquad\qquad\qquad \neg h \leftarrow[\mathit{fair}]— c, d \\
\qquad \neg l \leftarrow[\mathit{conclusive}]— \neg e, f
\end{array}
$$

This argument would be defeated by

$$
\begin{array}{l}
h \leftarrow[\mathit{strong}]— \\
\qquad \neg b \\
\qquad \neg l \leftarrow[\mathit{conclusive}]— \neg e, f
\end{array}
$$

if the latter is constructible, i.e. if $\neg b$, $\neg e$, and f are facts and $\neg b, \neg l —[\mathit{strong}]\rightarrow h$ and $\neg e, f —[\mathit{conclusive}]\rightarrow \neg l$ are known as rules. The argument for *h* defeats the argument for *m* because *h* contradicts a conclusion of a subargument of the argument for *m*, and the strength of the argument for *h* supersedes the strength of the subargument for $\neg h$. If the strength of the argument for *h* would be *fair*, instead of *strong*, then the argument for *h* would interfere with the argument for *m*. And if the strength of the argument for *h* would be weaker than *fair*, then the argument for *h* would degrade into an (ineffective) counterargument.

It is important to see that a defeater is not only a destructive concept. A defeater may also reinstate arguments by defeating a defeater. Arguments may be repeatedly defeated an reinstated before reaching their ultimate logical status. It follows that the logical interaction among arguments is a complex phenomenon.

In defining which argument, given the facts, are ultimately undefeated, I follow John Pollock (1987). His definition of level-*n* arguments is a convenient way to do this.

6.1.2. DEFINITION. (John Pollock, 1987). Every argument is *in at level* 1. Let $n>1$. An argument is *in at level n*, if it has no defeaters at level $n-1$. An argument is *ultimately undefeated* if it is in for all but a finite number of levels; it is said to be *ultimately defeated* if it is out, i.e. not in, for all but a finite number of levels; in all other cases, the argument is said to be *provisionally defeated*. A claim is *warranted* if it is the conclusion of an argument that is ultimately undefeated.

The goal of a dispute is to find out, as fast and as fair as possible, whether the main point of contention is warranted. The protocol is responsible for this, and the ideal protocol will always be leading to the right conclusion in the most economic way. Unfortunately, ideal protocols do not exist. This is because speed and profoundness are conflicting interests that cannot be realised simultaneously. Particularly in situations where speed is preferred at the expense of profound search, the outcome of a dispute is not always completely rational, i.e., does not always agree with the definition of warrant, so that it may happen that a warranted claim is defended without success, while other claims survive dispute unjustly. This is an important point of distinction between warrant and dispute.

6.2. Promptness and correctness

Agents must agree on a protocol that, on the one hand, maintains discipline but, on the other hand, allows searching for the right conclusions. These two criteria translate as follows.

6.2.1. DEFINITION. A protocol is said to be *correct* if precisely those queries are confirmed of which the corresponding claim is warranted. It is said to be *prompt* if the transactions implied by the protocol lay a relatively modest claim on the resources available.

Informally, a correct protocol leads the dispute towards the right conclusions. If a protocol is prompt, the dispute is conducted in an economic way to save time, memory, communication facilities, or other resources.

Prompt and correctness are global and independent desiderata that neither imply nor include each other. Correctness is achieved by pursuing subgoals like fairness, accuracy, impartiality, independence, equality, uniformity, rationality, openness, certainty, and universality (Marshall, 1977). Promptness is achieved by shallow search, limited consultancy, and quick time-out's. Correct protocols are typically slow; prompt protocols are typically fast.

Promptness and correctness are properties that are mutually exclusive. They cannot be achieved within one and the same protocol. Thorough deliberation, for example, lies a severe claim on valuable resources. And cutting on resources yields prompt but hesitant conclusions. Therefore, it makes no sense in trying to prove that a protocol is both prompt and correct. At best, an open protocol can be steered back and forth between prompt and correctness. Thus, in addition, it makes no sense in trying to prove that an open protocol is *either* prompt or correct. These are properties that are determined in dispute, by arguments on, e.g., fairness, relevancy, time-out thresholds, and termination. In dispute, every open protocol can easily be directed into a state of superficial (and incomplete) reasoning. It is not self-evident, however, that an open protocol can be directed a priori into a state of complete correctness.

An important requirement of an open protocol is that, if order allows, it must be able to be directed into a state of complete correctness. A second important requirement is that there is an optimal tradeoff between the grade of correctness and the resources consumed. Thus, with good protocols, a loss in momentum is substantially compensated by a gain in correctness, and a loss in correctness is substantially compensated by prompt settlements.

Below I will prove that it is possible to direct the present notion of open protocol into a state of complete correctness. As for the second requirement, however, I am unable to formulate a positive result about the tradeoff between the grade of correctness and the resources consumed. This is a practical constraint, rather than a shortcoming of the theory. In (Vreeswijk, 1994) it is proven that, in a relatively straightforward two-person protocol, the outcome of a curtailed dispute does, indeed, approximate the outcome of a complete dispute. It was also shown, however, that the quality of these approximations is determined entirely by the criteria according to which dispute is curtailed. More to the point, if a specific criterion causes the group to provisionally accept arguments that are refuted in a later stage, then it will be clear that the provisional conclusions are poor approximations of the final result, and become useful only when dispute is drawing to a close. Moreover, if the group is very unlucky, it may even happen that, at the time when the final moves are to be made, the last argument might cause a drastic turnover in the collection of arguments provisionally established (which is in accordance with practice, in which each party may keep the best for last). Thus, a favorable tradeoff between correctness and promptness cannot be proved within the present framework.

Let us prove that open protocol can be correct. Before we can do this, we will have to hark back to John Pollock's theory of level-*n* arguments.

1. In Pollock's theory of level-*n* arguments, every argument is *in* or *out* at various levels. The

norm of an argument is the first level where this becomes definitive. An argument that is *in* at level 1,3, and 5, for example, and *out* at all other levels, has norm 6. A norm is a classification of complexity among arguments. Still, we need a refined classification. Let us refine our classification by putting an order of complexity on arguments. The complexity of an argument is determined by the norm first, and if arguments have an equal norm, then arguments are compared in size. Technically, the *complexity* of an argument is an ordered pair consisting of the norm and the size of the argument, respectively. Arguments are compared lexicographically. Thus, an argument of complexity (6,8) is less complex than an argument of complexity (6,9), and an argument of complexity (6,9) is less complex than an argument of complexity (7,1).

2. Like arguments, claims have a norm and a complexity. The norm of a claim depends on the norm of its supporters. If the claim is supported by at least one valid argument, the norm of the claim is defined as the minimal norm of the valid arguments supporting that claim; else, if there is a provisionally defeated supporter, the norm is undefined; if there are only supporters that are ultimately defeated, then the norm of the claim is defined as the maximal norm of its supporters. Claims are ordered lexicographically, in a similar way as with arguments.
3. There are two results that we need. Both results follow with the help of Definition 6.1.1. The first result is that every valid (i.e. ultimately undefeated) argument has a strictly higher norm than each of its defeaters. The second result is that an ultimately defeated argument has at least one valid defeater with a strictly lower norm. These results may be examined by the interested reader. From these two results it follows that every valid argument is more complex than each of its defeaters *and* each of its subarguments. Moreover, it trivially follows that an ultimately defeated argument is more complex than at least one valid defeater.

Having put an order of complexity on arguments, we are now ready to prove the main proposition.

6.3. PROPOSITION. (Correctness of the basic protocol.) *Assuming that multi-agent deliberation is not curtailed by amendments in the order of procedure, queries are confirmed if and only if they are warranted.*

PROOF. We prove this equivalence with induction on the complexity of the claims involved. Claims of basic complexity have a norm and size equal to one, and thus are facts. The absence of restrictive amendments enables agents to follow the basic protocol, which instructs them to confirm facts. This establishes the proposition for basic claims.

Now suppose the equivalence holds for all claims upto, but not including, a certain complexity. Consider a fixed claim of the complexity considered. This claim can be warranted or not. Suppose it is. In that case, there exists a valid argument supporting that claim. If this argument is a fact, we have a case like above. If it is not a fact, its top rule must be in the knowledge base of some (other) agent. Let us consider the activities of this agent. This agent has to find support for the antecedent of the top rule, and it has to search for information against the consequent of the top rule. Using the two results above, it follows that all the arguments involved are less complex than the argument considered. According to the induction hypothesis this means that all the antecedent queries are confirmed, and that the consequent query is confirmed with an inadequate modality, or is not confirmed at all. In the absence of amendments stipulating the contrary, this suffices to confirm the main query.

If the claim is not warranted, it can be supported by an argument or not. If not, failure of query confirmation will be clear. If support does exist, however, then the complexity of this argument must be equal to the complexity of the claim considered. Using the two results above,

there must exist a valid defeater that is less complex. According to the induction hypothesis this means that every query that corresponds to the conclusion of this defeater is confirmed. This holds in particular for the query that is generated to see if there may be information against the main query. Thus, the main query will not be confirmed. □

It is important to remember that the entire proof rests on the assumption that, throughout the whole run, every agent has ample resources to perform exhaustive search. It is assumed that communication is perfect, that search is properly distributed and that consultancy may be iterated to arbitrary depth. It is assumed that consultation is adequate, and that time-out mechanisms do not apply. In summary, it is assumed that agents follow the basic protocol unconditionally, ad infinitum.

If agents follow the basic protocol 'wherever it goes', conclusions must be awaited for. Therefore, correct protocols are impractical. Correct protocols are available in low level communication environments (Tanenbaum, 1989), but do not exist in advanced multi-agent systems. This, once again, underlines the need for open protocol.

7. Conclusion

In this article, I explored the notion of open protocol in multi-agent systems. First in theory and then by example I showed how protocol may alter itself by directing dispute towards amendments. I argued that performance of open protocols is determined by prompt and correctness, two desiderata that cannot be achieved simultaneously. Finally, I showed that every open protocol must be correct if promptness is considered irrelevant. (And I argued that the converse is trivial.)

The need for open protocol has already been recognized in the literature but the discussions never materialized in formal definitions. I pretend to have provided a premature formalization.

Three characteristics of open protocol motivate further elaboration of the theory. First, highly developed systems demand open interaction structures. Intelligent agents typically operate in a concurrent, asynchronous, decentralized, and sometimes inconsistent environment for which no a priori patterns of interaction exist. This obliges further study of open protocol. Second, open protocol is self-directing. Self-directing control and self-government is considered to be an essential property of intelligence and autonomy. Third, the presence of self-regulating protocol circumvents the difficult task of protocol specification. In the ideal case, the protocol's specification develops out of the context, just like what happens in the organization of social talk (Taylor and Cameron, 1987). We are in a remote position in this respect but, on the other hand, the appearance of specifications emerging out of the context is a current phenomenon for those who engage in genetic programming (Goldberg, 1989), (Koza, 1992). These three characteristics make open protocol to be an important, promising, and challenging subject, a subject that certainly warrants further study and exploration.

Acknowledgements. The authors like to thank Frank de Boer, Gerhard Brewka, Tom Gordon, Wiebe van der Hoek, Ron Loui, John-Jules Meyer, Henry Prakken, Yao-Hua Tan, and Bart Verheij for their efforts in helping to organize this material. Thanks are also due to the organizers and participants of FoReST, the Dutch Colloquium on Formal Representation and Specification Techniques.

References

1. Burmeister, B., Haddadi, A., and Sundermeyer, K., Generic Configurable Cooperation Protocols for Multi-Agent Systems, (1993). unpublished manuscript, Daimler-Benz AG Research and Technology.
2. Carver, C.S. and Scheier, M.F., (1981). *Attention and self-regulation: a control-theory approach to human behavior*, Springer Series in Social Psychology, Springer, New York.
3. Conry, S., Meyer, R., and Lesser, V., Multistage Negotiation in Distributed Planning, (1986). COINS technical report 86-67, University of Massachusetts. Also appeared in 'Readings in Distributed Artificial Intelligence', A.H. Bond, L. Gasser (eds.), Morgan Kaufmann Publishers, pp. 367-384 (1988).
4. Davis, R. and Smith, R.G., (1983). Negotiation as a metaphor for distributed problem solving, in: *Artificial Intelligence*, Vol. 20, Nr. 1, pp. 63-109, North Holland.
5. Doyle, J.C., Francis, B.A., and Tannenbaum, A.R., (1992). *Feedback Control Theory*, Springer Series in Social Psychology, Macmillan, New York.
6. Durfee, E.H. and Lesser, V.R., (1987). Using Partial Global Plans to Coordinate Distributed Problem Solvers, in: *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 875-883.
7. Georgeff, M.P., (1987). Planning, in: *Annual Reviews of Computer Science*, pp. 359-400. Also published in 'Readings in Planning,' J. Allen, J. Hendler and A. Tate (eds.), Morgan Kaufman, 1990.
8. Goldberg, D.E., (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Springer Series in Social Psychology, Addison-Wesley, New-York.
9. Gordon, T.F., The Pleadings Game: An Artificial Intelligence Model of Procedural Justice, (1993). docotoral dissertation, Technische Hochschule Darmstadt.
10. Hage, J.C., (1993). Monological Reason Based Logic, in: *Proceedings of the 4th International Conference on AI and Law*, pp. 30-39, ACM Press.
11. Hewitt, C., (1991). Open Information Systems Semantics for Distributed Artificial Intelligence, in: *Artificial Intelligence*, Vol. 47, pp. 79-106.
12. Hewitt, C.E., (1986). Offices are Open Systems, in: *ACM Transactions on Office Information Systems*, Vol. 4, Nr. 3, pp. 271-287.
13. Hofstadter, D.R., (June 1982). About Nomic: A Heroic Game That Explores the Reflexivity of the Law, in: *Scientific American*, Nr. 246, pp. 16-28.
14. Koza, J., (1992). *Genetic Programming: On the Programming of Computers By Means of Natural Selection*, Springer Series in Social Psychology, MIT Press.
15. Ladner, R.E. and Reif, J.H., (1986). The Logic of Distributed Protocols (Preliminary Report), in: *Proceedings of the First Conference on Theoretical Aspects of Reasoning about Knowledge*, ed. J.Y. Halpern, Springer Series in Social Psychology, pp. 207-222.
16. Lansky, A.L., (1991). Local Search for Multi-Agent Planning, in: *Proceedings of the International Joint Conference on Artificial Intelligence*.
17. Loui, R.P., Process and Policy: Resource-Bounded Non-Demonstrative Argument, (October 1992). draft, Department of Computer Science, Washington University, St. Louis.
18. Lynch, N.A. and Fischer, M.J., (1981). On Describing the Behavior and Implementation of Distributed Systems, in: *Theoretical Computer Science*, Vol. 13, pp. 17-43.

19. Marshall, G., (1977). Due Process in England, in: *Due Process*, ed. J.W. Chapman, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, pp. 69-89, New York University Press.

20. Miller, C.A., (1977). The Forest of Due Process of Law: The Americal Constitutional Tradition, in: *Due Process*, ed. J.W. Chapman, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, pp. 3-68, New York University Press.

21. Mohan, C., (1984). *Tutorial: Recent Advances in Distributed Database Management*, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, IEEE Computer Society Press.

22. Page, C.V., (1991). Principles for Democratic Control of Bounded-Rational, Distributed, Knowledge Agents, in: *Proceedings of the European Simulation Multiconference*, pp. 359-361.

23. Pollock, J.L., (1987). Defeasible Reasoning, in: *Cognitive Science*, Vol. 11, pp. 481-518.

24. Pollock, J.L., (1992). How to Reason Defeasibly, in: *Artificial Intelligence 57*, pp. 1-42.

25. Rescher, N., (1977). *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, State University of New York Press, Albany.

26. Robert, H.M., (1971). *Robert's rules of order Revised—The classic 1915 edition*, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, Morrow paperback editions, New York. Foreword by General Henry M. Robert III.

27. Sartor, G., (1993). *A Simple Computational Model for Nonmonotonic and Adversarial Reasoning*, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, p. Proceedings of the 4th International Conference on AI and Law, huisdrukkerij Vrije Universiteit, Amsterdam.

28. Stary, C., (1991). Modeling Decision Support for Rational Agents, in: *Proceedings of the European Simulation Multiconference*, ed. E. Mosekilde, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, pp. 351-356.

29. Stirling, W., (1992). Multi-Agent Coordinated Decision-Making Using Epistemic Utility Theory, in: *Proceedings of the MAAMAW*, National Research Council, Institute of Psychology, Rome, Italy.

30. Suber, P., (1990). *The Paradox of Self-Amendment, A Study of Logic, Law, Omnipotence, and Change.*, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, Peter Lang Publishing.

31. Tanenbaum, A.S., (1989). *Computer Networks*, NOMOS XVIII, Yearbook of the American Society for Political and Legal Philosophy, Prentice Hall.

32. Tenney, R.R. and Jr., N.R. Sandell, (1981). Strategies for Distributed Decisionmaking, in: *IEEE Transactions on Systems, Man and Cybernetics*, pp. 527-538.

33. Vreeswijk, G.A.W., (1993). Defeasible Dialectics: A Controversy-Oriented Approach towards Defeasible Argumentation, in: *The Journal of Logic and Computation*, Vol. 3, Nr. 3, pp. 3-27. Available via anonymous ftp from ftp.cs.rulimburg.nl :/pub/papers/vreeswyk.

34. Vreeswijk, G.A.W., Studies in Defeasible Argumentation, (1993). doctoral dissertation, Department of Mathematics and Computer Science, Vrije Universiteit Amsterdam.

35. Vreeswijk, G.A.W., Self-modifying protocol games: A Preparatory Case Study, (1994). technical report, Vakgroep Informatica (FdAW), Rijksuniversiteit Limburg, Maastricht,

The Netherlands. In preparation. Available via anonymous ftp from ftp.cs.rulimburg.nl :/pub/papers/vreeswyk.

36. Vreeswijk, G.A.W., Searching for the Impossible in Multi-agent Systems, (1994). technical report, Vakgroep Informatica (FdAW), Rijksuniversiteit Limburg, Maastricht, The Netherlands. In preparation. Available via anonymous ftp from ftp.cs.rulimburg.nl :/pub/papers/vreeswyk.

37. Vreeswijk, G.A.W., The Computational Value of Debate in Defeasible Reasoning, (1994). report IR-309, Faculteit Wiskunde en Informatica, Vrije Universiteit Amsterdam. To appear in *Argumentation*. Available via anonymous ftp from ftp.cs.rulimburg.nl :/pub/papers/vreeswyk.

38. Werner, E., (1989). Cooperating Agents: A Unified Theory of Communication and Social Structure, in: *Distributed Artificial Intelligence*, ed. M.N. Huhns, Research Notes in Artificial Intelligence, vol. II, pp. 3-36, Morgan Kaufmann Publishers, San Mateo, CA.

39. Zlotkin, G. and Rosenschein, J.S., (1992). Cooperation and Conflict Resolution via Negotiation Among Autonomous Agents in Noncooperative Domains, in: *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 21, Nr. 6, pp. 1317-1324.

Appendix

This appendix is a list of concepts introduced in this paper. Each concept is followed by a short description.

Amendment.

A procedural claim, intended to alter the protocol.

Amendment, Successful

→ Successful amendment.

Argument.

A set of rules, connected to each other in the form of a tree.

Argument, Constructible

→ Constructible argument.

Argument, Counter

→ Counterargument.

Argument, Interfering

→ Interfering argument.

Belief.

A belief is an element of the set of working beliefs.

Claim.

A modalized literal (see also: set of working beliefs).

Claim, Procedural

→ Procedural claim.

Claim, Proper

→ Proper claim.

Collective knowledge.

The union of all knowledge basis in the multi-agent epistemic environment.

Constructible argument.

An argument of which all the premises are facts.

Counterargument.

An argument of which the conclusion contradicts the conclusion of a subargument of the argument in question.

Counterargument, Defeating

→ Defeating counterargument

Defeater.

A defeating counterargument.

Defeating counterargument.

A counterargument that is stronger than the target it opposes.

Discourse, Field of

→ Field of discourse.

Domain knowledge.

Part of the knowledge base that pertains to the field of discourse.

Fact.

A claim that is an element of the knowledge base.

Field of discourse. (Domain of discourse, Topic of discourse.)

The topic on which an agent, or a multi-agent system, reasons about.

Interfering counterargument.

A counterargument that is as strong as the target it opposes.

Knowledge, Domain

→ Domain knowledge.

Knowledge, Procedural

→ Procedural knowledge.

Knowledge base.

A set of facts and rules. A knowledge base is denoted by K.

Modality

Primarily, a label used to indicate the strength of a rule. Using the weakest link principle for arguments, modalities of rules can be used to attribute modalities to claims.

The set of modalities is denoted by m.

Norm (of a claim).

The formal definition of the norm of a claim is a bit involved (cf. the text). Informally, the norm of a claim represents its 'inherent dialectical complexity', the number of turns pro and con it takes before its status is definitely established.

Norm (of an argument).

The first level at which an argument stops alternating between *in* and *out*.

Open protocol.

A protocol lying open to further alteration and adjustment.

Order (of procedure).

Set of procedural claims held by a particular agent; the order held by agent i is denoted by $B_i \cap P$.

Procedural claim.

Claim pertaining to the protocol.

The set of all procedural claims existing is denoted by P.

Procedural knowledge.

Knowledge (= facts and rules) pertaining to procedure.

Knowledge, Collective

→ Collective knowledge

Protocol.

A function mapping local states of agents to (informal) action descriptions.

Protocol, Open

→ Open protocol.

Open rule.

A rule without a modality indicating its strength.

Rule.

Symbolic representation of a reason, meant to be used by (artificial) reasoners in the construction of arguments; a reason is an irreducible step of reasoning.

Rule, Open

→ Open rule.

Run.

The sequence of states through which the system passes as the result of an initiating action (like issuing a query).

Successful amendment.

Amendment that is defended successfully in dispute.

Target

A subargument of a larger argument, aimed at by a counterargument.

Working beliefs, Set of

A working belief is either a claim or an element of the knowledge base. Intuitively, a working belief is a epistemic commitment on the basis of which can be operated if necessary. A proper working belief, i.e. a working belief that is not an element of the knowledge base, is a tentative epistemic commitment.

The set of working beliefs held by agent i is denoted by B_i.