Revising Beliefs Through Arguments: Bridging the Gap Between Argumentation and Belief Revision in MAS

Fabio Paglieri[1] and Cristiano Castelfranchi[2]

[1] University of Siena, Piazza S. Francesco 8, 53100 Siena, Italy
paglieri@media.unisi.it
[2] Institute of Cognitive Sciences and Technologies, National Research Council (ISTC-CNR), Viale Marx 15, 00137 Roma, Italy
c.castelfranchi@istc.cnr.it

Abstract. This paper compares within the MAS framework two separate threads in the formal study of epistemic change: belief revision and argumentation theories. Belief revision describes how an agent is supposed *to change his own mind*, while argumentation deals with persuasive strategies employed *to change the mind of other agents*. These are two sides (cognitive and social) of the same epistemic coin: argumentation theories are incomplete, if they cannot be grounded in belief revision models – and vice versa. Nonetheless, so far the formal treatment of belief revision mostly neglected any systematic comparison with argumentation theories. In MAS such problem becomes evident and inescapable: belief change is usually triggered by communication and persuasion from other agents, involving deception, trust, reputation, negotiation, conflict resolution (all typical issues faced by argumentation-based models). Therefore, a closer comparison between belief revision and argumentation is a necessary preliminary step towards an integrated model of epistemic change in MAS.

1 Belief Revision Without Argumentation

Following the seminal work in [14], belief revision has recently become an extremely active area of research at the confluence between AI, logic, cognitive science, and philosophy. Notwithstanding the impressive amount and quality of studies devoted to this topic (including many researches in the MAS community, e.g. [1, 9, 12, 13, 31]), belief revision has been mainly addressed in a rather single-minded fashion, isolating the issue of belief change from other related features of cognitive processing. As remarked in [26], current theories of belief revision have been put forward and discussed in a sort of epistemological vacuum, without providing a more comprehensive account of epistemic states and dynamics. Moreover, the process of belief change has been usually conceived as an isolated activity, neglecting even the most obvious connections with other cognitive tasks: e.g. inferential reasoning, communication, argumentation (significant exceptions to this trend are in [10, 13]). On the contrary, we claim that belief revision should be investigated as a specific function (albeit a crucial one) in the cognitive processing of epistemic states, integrating formal models of belief change in a more comprehensive epistemological theory, and providing systematic connections with related cognitive tasks.

I. Rahwan et al. (Eds.): ArgMAS 2004, LNAI 3366, pp. 78–94, 2005.

1.1 Limitations of Current Theories

The AGM paradigm [14] has been the most influential model of belief revision so far, serving as a frame of reference for both refinements and criticisms of the original proposal. Roughly summarizing (see [21] for further discussion), this model was first conceived as an idealistic theory of rational belief change: belief states were characterized as sets of propositions (infinite and deductively closed), three basic types of change were described (expansion, contraction, revision), and rationality was expressed by a set of postulates binding these operators. To decide between different outcomes of the revision process (i.e. different sets of propositions consistent with the rationality postulates), an ordering criterion was introduced in the original belief state, ranking propositions for their importance (epistemic entrenchment).

This approach to belief revision fails to integrate with argumentation theories for two reasons: (1) it does not make any predictions or assumptions about how and why some propositions come to be believed, rather than others; (2) there is a deliberate lack of structural properties in the characterization of epistemic states. Argumentation theories capture how a desired change in the audience's beliefs is brought about by the arguer: therefore, without an explicit theory of *the reasons to believe something*, the whole point of argumentation is lost. AGM-style approaches to belief revision simply lack the necessary internal structure to describe argumentative strategies (for a philosophically oriented discussion of justification in belief revision, see [15]).

In this respect, the so called *foundation theories* of belief revision fare better than AGM, since they provide a precise account of the reasons supporting a given belief, e.g. using Truth Maintenance Systems [8]. Similar proposals have also been advanced in the field of multi-agent systems [9, 12, 13, 18], and there are several analogies between the criticisms to the AGM approach discussed in this paper and objections raised within the TMS community (e.g. the need for detailed analysis of the reasons that support and determine the agent's beliefs), although our approach is more cognitive-oriented, while TMS put greater emphasis on computational issues.

Since a detailed comparison between our approach and TMS is beyond the aim of this work (cf. 4 on future developments in the direction), here we will provide only a short comment on belief change and argumentation in TMS. In spite of the richer framework outlined by TMS for belief revision in MAS, only few of these theories explicitly address argumentation and/or communication (e.g. [18]), and the structural properties of epistemic states are restricted to factual supports for the agent's beliefs, to ensure an accurate weighting of unreliable and/or contrasting sources of information. Although such structures are essential to integrate belief revision and argumentation, they are not enough: a fairly rich picture of argumentative strategies must include motivational and emotional features [7, 11, 16, 17], not only factual credibility. Since also belief revision is affected by similar considerations, a more comprehensive cognitive model of epistemic change must be devised (cf. 2.1-2.4).

2 A Cognitive Model of Data-Oriented Belief Revision (DBR)

The following sections provide a short outline of an alternative model of belief revision, i.e. Data-oriented Belief Revision (DBR): for further details, see [6, 21]. Although this model is still mainly theoretical and far from implementation in MAS, it is conceived as a realistic cognitive framework for understanding belief revision in agent-based social simulation[1]. Special emphasis is given to the representation of *individual variation in belief revision* (cf. 2.2, 2.4): it is extremely important, for the sake of cognitive plausibility and social simulation, to be able to model different strategies of epistemic change for different agents, and to represent all of them within the same conceptual framework. This has a significant impact on argumentation as well, since it allows to model different argumentation strategies and to distinguish between local and global arguments (cf. 3.5).

2.1 Data and Beliefs: Properties and Interactions

Two basic epistemic categories, *data* and *beliefs*, are put forward in this model, to account for the distinction between pieces of information that are simply *gathered and stored* by the agent (data), and pieces of information that the agent considers *reliable bases for action, decision, and specific reasoning tasks*, e.g. prediction and explanation (beliefs). Clearly, the latter are a subset of the former: the agent might well be aware of a datum that he does not believe (i.e. he does not consider reliable enough); on the other hand, the agent should not be forced to forget (i.e. to lose as a datum) a piece of information which he temporarily rejects as a belief [6]. Moreover, a rejected piece of information retains significant epistemic properties (e.g. its own unreliability, and the reasons for it) that will often be crucial in future revisions and should be preserved by a formal model of belief change [9, 26].

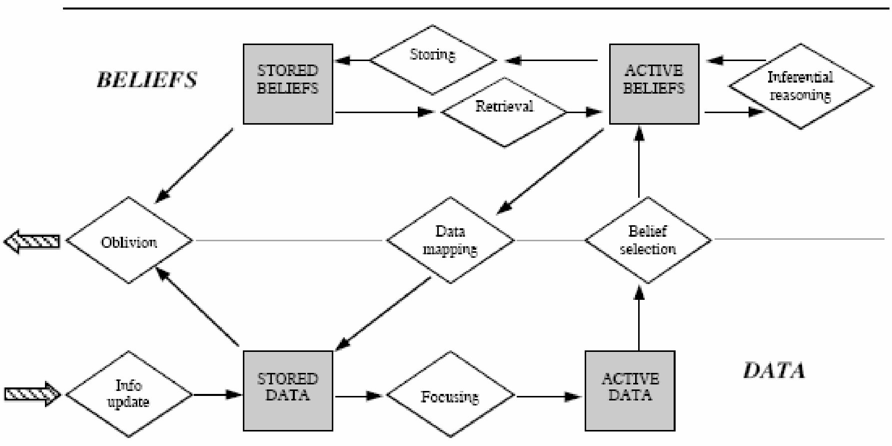

Fig. 1. Epistemic processing in Data-oriented Belief Revision

The distinction between data and beliefs yields a number of consequences for the formal study of epistemic dynamics: to start with, it leads to conceive *belief change as a two-step process*. Let us consider external belief change (cf. 2.3): whenever a new piece of evidence is acquired through perception or communication, it affects *directly*

[1] Broader accounts of belief revision have been advocated also for epistemic change in communication [13] and in defeasible reasoning [10, 26].

the agent's data structure, and only *indirectly* his belief set. In other words, the effects (if any) of the new datum on the agent's beliefs depend (1) on its effects on the other data, and (2) on the process of belief selection applied by the agent over such data (cf. 2.2). We call this procedure *Data-oriented Belief Revision (DBR)*.

More generally, data and beliefs define the two basic cognitive layers of the whole epistemic processing performed by the agent, as summarized in Figure 1. An exhaustive discussion of this general model is beyond the aim of this paper: here we will focus mainly on the treatment of data, with special reference to information update, data properties and assessment, and belief selection (cf. 2.2-2.3), since these are the features most directly involved in belief revision. However, it is important to keep in mind the overall epistemic processing, if we want to provide a formal model adequate to express belief change in cognitive agents.

In this model data are selected (or rejected) as beliefs on the basis of their properties, i.e. the possible *cognitive reasons to believe* them. DBR accounts for four distinct properties of data [6, 21]:

I. *Relevance*: a measure of the pragmatic utility of the datum, i.e. the number and values of the (pursued) goals that depends on that datum.
II. *Credibility*: a measure of the number and values of all supporting data, contrasted with all conflicting data, down to external and internal sources;
III. *Importance*: a measure of the epistemic connectivity of the datum, i.e. the number and values of the data that the agent will have to revise, should he revise that single one;
IV. *Likeability*: a measure of the motivational appeal of the datum, i.e. the number and values of the (pursued) goals that are directly fulfilled by that datum.

The assessment of credibility is discussed in 2.3, while the assessment of importance, relevance and likeability is detailed in [21]. In DBR, credibility, importance and likeability determine the outcomes of belief selection, i.e. whether a candidate data is to be believed or not, and with which strength (cf. 2.2), while relevance is crucial in pre-selecting the sub-set of active data (focusing), i.e. determining which data in the agent's data base are useful/appropriate for the current task, and should therefore be taken in consideration as candidate beliefs (an in-depth discussion of focusing is given in [21]). While relevance and likeability depend on a comparison between data and goals, credibility and importance basically *depend on structural relations between data* [6]. In fact, in DBR data bases are highly structured domains, best conceived as *networks*: data are represented as *nodes*, interconnected through characteristic functional relations (cf. 2.3), i.e. *links* in the network.

Table 1. Data and beliefs: an overview

	Basic properties	Organization principle	Internal dynamics	Interaction principle
DATA	*Relevance, credibility, importance, likeability*	*Networks*	*Updates, propagation*	*Belief selection*
BELIEFS	*Strength*	*Ordered sets*	*Inferential reasoning*	*Data mapping*

The agent's beliefs emerge from his data base through the selection process (cf. 2.2). Beliefs are characterized by *strength*, which reflects their implicit ordering. Strength is determined by the selection process from the values of credibility, importance, and likeability of the corresponding active datum. Therefore beliefs are organized in *ordered sets*, rather than networks [14, 21].

The basic distinction between data and beliefs yields a rich picture of epistemic dynamics (Fig. 1 and Table 1). From a computational viewpoint, such distinction opens the way for *blended approaches to implementation* [21]: data structures present remarkable similarities with Bayesian networks and neural networks, while belief sets are a well-known hallmark of AGM-style belief revision [14]. Moreover, data and beliefs are here conceived as *different stages, roles, and functions in the processing of internal epistemic states*, to be accounted for in the agent architecture.

2.2 Belief Selection

Once the informational values of the active data are assessed (cf. 2.3), a selection over such data is performed, to determine the subset of reliable information (i.e. beliefs) and their degree of strength. Every time new relevant information is gathered by the agent, modifying his data network and the subset of active data, the belief selection takes place anew, possibly (but not necessarily) changing the agent's belief set.

This process of belief selection regulates the interaction from data to beliefs, determining (1) what data are to be believed, given the current informational state, and (2) which degree of strength is to be assigned to each of them. The outcome of belief selection is determined by the informational values of the candidate data (credibility, importance, likeability) and by the nature of the agent's selection process.

In DBR the agent's belief selection is represented by a mathematical system, including a *condition* C, a *threshold* k, and a *function* F. Condition C and threshold k together express the minimal informational requirements for a datum to be selected as belief. The function F assigns a value of strength to the accepted beliefs. Both C and F are mathematical functions with credibility and/or importance and/or likeability as their arguments. Given a datum ϕ, c^ϕ, i^ϕ, l^ϕ are, respectively, its credibility, importance, and likeability. Let $\boldsymbol{B}$ represents the set of the agent's beliefs, and $B^s\phi$ represents the belief ϕ with strength s. The general form of the selection process is:

if $C(c^\phi, i^\phi, l^\phi) \leq k$	then	$B^s\phi \notin \boldsymbol{B}$
if $C(c^\phi, i^\phi, l^\phi) > k$	then	$B^s\phi \in \boldsymbol{B}$ with $s^\phi = F(c^\phi, i^\phi, r^\phi)$

The setting of C, F and k is an individual parameter, which might vary in different agents (cf. 2.4). Examples of individual variation in belief selection are the following:

$C: c^\phi > k$	$k: 0.5$	$F: c^\phi$
$C: c^\phi > k$	$k: 0.6$	$F: (c^\phi + i^\phi + l^\phi) / 3$
$C: c^\phi > k \times (1 - l^\phi)$	$k: 0.8$	$F: c^\phi \times (i^\phi + l^\phi)$

All these parametrical settings assign to data credibility the main role in determining belief selection, but they do so in widely different ways. The first parametrical setting expresses a thoroughly realistic attitude towards belief selection, regardless of any considerations about importance or likeability. At the same time, the minimal threshold is set at a quite tolerant level of credibility (0.5). The threshold is slightly higher in the second parametrical setting, and the condition is identical: on the whole, this reflects a more cautious acceptance of reliable data. But once a datum is indeed accepted as belief, its strength is now calculated taking in account also importance and likeability, in contrast to the previous setting. The same happens in the third parametrical setting, although along different lines. Here the threshold is extremely high (0.8), but the condition is influenced by likeability as well: assuming that likeability ranges in the interval **[0, 1]**, here the minimal threshold over credibility is conversely proportional to the likeability of the datum (e.g. it is 0.08 for a datum with likeability 0.9 vs. 0.72 for a datum with likeability 0.1). That expresses a systematic bias towards the acceptance of likeable (i.e. pleasant) data, in spite of their credibility. In other words, these parametrical settings define three agents with different personalities, with respect to belief selection: a *tolerant full realist* (the first), a *prudent open-minded realist* (the second), and a *wishful thinking agent* (the third).

Allowing several parametrical settings in belief selection (as well as in other features of DBR, cf. 2.4) serves to capture *individual variation in epistemic dynamics*, i.e. specifying different strategies of belief change for different agents and/or for different contexts and tasks[2]. It also shows that, although the selection process in DBR is just a mathematical simplification of the cognitive process of belief selection, it is extremely flexible and expressive, since we can manipulate and set condition, function and threshold in such a way to express different selection strategies, with an high degree of sophistication. Moreover, a mathematically straightforward treatment of individual variation will prove essential for investigating evolutionary dynamics in shaping belief revision strategies in MAS, e.g. applying genetic algorithms over population of agents with randomized internal settings (cf. 4).

2.3 Information Update and Data Assessment

Belief revision is usually triggered by *information update* either on a fact or on a source: the agent receives a new piece of information, rearranges his data structure accordingly, and possibly changes his belief set, depending on the belief selection process. Information update specifies the way in which new evidences are integrated in the agent's data structure. We define *external belief selection* the process of epistemic change triggered by information update, in contrast to *internal belief revision*, i.e. belief change due to inferring a new piece of information from old premises (on internal belief revision, see [21]).

Data structures are conceived as networks of nodes (data), linked together by characteristic relations. For the purposes of the present discussion, it will suffice to define three different types of data relations: support, contrast, and union.

---

[2] Individual variation in MAS is a major concern also for argumentation studies, e.g. as a way of framing a theory of personality in multi-agent platforms (see for instance [19]).

I. *Support*: ϕ supports ψ ($\phi \Rightarrow \psi$) iff $c^{\psi} \propto c^{\phi}$, the credibility of ψ is directly proportional to the credibility of ϕ.
II. *Contrast*: ϕ contrasts ψ ($\phi \perp \psi$) iff $c^{\psi} \propto 1/c^{\phi}$, the credibility of ψ is conversely proportional to the credibility of ϕ.
III. *Union*: ϕ and ψ are united (ϕ & ψ) iff c^{ψ} and c^{ϕ} jointly (not separately) determine the credibility of another datum γ.

New external information generates not only a datum concerning its *content*, but also data concerning *source attribution* and *source reliability*, and the *structural relations* among them. More precisely, information update brings together:

I. a datum concerning the content (object datum, *O-datum*);
II. a datum identifying the information source (*S-datum*);
III. a datum concerning the reliability of the source (*R-datum*).

These data are closely related, since the credibility of the new information depends on the jointed credibility of the other two data: i.e. the union of the S-datum and the R-datum supports the O-datum (Fig. 2). Once an agent has been told by x that ϕ holds, his confidence in ϕ will depend on the reliability he assigns to x, provided he is sure enough that the source of ϕ was indeed x. The environmental input is characterized by a content ϕ (e.g. its propositional meaning), a source x (e.g. another agent), and a noise n (affecting both source identification and content understanding)[3].

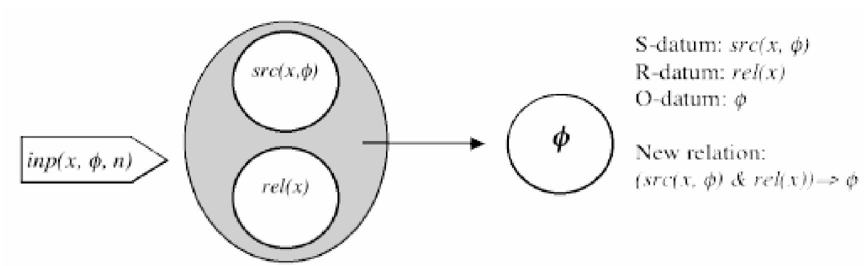

Fig. 2. Information update: integrating new external data

While pragmatic relevance, epistemic importance and motivational likeability of a datum are further discussed in [6, 21], here we focus on credibility. The credibility of a given datum depends on the credibility of its supports, weighted against the credibility of its contrasts [6, 12, 26]. Each agent must be equipped with a specific algorithm to determine such value. Although this algorithm is an individual parameter (different agents can use different heuristics), it must obey the general definition of support and contrast relations. This is an example of *credibility algorithm*[4]:

[3] More sophisticated models (e.g. [9]) might take in account also the degree of certainty over the content expressed by the source, allowing agents to communicate information with different shades of confidence.

[4] It is convenient to range credibility in the close interval **[0, 1]**, but of course this does not necessarily lead to probabilistic accounts of epistemic dynamics.

$$c^{\alpha} = (1 - \Pi_{\mu \in S\alpha}(1 - c^{\mu})) \times \Pi_{\chi \in K\alpha}(1 - c^{\chi})$$

with $S\alpha$ = the set of all data supporting α
$K\alpha$ = the set of all data contrasting α

Support and contrast determine the credibility of one *relatum* in terms of the credibility of the other. Union takes in account the credibility of both *relata* at the same time, in order to assess the credibility of a third datum – either supported or contrasted. An example is given by information update (Fig. 2): the credibility of the O-datum depends on the credibility of the union of S-datum and R-datum. Therefore we need to specify a *union algorithm* for each agent [21]: i.e. a procedure to assess the credibility of $(\phi \,\&\, \psi)$, given the credibility of ϕ and ψ. For instance:

$$c^{\phi \& \psi} = min(c^{\phi}, c^{\psi})$$

Now we have enough elements to provide a quantitative description of information update, and not only a qualitative one. The credibility of the O-datum will depend on the credibility of the union of the S-datum (here with $c = 1$, assuming noiseless communication by hypothesis) and the R-datum, weighted against the credibility of all contrasting evidences (if any), according to the credibility algorithm of that particular agent. Assessment of source reliability is thoroughly discussed in [6, 12].

2.4 Principles and Parameters

The model of belief revision presented so far is based on a conceptual distinction between *principles* and *parameters*. Principles are *general* and *qualitative* in nature, defining the common features which characterize epistemic processing in every agent. Parameters, instead, are *individual* and *quantitative*, specifying in which fashion and measure each agent applies the universal principles of belief revision. The cognitive and social framework of the model is captured by its principles, while individual variation is represented through parametrical setting (as already showed in 2.2 concerning belief selection).

For instance, the overall two-steps dynamic of belief revision is a universal principle, while the mathematical nature of the selection process is an individual parameter. Credibility assessment will always be positively affected by supporting evidence and negatively affected by contrasting data, but the credibility algorithm might vary from one agent to another. All agents perform inferential deduction at the level of beliefs, but the specific axioms applied are a matter of individual variation – and so on. More details on parametrical setting are given in [21]: here we will only discuss their impact over argumentation (cf. 3.5).

3 Argumentation and Belief Revision

This section is devoted to highlight several connections between our model of belief change and argumentation theories: the impact of rhetorical arguments over the audience's beliefs (cf. 3.1), the different stages in Toulmin's model of argumentation (cf. 3.2), the treatment of defeasible reasoning (cf. 3.3), the role of contradictions in arguments (cf. 3.4), and the effects of individual parameters over argumentation strategies and outcomes (cf. 3.5). Presenting and discussing this variety of topics, we aim to verify the expressive power of our model of belief revision on several argumentation-related features of MAS (see also [10] for a similar attempt). A failure at this stage would testify the inadequacy of the formal model in dealing with argumentation – as it is the case with the AGM framework (cf. 3.2, 3.4, 4). On the contrary, the satisfactory results achieved by our model sound promising for future developments in the same direction [4]. However, it must be understood that this first survey is meant as preliminary recognition of a complex and exciting landscape, to test the chances of success (that we find quite favorable) for more ambitious and exhaustive attempts of integrating belief revision and argumentation (cf. 4).

3.1 Rhetoric and Audience's Beliefs

Aristotle's definition of rhetorical argument characterizes it as being especially focused on the *audience's beliefs*, rather than general acceptability. This definition is usually referred to in formal studies of rhetorical argumentation, e.g. [16], where the need for a model of belief revision (and more generally belief processing) is quite self-evident. However, as far as cognitive agents are concerned, even the most general and uncontroversial argument requires a process of belief revision in the mind of the audience: it is not the fact that p follows from q and q is the case which makes me believe p, but rather my beliefs that "p follows from q" and "q is the case". An integrated framework naturally emphasizes that any form of argumentation (including strictly logical ones) must be strongly focused on the audience's beliefs.

In our model, a crucial factor in determining whether a new piece of information will be accepted or rejected as belief is its *importance* [14, 21], i.e. the degree of connectivity (integration) of the new datum in the audience's data structure. An effective argument not only presents new information to the audience, but also provides the relevant connections with data already available to (and possibly believed by) that audience. Such connections vouch for the *plausibility* of the new datum [6] and are crucial in persuading the audience to accept it. In data networks, we distinguish two cases of argumentation through plausibility:

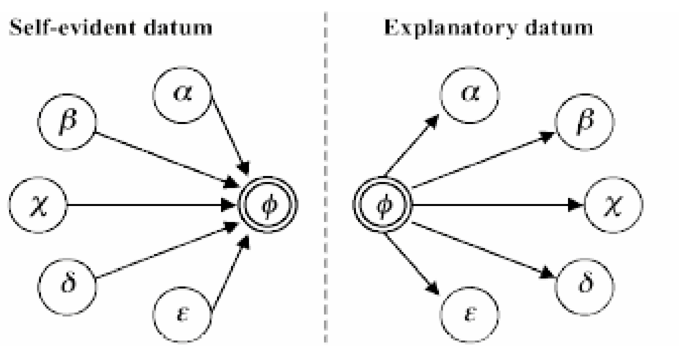

Fig. 3. Two plausibility arguments: self-evident and explanatory data

I. *Self-evident data*: the new datum is presented as following from what the audience already knew – the datum had not yet been inferred, but it might have been, and the audience is likely to remark: «Sure! Of course! Obviously!» etc.;

II. *Explanatory data*: the new datum is presented as supporting and explaining data already available – since such explanation was missing so far, it produces reactions like: «Now I see! That's why! I knew it! » etc.

This distinction is easily represented by a structured data-domain: in our model, self-evident data are data with a high number of supports, while explanatory data in turn support many other data (Fig. 3). Different degrees of self-evidence and explanatory power are expressed by epistemic importance (cf. 2.1).

3.2 Toulmin Revis(it)ed

One of the most influential account of argumentation is Toulmin's model [28], which analyzes six features of an argument: data, claim, warrant, backing, qualifier, rebuttal. *Data* are the facts (e.g. John loved his wife) which support the arguer's *claim* (e.g. John did not murder her), while the *warrant* ensures the connection between data and claim (e.g. people do not murder the ones they love), on the basis of some *backing* (e.g. murderers hate their victims); the *qualifier* specifies to what extend the warrant applies (e.g. usually), and the *rebuttal* describes special conditions which undermine the warrant (e.g. John is in bad need of money and will benefit from her insurance).

This schema is liable of immediate implementation in our model of belief revision, since it defines a specific data structure (Fig. 4). The union of data and warrant supports the claim, and the warrant is in turn supported by its backing and contrasted by the rebuttal, i.e. supports to the rebuttal make the warrant less reliable. The qualifier is represented by the degree of credibility assigned to the claim by this data structure – while more sophisticated models of source integration also distinguish between the claim's credibility and the confidence expressed by the arguer [12].

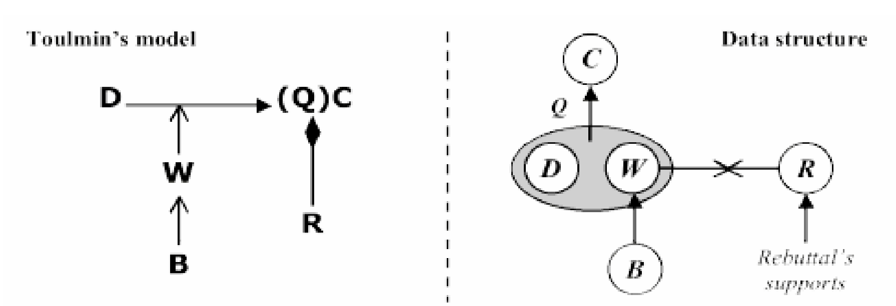

Fig. 4. Toulmin's model in data structure

This convergence is not surprising, since our model is built over the intuition that epistemic processing requires "reasons to believe" [6], and indeed argumentation is mainly concerned with the manipulation of reasons in order to change the audience's beliefs. However, it is worth noticing that other theories of belief revision fail to incorporate Toulmin's model: e.g., the AGM approach has no way to capture similar argumentative structures, without undertaking major modification of the model.

3.3 Defeasible Reasoning in Data Networks

Argumentation is often modeled in the formal framework of defeasible reasoning [2, 26], distinguishing between two kinds of defeaters (i.e. possible counterarguments against a reason-schema): *rebutting* vs. *undercutting defeaters*. Applying the terminology proposed in [28], a rebutting defeater is any reason which directly denies the claim of the argument, while an undercutting defeater is a reason which undermine the validity of the relevant warrant.

In our model, different defeaters target different nodes in the data network (Fig. 5): rebutting defeaters are data which contrast the claim-node (e.g. John has been seen shooting his wife), while undercutting defeaters are data contrasting the warrant-node (e.g. jealousy can make you kill the ones you love most). Moreover, a third category of defeaters can be expressed: *premise defeaters*, i.e. reasons which contrast the data-node (e.g. John did not love his wife). Undercutting and premise defeaters have similar function but different targets: the former attack the connection between data and claim, while the latter question the statement of fact supporting the conclusion[5].

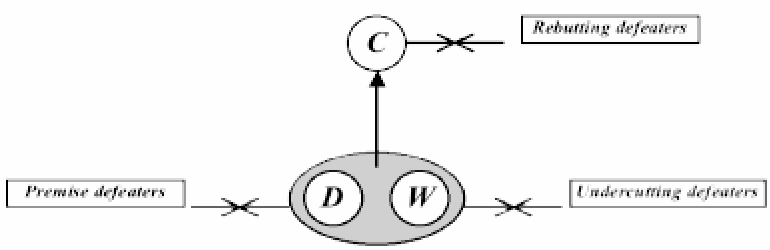

Fig. 5. Defeasible reasoning in data structure

3.4 Revising Contradictions in Argumentation

AGM-style approaches to belief revision exclude contradictions in principle, assuming belief states to be fully consistent – an untenable assumption, as far as

[5] Here we follow the terminology used in [26], but actually the expression 'rebutting defeater' is quite misleading, when compared with Toulmin's model. The rebuttal, as defined in [28], specifies the conditions which undermine the validity of the warrant, not of the claim – i.e. rebuttals are in fact undercutting defeaters. So the expression *direct defeaters* would be less ambiguous, to indicate defeaters which directly affect the claim.

cognitive agents are concerned. On the contrary, argumentation theories have been quite successful in handling inconsistency and conflicts [2, 7, 26, 28, 30], since the very idea of defeating an argument implies that such argument can be showed to be inconsistent with respect to a better one. Moreover, the AGM paradigm assumes belief states to be deductively closed, therefore infinite. This is not only a computational problem, but also a conceptual mistake: cognitive agents do not derive all the consequences from available data not only because they are resource-bounded [1, 21, 31], but mainly because they have no need to derive irrelevant consequences from accepted claims.

In our model, epistemic states are both finite and deductively open, and there is no universal insurance against contradictions. Instead, we are able to capture two relevant distinction concerning inconsistency: *implicit* vs. *explicit contradictions*, and *data contrasts* vs. *beliefs contradictions*. Agents are likely to entertain a certain number of implicitly contradictory beliefs, i.e. beliefs from which a contradiction could be derived, although the agent has not yet done so. As long as the contradiction remains implicit, the agent has no problem in handling it – just by ignoring it altogether! In fact, one of the most common strategy in argumentation consists in confronting the audience with their own contradictions, i.e. forcing them to draw contradictory conclusions from what they already believe.

In data structures, contrast relations capture 'contradictions' between data (cf. 2.3). However, these are not contradictions in the proper sense, since the contrasting data are not necessarily believed by the agent, and not necessarily at the same time: they are just information on mutually excluding states of the world. Moreover, contrasts among data are actually beneficial to the agent, since they provide him with crucial information on the credibility of both *relata*: information on $\sim p$ are useful to assess the credibility of p exactly because a contrast relation is defined between p and $\sim p$, in the form $(p \perp \sim p)$. Without such relation, negative evidence would not be evidence at all, and the efficiency of our epistemic processing would be severely impaired[6].

In other terms, *contradictions need to be solved only if they arises at the level of beliefs*, i.e. if the selection process (cf. 2.2) accepts two contrasting data as beliefs. This is rare, since credibility plays a crucial role in belief selection, and the credibility of contrasting data is conversely proportional (cf. 2.3). However, under specific circumstances (e.g. a selection which emphasizes importance and likeability over credibility) it might happen that an agent is confronted with contradictory beliefs. In this case, the contradiction is solved through reasoning, e.g. applying an axiom to reject one of the contradictory beliefs, or both.

Here we are faced with an intriguing parallel between epistemic and motivational dynamics. In BDI models of agency it has been correctly postulated that an important difference between the level of mere Desires (or wishes) and the level of Intentions, i.e. goals actually directing the agent's behavior, is that while Desires can be

[6] We agree with Aristotle that the human mind refuses contradiction – the point is, what is a contradiction, and under which conditions contradictions arise? Contrasting data *per se* do not generate contradictions, since the agent is not yet committed to their propositional content. Hence data bases are expected to be typically inconsistent, without bothering in the least the agent. Being informed of p and $\sim p$, e.g. by being exposed to conflicting sources of information, is not a contradiction: only *believing at the same time both contrasting claims* would produce a contradiction in the agent mind and require a solution at the level of beliefs.

subjectively contradictory and the subject can entertain them as such before being obliged to choose, on the contrary Intentions – i.e. what one has decided to pursue and to do – must be subjectively non contradictory. In other words, conflicts between possible/candidate goals must have been solved at the deliberation stage. Exactly the same happens between data and beliefs: after the selection process, inconsistency cannot be tolerated anymore and contradictions have to be solved (Fig. 6). This parallel between the processing of epistemic and motivational representations yields a convincing picture of human mind as a *coherence-seeking device*.

Contradiction management is further discussed in [21]. Here we only want to emphasize that rational agents are not preserved from contradictions for some benevolent 'law of nature': they are rather equipped *to handle contradictions efficiently* both in the epistemic and motivational processing, e.g. exploiting the informational value of contrasting evidences and balancing conflicting desires. If we fail to acknowledge inconsistency in belief change, we miss the core of argumentation: weighting against each other contradictory claims.

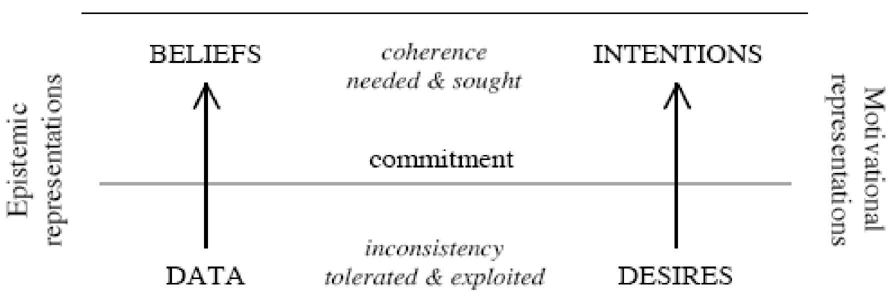

Fig. 6. The mind as a coherence-seeking device

3.5 Parameters and Argumentation

In DBR, parameters (cf. 2.2 and 2.4) provides a computational description of individual variation [21]. They also have consequences over the treatment of argumentation, capturing the relevant distinction between *local* and *global persuasion*, and the *multi-layered nature of argumentative strategies*.

An argument can either aims to change single beliefs in the mind of the audience (local persuasion), or it might address the basic processes which define the outcome of belief revision for that audience (global persuasion). Whenever persuasive argumentation is a major issue (e.g. political campaigns, advertising, religious events), global persuasion is the key feature: it is not enough to change some specific beliefs, the arguer is basically trying *to make the audience accept a different way of thinking* – i.e. different revision procedures, to be applied autonomously from now on.

Local and global strategies are grounded in our model, respectively, in *argumentation over data network* and *argumentation over parameters*. The examples discussed in 3.1-3.4 are instances of local persuasion, which attack or support nodes in the data structure. On the contrary, global persuasion questions the validity of individual parameters concerning belief revision, e.g. the selection process («You should not pay so much attention to explanatory power, otherwise you are prone to wishful thinking! »), the assessment of data values («Do not underestimate contrasting evidences, or you will be biased toward confirmation! »), the reliability assigned to new sources («Why do you trust so much somebody you does not know? ») [21].

Perhaps the most famous instance of the interplay between belief revision parameters, argumentation and global persuasion is from the Gospels: that is, the incredulity of St Thomas. When Jesus, after his resurrection, appeared for the first time to the apostles, Thomas was not there. Once he had been told of the miracle by his companions, he refused to believe in their account, claiming that "unless I see in his hands the print of the nails, and place my finger in the mark of the nails, and place my hand in his side, I will not believe" (St John, 20: 25). This bold statement was challenged when Jesus appeared again, and explicitly insisted that Thomas should probe Jesus' wounds with his incredulous finger. After that, the apostle was convinced and repentant, but Jesus was after a global persuasion, rather than a local one. Hence his final comment: "Have you believed because you have seen me? Blessed are those who have not seen and yet believe" (St John, 20: 29).

In this episode a whole attitude (skepticism) is stigmatized as inadequate within a given context (matters of faith)[7], and the misbehaving agent is required for the future to apply different parameters to his processes of belief selection and change. The positive counterpart of Thomas is exemplified by Mary Magdalene, who immediately believed in the resurrection of Jesus once she was told by him, although she was not able to distinguish his features and his voice. Nevertheless, the testimony of a stranger standing next to the sepulcher of Jesus was enough for her to believe in the miracle. Both these attitudes can be captured (in a simplified form) within the framework of DBR, as the computational analogous of Mary Magdalene and St Thomas summarized in Table 2 (for details on each parameter listed in the table, see [21]). In the MAS counterpart of the biblical episode, the argumentative strategy applied by Jesus would aim to make Thomas shift his parameters towards the ones of Mary, i.e. developing a more trustful epistemic attitude through several minor changes: e.g. a less pessimistic assessment of credibility value (the first two parameters), more refined processes to evaluate importance (the third, fourth and fifth parameter), a less realistic process of belief selection (the sixth, seventh and eighth parameter), and more reliance in new sources of information (the last parameter in Table 2).

---

[7] The episode serves also to illustrate *the relevance of context* in defining whether a given belief revision strategy is adaptive or not. In fact, Thomas' skepticism is inadequate only with respect to the situation portrayed here, i.e. the faith of a disciple towards his religious leader. In other contexts, the skeptical attitude embodied by Thomas would indeed constitute the only sensible strategy to use: this is exactly the reason why the mistreated Thomas became, starting from the Renaissance, an icon of scientific inquiry and human curiosity.

Table 2. Parameters in DBR and argumentation: Mary Magdalene vs. St Thomas

parameters	**Trustful (Mary Magdalene)**	**Skeptic (St Thomas)**
Credibility alg.	$c^{\alpha} = (1 - \Pi_{\sigma \in S\alpha}(1 - c^{\sigma})) \times \Pi_{\varepsilon \in K\alpha}(1 - c^{\varepsilon})$	$c^{\sigma} = pr^{\sigma} \times \Pi_{\varepsilon \in K\sigma}(1 - pr^{\varepsilon})$ with $\sigma \in S$ $c^{\alpha} = 1 - \Pi_{\sigma \in S\alpha}(1 - c^{\sigma})$ with $\alpha \notin S$
Union algorithm	$c^{\alpha\&\beta} = \min(c^{\alpha}, c^{\beta})$	$c^{\alpha\&\beta} = c^{\alpha} \times c^{\beta}$
Importance alg.	$\mu < 5,\ i^{\phi} = \mu/5 \times (1 - \Pi_{\psi \in N\phi}(1 - c^{\psi}))$ $\mu \geq 5,\ i^{\phi} = 1 - \Pi_{\psi \in N\phi}(1 - c^{\psi})$	$\mu < 5,\ i^{\phi} = \mu/5 \times (1 - \Pi_{\psi \in N\phi}(1 - c^{\psi}))$ $\mu \geq 5,\ i^{\phi} = 1 - \Pi_{\psi \in N\phi}(1 - c^{\psi})$
Depth λ	*2*	*1*
Consid. thres. w	*0.3*	*0.6*
Condition C	$c^{\phi} / (1 - i^{\phi})$	c^{ϕ}
Accept. thres. k	*0.6*	*0.8*
Function F	$c^{\phi} + i^{\phi} - (c^{\phi} \times i^{\phi})$	c^{ϕ}
Reliability default	*0.7*	*0.3*

Finally and more generally, it is worth noticing that parameters play a crucial role in any instance of argumentation, since the arguer is required to understand, at least partially, the parameters governing belief revision in his audience. This reflects the multi-layered nature of argumentation: for the arguer to be effective, it is not enough to figure out the audience's beliefs (the data structure and the resulting belief set), but also the way in which beliefs are processed (the audience's parameters on belief revision, e.g. how they assess data values, how they select beliefs from data, etc.). Factual evidences are useless, if the audience do not care for credibility in belief selection; on the other hand, alluring picture of highly desirable states of things does not work with matter-of-fact types – and so on. Formal models of belief change which fail to account for individual variation are implying that every audience will have identical reactions to the same base of data: an untenable assumption [5, 7, 21, 28].

4 Conclusions and Future Works

The integrated framework sketched here strongly supports a general methodological claim: a model of belief revision, in order to deal effectively with argumentation in MAS, must ensure *a proper degree of structural analysis* – i.e. it must emphasizes the relational properties which characterize epistemic processing, rather than its overall principles. Ordering criteria over propositions or sets, like in AGM-style approaches, are not expressive enough to model argumentation – nor belief revision.

Therefore, the main implication of this preliminary proposal is to initiate a systematic effort of integrating research areas necessarily connected with each other, i.e. argumentation studies and belief revisions, but that only rarely have been so far modeled within the same framework [10, 26]. Even more important, the DBR theory presented here constitutes a first step towards formal and computational models of epistemic change (both intra- and inter-agents) to express the complex socio-cognitive dynamics involved in belief revision in MAS, in contrast with the idealistic approach which dominated this field so far (see analogous considerations in [6, 21, 26, 27]).

In our future work we intend to refine the DBR model of belief revision (e.g. extending the computational treatment of data properties to motivational and emotional features, i.e. relevance and likeability [6, 11, 21]), to provide more systematic connections with argumentation theories [2, 4, 16, 17, 28, 30], especially within the MAS community [3, 19, 20, 22, 23], to explore the interaction between TMS-based belief revision and argumentation models [8, 18], and to move towards implementation in agent-based cognitive and social simulation (preliminary work in this direction is being developed within the AKIRA framework [24]), exploiting random parametrical variation to study evolutionary dynamics in belief revision and argumentation development. As a starting point, we plan to use argumentation tasks as testing ground for belief revision algorithms, and vice versa – building on the general results discussed here. Finally, we also aim to investigate the more radical idea of modeling the whole process of epistemic change as a form of *internal argumentation* [4, 21], as long ago suggested in developmental psychology by Jean Piaget [25] and Lev Vygotsky [29].

References

1. Alechina, N., Logan, B.: Ascribing Beliefs to Resource Bounded Agents. Proceedings of AAMAS'02 (2002)
2. Amgoud, L., Cayrol, C.: Inferring from Inconsistency in Preference-Based Argumentation Frameworks. Journal of Automated Reasoning 29 (2002) 125-169
3. Amgoud, L., Parsons, S., Maudet, N.: Arguments, Dialogue and Negotiation. Proceedings of ECAI 2000, Berlin (2000)
4. van Benthem, J., van Eemeren, F. H., Grootendorst, R., Veltman, F. (eds.): Logic and Argumentation. North-Holland, Amsterdam (1996)
5. Castelfranchi, C.: Guarantees for Autonomy in Cognitive Agent Architecture. Proceedings ECAI 1994 Workshop on Agent Theories, Architectures, and Languages, Springer, Berlin (1995) 56-70
6. Castelfranchi, C.: Reasons: Belief Support and Goal Dynamics". Mathware & Soft Computing 3 (1996) 233-247
7. De Rosis, F., Grasso, F., Castelfranchi, C., Poggi, I.: Modeling Conflict-Resolution Dialogues. J. H. M¸ller, R. Dieng (eds.): Computational Conflicts. Springer, Berlin (2000) 41-62
8. Doyle, J.: Reason Maintenance and Belief Revision: Foundations vs. Coherence Theories. P. G‰rdenfors (ed.): Belief Revision. Cambridge University Press, Cambridge (1992) 29-51
9. Dragoni, A. F., Giorgini, P.: Distributed Belief Revision. Autonomous Agents and Multi-Agent Systems 6 (2003) 115-143
10. Falappa, M. A., Kern-Isberner, G., Simari, G. R.: Explanations, Belief Revision and Defeasible Reasoning. Artificial Intelligence 141 (2002) 1-28
11. Frijda, N. H., Manstead, A., Bem, S. (eds.): Emotions and Beliefs: How Feelings Influence Thoughts. Cambridge University Press, Cambridge (2000)
12. Fullam, K.: An Expressive Belief Revision Framework Based on Information Valuation. MS thesis, University of Texas at Austin (2003)
13. 13.Galliers, J. R.: Autonomous Belief Revision and Communication. P. G‰rdenfors (ed.): Belief Revision. Cambridge University Press, Cambridge (1992) 220-246
14. 14.G‰rdenfors, P.: Knowledge in Flux: Modeling the Dynamics of Epistemic States. The MIT Press, Cambridge (1988)

15. Gillies, A. S.: Two More Dogmas of Belief Revision: Justification and Justified Belief Change. Working paper, Harvard University (2003) consulted on May 2004, available at: http://www.people.fas.Harvard.edu/~gillies/dogmas.pdf
16. 16.Grasso, F.: A Mental Model for a Rhetorical Arguer. Proceedings of EuroCogsci 03, LEA (2003)
17. Guerini, M., Stock, O., Zancanaro, M.: Persuasion Models for Intelligent Interfaces. Proceedings of IJCAI Workshop on Computational Models of Natural Argument (2003)
18. Huns, M. N., Bridgeland, D. M.: Multiagent Truth Maintenance. IEEE Transactions on Systems, Man, and Cybernetics 21 (1991) 1437-1445
19. Kakas, A., MoraÔtis, P.: Argumentation Based Decision Making for Autonomous Agents. Proceedings of AAMAS 2003, Melbourne (2003), 883-890
20. McBurney, P., Parsons, S., Wooldridge, M.: Desiderata for Agent Argumentation Protocols. Proceedings of AAMAS 2002, Bologna (2002)
21. Paglieri, F.: Data-oriented Belief Revision: Towards a Unified Theory of Epistemic Processing. Proceeding of STAIRS 2004, Valencia, IOS Press (2004)
22. Parsons, S., Petterson, O., Saffiotti, A., Wooldridge, M.: Intention Reconsideration in Theory and Practice. Proceedings of ECAI 2000, Berlin (2000)
23. Parsons, S., Wooldridge, M., Amgoud, L.: Properties and Complexity of Some Formal Inter-agent Dialogues. Journal of Logic and Computation 13 (2003) 347-376
24. Pezzulo, G., Calvi, G.: AKIRA: a Framework for Multi-Agent Based Simulation. Proceedings of MAMABS 2004 (2004)
25. Piaget, J.: The Language and Thought of the Child. Kegan Paul, Trench, Trubner (1926)
26. 26.Pollock, J. L., Gillies, A. S.: Belief Revision and Epistemology. Synthese 122 (2000) 69-92
27. Segerberg, K.: Two Traditions in the Logic of Belief: Bringing Them Together. H. J. Ohlbach, U. Reyle (eds.): Logic, Language and Reasoning. Kluwer, Dordrecht (1999) 135-147
28. 28.Toulmin, S.: The Uses of Argument. Cambridge University Press, Cambridge (updated edition 2003, orig. ed. 1958)
29. Vygotsky, L. S.: Thought and Language. The MIT Press, Cambridge MA (1962)
30. Walton, D. N.: Argumentation Schemes for Presumptive Reasoning. LEA, Mahwah (1996)
31. 31.Wassermann, R.: Resource-Bounded Belief Revision. ILLC dissertation series DS-2000-01, Amsterdam (2000)