Available online at www.sciencedirect.com
ScienceDirect
Artificial Intelligence 171 (2007) 855–874
Artificial Intelligence
www.elsevier.com/locate/artint

Practical reasoning as presumptive argumentation using action based alternating transition systems

Katie Atkinson *, Trevor Bench-Capon

Department of Computer Science, University of Liverpool, Liverpool, L69 3BX, UK

Received 8 November 2006; received in revised form 4 April 2007; accepted 16 April 2007

Available online 29 April 2007

Abstract

In this paper we describe an approach to practical reasoning, reasoning about what it is best for a particular agent to do in a given situation, based on presumptive justifications of action through the instantiation of an argument scheme, which is then subject to examination through a series of critical questions. We identify three particular aspects of practical reasoning which distinguish it from theoretical reasoning. We next provide an argument scheme and an associated set of critical questions which is able to capture these features. In order that both the argument scheme and the critical questions can be given precise interpretations we use the semantic structure of an Action-Based Alternating Transition System as the basis for their definition. We then work through a detailed example to show how this approach to practical reasoning can be applied to a problem solving situation, and briefly describe some other previous applications of the general approach. In a second example we relate our account to the social laws paradigm for co-ordinating multi-agent systems. The contribution of the paper is to provide firm foundations for an approach to practical reasoning based on presumptive argument in terms of a well-known model for representing the effects of actions of a group of agents.

Keywords: Argumentation; Practical reasoning; Value based argumentation; Argument schemes; AATS

1. Introduction

Although many of the arguments that are deployed in everyday life are concerned with what it is sensible or practical to do, the topic of practical reasoning has been rather neglected by philosophers. Practical reasoning has, of course, been addressed (see, e.g. [19] for a collection of essays and [20] for a monograph), but it has received nothing like the attention that has been paid to reasoning about beliefs. When action has been considered, it has most often been in the context of ethics, considering what is morally right or wrong, rather than what is prudentially or practically useful. Until recently it has been similarly neglected in computer science, where practical reasoning has traditionally been treated as little different from deduction, standard backward chaining techniques being applied to rules with goals as consequents and pre-conditions and actions as antecedents, to form plans according to some variety of means-end analysis.

* Corresponding author.
E-mail addresses: katie@csc.liv.ac.uk (K. Atkinson), tbc@csc.liv.ac.uk (T. Bench-Capon).

0004-3702/$ – see front matter
doi:10.1016/j.artint.2007.04.009

In Section 2 we discuss some of the differences between reasoning about belief and reasoning about action which cause problems with approaches based on the practical syllogism. In Section 3 we discuss the approach to practical reasoning in [22] which makes use of an argumentation scheme and associated critical questions. We extend this scheme and identify additional critical questions. Section 4 introduces the notation of Action Based Alternating Transition Systems (AATSs) to ground the approach. Section 5 presents definitions of the scheme and questions in terms of the AATS. Section 6 gives a detailed example which illustrates the use of the argument scheme and provides instances of each of the critical questions. Section 7 relates this work to the social laws approach to co-ordinating multi-agent systems. Finally, we offer some concluding remarks in Section 8.

2. Practical reasoning

Practical reasoning in computer science can predominately be seen as based on a form of the practical syllogism. An example of the practical syllogism taken from [15] is:

K1 I'm to be in London at 4.15
If I catch the 2.30, I'll be in London at 4.15
So, I'll catch the 2.30.

This is the basis for means-end reasoning in knowledge-based planning systems: given the goal in the first premise and the rule in the second, the antecedent of the rule becomes the new goal. Considered as an argument justifying the action, however, this cannot be quite right. It may well be possible to accept both the premises and deny the conclusion. There are at least three bases for criticism:

C1 K1 represents a species of abduction, and so there may be alternative ways of achieving the goal.
C2 Performing an action typically excludes the performance of other actions, which might have other desirable results; these may be more desirable than the stated goal.
C3 Performing an action typically has a number of consequences. If some of these are undesirable, they may be sufficiently bad to lead us to abandon the goal.

While C1 has inspired much work in AI, C2 and C3 have received less attention, the focus in AI having been on making the argument sound through the exclusion of alternatives rather than the provision of justifications in terms of the desirability or otherwise of the state of affairs brought about for the particular recipient of the argument. In planning systems, for example, the desired goal is typically given rather than chosen: choice being reserved for subgoals.

In order to act on the basis of an argument such as K1 therefore, we need to consider alternative actions, alternative goals and any additional consequences, and then choose the best of these alternatives. Note the element of choice: we can choose our goals and actions in a way in which we cannot choose our beliefs, and *different people* may rationally make *different choices*. As Searle puts it:

> "Assume universally valid and accepted standards of rationality, assume perfectly rational agents operating with perfect information, and you will find that rational disagreement will still occur; because, for example, the rational agents are likely to have different and inconsistent values and interests, each of which may be rationally acceptable" [20, xv].

In a sense therefore any practical argument is directed to a specific person at a specific time, to encourage them towards a particular choice (or to justify them in making a particular choice) and the objectivity that we can find in factual matters cannot in general be attained in practical reasoning. An attempt to modify K1, similar to one put forward by Searle in [20] (although not regarded by him as satisfactory), is:

S1 I want, all things considered, to achieve E
The best way, all things considered, to achieve E is to do M
So, I will do M.

There are problems with this: we cannot in general consider all things, because we have limited reasoning resources and imperfect information. Nor is it easy to say what is meant by "best" here. In computer science there are often attempts to define best using some kind of utility function but, in practise, the preference used by an agent to select an action ordering is often the *product* of practical reasoning than an input to it. Coming to understand what we think is best *for us* in the particular situation in which we find ourselves is part of what we do in practical reasoning. We will discuss this further below.

One way of addressing these problems is to regard practical reasoning as a species of presumptive argument. Given an argument like K1, we have a presumptive reason for performing the action. This presumption can, however, be challenged and withdrawn. Subjecting our argument to appropriate challenges is how we hope to identify and consider the alternatives that require consideration, and determine the best choice for us, in the particular context.

One account of presumptive reasoning is in terms of argument schemes and critical questions, as given in [22]. The idea here is that an instantiation of an appropriate argument scheme gives a presumption in favour of its conclusion. Whether this presumption stands or falls depends on satisfactory answers being given to the critical questions associated with the scheme. We present such a scheme in Section 3.

When applying practical reasoning to the design of autonomous computer systems we need some means by which these variations in preferences can be expressed. In agent systems this has usually been done through the economic theoretic notion of utility functions ascribed to states. Utility functions represent the desirability of states, and agents act in order to try and maximise the perceived utility they expect to get from executing actions that lead to the most desirable states. However, this economic notion forces the assignment of rankings over states of affairs in a manner that seems counter-intuitive to the nature of practical reasoning. Returning to the work of Searle in [20], we can see that he holds the view that practical reasoning in humans does not make use of pre-existing utility functions:

> "This answer, [that an audience can provide a ranking for goals] though acceptable as far as it goes, mistakenly implies that the preferences are given prior to practical reasoning, whereas, it seems to me, they are typically the product of practical reasoning. And since ordered preferences are typically products of practical reason, they cannot be treated as its universal presupposition" [20, p. 253].

Thus, according to Searle, any theory of practical reasoning must take into account that choices concerned with the selection of actions should be made during the reasoning process and not form an input to it. The account given in this paper aims at satisfying this criterion. The justification for action that we present makes use of *values* and shows how preferences based upon individual values emerge through the practical reasoning process. Values, as we use the term, denote some actual descriptive social attitude/interest which an agent may or may not wish to uphold or subscribe to and they provide an actual subjective reason for wanting to bring about a particular state.[1] In this sense values are not some quantitative measure of states, but they provide qualitative, subjective reasons as to why particular individuals should wish to reach or avoid certain states. In some cases, however, we may wish to add more quantitative measures to indicate the degree to which a particular value is promoted. Note that this is *not* a ranking on values, but a way of comparing alternatives based on the *same* value. In the next section we will show how we use values in argumentation for practical reasoning.

3. Argument schemes and critical questions

An argument scheme and associated critical questions was presented in [3] which enables agents to propose, attack and defend justifications for action. This argument scheme followed Walton [22] in viewing reasoning about action (practical reasoning) as presumptive justification—*prima facie* justifications of actions can be presented as instantiations of an appropriate argument scheme, and then critical questions characteristic of the scheme used can be posed to challenge these justifications. The argument scheme developed in [3] is an extension of Walton's *sufficient condition scheme for practical reasoning* [22] and is stated as follows:

[1] Values have nothing whatsoever to do with quantitative measures. This conforms with current English usage: for example, US and UK politicians continually appeal to their values to persuade people to adopt their policies.

AS1 In the current circumstances R
We should perform action A
Which will result in new circumstances S
Which will realise goal G
Which will promote some value V.

In this scheme we have disambiguated Walton's notion of a goal by separating it into three elements: the state of affairs brought about by the action; the goal (the desired features in that state of affairs); and the value (the reason why those features are desirable). For example, I may diet to lose weight, with the goal of not being overweight, to promote the value of health. Our underlying idea in making this distinction is that the agent performs an action to move from one state of affairs to another. The new state of affairs may have many differences from the current state of affairs, and it may be that only some of these differences are desired by the agent. The significance of these differences is that they make the new state of affairs better with respect to some good valued by the agent.

Now an agent who does not accept this presumptive argument may challenge the contentious elements in the instantiation through the application of critical questions. We have elaborated Walton's original four critical questions associated with his scheme by extending them to address the different elements identified in the goal in our new argument scheme. The extension in [3] resulted in sixteen different critical questions. An unfavourable[2] answer to a critical question will identify a potential flaw in the argument. By putting forward arguments giving a reason to believe that there may be an unfavourable answer, therefore, agents can attack the validity of the various elements of the argument scheme and the connections between them, suggest alternative possible actions, and draw attention to side effects of the proposed action. The critical questions thus identify the ways in which the presumption may be challenged, and arguments grounding negative answers to them can be seen as attacks on the original argument. The critical questions are as follows:

CQ1: Are the believed circumstances true?
CQ2: Assuming the circumstances, does the action have the stated consequences?
CQ3: Assuming the circumstances and that the action has the stated consequences, will the action bring about the desired goal?
CQ4: Does the goal realise the value stated?
CQ5: Are there alternative ways of realising the same consequences?
CQ6: Are there alternative ways of realising the same goal?
CQ7: Are there alternative ways of promoting the same value?
CQ8: Does doing the action have a side effect which demotes the value?
CQ9: Does doing the action have a side effect which demotes some other value?
CQ10: Does doing the action promote some other value?
CQ11: Does doing the action preclude some other action which would promote some other value?
CQ12: Are the circumstances as described possible?
CQ13: Is the action possible?
CQ14: Are the consequences as described possible?
CQ15: Can the desired goal be realised?
CQ16: Is the value indeed a legitimate value?

As described in [3], each of the above critical questions represents a source of disagreement about a particular element of an instantiation of the argument scheme AS1. In arguing for an unfavourable answer to a critical question an opponent is making an 'attack' on the element of the position in question and this attack may be stated with varying degrees of force. This leads to a number of variants that can be associated with particular critical questions. For example, with CQ1 an attacker may simply disagree with the description of the circumstances, i.e., arguing that R is not the current state of the world. Beyond this minimalist attack, an attacker may also state an alternative position to that proposed, for example, expressing not only that R is not the current state of the world, but also that T is the

[2] By "unfavourable" we mean an answer that casts doubt on the presumption. This answer may in fact be 'no', as in e.g. CQ1, or 'yes', for e.g. CQ7.

real current state. A full list and description of the attacks and their variants is given in [3] and although we will not formalise all variants of attacks in this paper, we will provide examples to illustrate the point.

In addition to the list of critical questions given above, there is one additional question when using a representation that models joint actions of agents. In Section 4 we will use Action-based Alternating Transition Systems which allow for the representation of joint actions whereby two or more agents each execute some component of the action. This means that the choice of one agent may not determine which joint action is performed: the other agent(s) may make choices which lead to other joint actions. This introduces the need for a further critical question: is the other agent guaranteed to execute its part of the desired joint action? In [3] this objection was subsumed under CQ2 since that question also claims that the action will lead to a different state from that proposed in the argument. There is, however, a point in distinguishing cases where the different state is reached through the choice of another agent, since the reasons why the desired state is not reached differ. For CQ2 the transition relied on does not exist, whereas for CQ17 it exists, but ensuring that it is followed is not within the sole control of the agent concerned. This means that we may answer this critical question by explaining why it is in the interests of the other agent to participate in this joint action. This distinction is used extensively in, for example, [13]. We therefore add this additional critical question as CQ17.

Now, given that we can test the presumptions present in instantiations of the argument scheme by posing the critical questions, we also need to consider the responses that can be given to the critical questions. There are essentially two options here: the response can either be an acceptance of the challenge, or a rebuttal to it. If the challenge is accepted the particular instantiation of the argument scheme no longer holds and the action is therefore no longer justifiable. However, if a rebuttal to a critical question is to be made, the nature of rebuttal will depend upon the particular critical question that is being rebutted. Certain critical questions invite alternative arguments which are themselves of the form of AS1, i.e., CQ5–CQ11, and so attacking arguments giving negative answers to these critical questions themselves form instantiations of the argument scheme and so are themselves subject to the same set of critical questions. Arguments based on the remaining critical questions do not pose alternatives as they simply attack a particular element of the argument scheme. Arguments posing these critical questions can be further sub-divided into two categories: those that dispute facts and those that dispute value preferences. Firstly, for those that dispute matters of facts (e.g., whether or not the circumstances are as described), if the proponent of the argument presented in the instantiation of AS1 refuses to accept this attack, then resolution of the issue must rely on some method to establish exactly what the facts are. Secondly, for attacks involving subjective elements of the scheme (e.g., whether or not a goal promotes a particular value), resolution may not be possible because value subscriptions and orderings are subjective and agents must be allowed to make different choices to reflect their different interests and aspirations. Thus, disagreements concerning values may not be capable of resolution and in such a case the participants have reached the point of rational disagreement. In order to try and resolve such disagreements over value orderings, reasoning at a meta level would be required. One approach to enabling participants to perform such reasoning has recently been described in [16].

One further point regarding the argument scheme and critical questions is that AS1 can also be stated in negative versions. CQ8 and CQ9 speak of a value being demoted. Where a value is demoted we have a reason to *refrain* from an action to avoid a state containing particular features (a "goal") which demote a value. Additionally, a further negative variant, used to pose CQ11, can be stated: given a particular set of circumstances, an action which leads to a particular state of affairs that entails the negation of some goal (i.e., $\neg$G) should not be performed, because it would prevent the promotion of the value associated with that goal. These negative versions of AS1 can thus be used to argue in terms of avoiding some undesirable outcome rather than achieving some positive outcome. The importance of arguments for avoiding undesirable outcomes has received emphasis from Amgoud and her colleagues (e.g. [2]), who use the term *bipolar* goals to distinguish goals to be pursued from "goals" to be avoided. We relate this bipolarity not to goals (in some cases a goal which demotes one value may promote another, and so goals need not be intrinsically negative or positive[3]), but to demotion or promotion of particular *values*. We use these negative variants of the argument scheme to generate arguments and critical questions, but once generated make no distinction between arguments based on demotion rather than promotion of values. These negative variants have also been used in [12].

[3] This is actually quite common: we speak of "swings and roundabouts" with regard to many possible goals. Owning a Porsche is a goal which promotes status and hedonism, but demotes expense and harms the environment. Where the balance lies depends on the aspirations and values of the individual.

The usefulness of the approach of [3] for the analysis of practical reasoning situations has been shown for a range of areas including law [10], e-democracy [7] and reasoning about the morally correct course of action [4,13]. Thus far, however, the generation of arguments and critical questions has been rather *ad hoc*. In order to provide rigour to the approach, it is necessary to make explicit the relation between the arguments and questions and some well founded underlying formal model. We will provide the required grounding in terms of an AATS, in the next section.

4. Action-based alternating transition systems

Much previous work on practical reasoning in agent systems has relied either on some particular agent architecture (e.g. [3]), or on some small set of rules expressing the consequences of actions, and the conditions under which a goal will be desired (e.g. [12] and [18]). The result has been that the modelling of the domain has been rather *ad hoc* or superficial. In particular the rules linking situations to desires do not fit well with our account of practical reasoning, where goals are desired in virtue of promoting values. Moreover, these approaches have focused on the decision of a single agent, and take no account of the effects of the choices of others on the outcome of the agent's actions. Here we want to recognise the importance of the fact that agents act in the context of a society of agents, and their choices must be made in the light of what these other agents may choose.

In order to be able to reason rigorously about actions and their effects, we need a well-defined structure in which we can represent how the actions of an agent will lead to transitions from one state to another. In particular we need to be able to contextualise these transitions so that the effects of actions can be made dependent on the action of other agents, and other events in the environment. One such structure is provided by Alternating Transition Systems (ATS), originally developed to underpin the Alternating-time Temporal Logic of [1]. These structures have also been used by van der Hoek et al. [21] to explore the social laws paradigm for describing coordination in multi-agent systems introduced largely through the work of Shoham, Tennenholtz and Moses (e.g. [17]). Like [21] we give the notions of actions and their pre-conditions a central role, so we adopt their version of ATS in which actions and pre-conditions are first class entities. This version is called an *Action Based Alternating Transition Systems* (AATS) in [21], and it is this structure that we will use in this paper.

Assume first that the systems we wish to model may be in any of a finite set Q of possible *states*, with some $q_0 \in Q$ designated as the *initial state*. Systems contain a set Ag of *agents*; and each agent $i \in Ag$ is associated with a set Ac_i of possible actions. It is assumed that these sets of actions are pairwise disjoint (i.e., actions are unique to agents).

A joint action j_C for set of agents, termed *a coalition*, C, is a tuple $\langle \alpha_1, \ldots, \alpha_k \rangle$, where for each α_j (where $j \leqslant k$) there is some $i \in C$ such that $\alpha_j \in Ac_i$. Moreover, there are no two different actions α_j and $\alpha_{j'}$ in j_C that belong to the same Ac_i. The set of all joint actions for coalition C is denoted by J_C, so $J_C = \prod_{i \in C} Ac_i$. Given an element j of J_C and an agent $i \in C$, i's action in j is denoted by j_i.

Thus, an *Action-based Alternating Transition System* (AATS) is an $(n+7)$-tuple $S = \langle Q, q_0, Ag, Ac_1, \ldots, Ac_n, \rho, \tau, \Phi, \pi \rangle$, where:

- Q is a finite, non-empty set of *states*;
- $q_0 \in Q$ is the *initial state*;
- $Ag = \{1, \ldots, n\}$ is a finite, non-empty set of *agents*;
- Ac_i is a finite, non-empty set of actions, for each $i \in Ag$ where $Ac_i \cap Ac_j = \emptyset$ for all $i \neq j \in Ag$;
- $\rho : Ac_{Ag} \to 2^Q$ is an *action pre-condition function*, which for each action $\alpha \in Ac_{Ag}$ defines the set of states $\rho(\alpha)$ from which α may be executed;
- $\tau : Q \times J_{Ag} \to Q$ is a partial *system transition function*, which defines the state $\tau(q, j)$ that would result by the performance of j from state q—note that, as this function is partial, not all joint actions are possible in all states (cf. the pre-condition function above);
- Φ is a finite, non-empty set of *atomic propositions*; and
- $\pi : Q \to 2^\Phi$ is an interpretation function, which gives the set of primitive propositions satisfied in each state: if $p \in \pi(q)$, then this means that the propositional variable p is satisfied (equivalently, true) in state q.

In addition to the elements of an AATS given in [21], we need to provide an extension to enable the representation of values from the underlying theory of practical reasoning. Firstly, we have a set Av of values for each agent (which are a subset of a set V of values). Every transition between two states from the set Q is either promoted, demoted,

or is neutral, with respect to each value. Note that values are not unique to agents: individual agents may or may not have values in common. Whether a value is promoted or demoted by a given action will be determined by comparing the state reached with the state left. More formal definitions of these elements are given below:

- Av_i is a finite, non-empty set of values $Av_i \subseteq V$, for each $i \in Ag$.
- $\delta : Q \times Q \times Av_{Ag} \rightarrow \{+, -, =\}$ is a *valuation function* which defines the status (promoted $(+)$, demoted $(-)$ or neutral $(=)$) of a value $v_u \in Av_{Ag}$ ascribed by the agent to the transition between two states: $\delta(q_x, q_y, v_u)$ labels the transition between q_x and q_y with one of $\{+, -, =\}$ with respect to the value $v_u \in Av_{Ag}$.

We can now extend the original specification of an AATS to accommodate the notion of values and re-define an AATS as a $(2n + 8)$-tuple $S = \langle Q, q_0, Ag, Ac_1, \ldots, Ac_n, Av_1, \ldots, Av_n, \rho, \tau, \Phi, \pi, \delta \rangle$.

5. Formalising the argument scheme and critical questions

In this section we will map the argument scheme and critical questions, identified in Section 3, onto the AATS structures introduced in the previous section. We can view practical reasoning as involving three stages, as follows:

- *problem formulation*: deciding what the propositions and values relevant to the particular situation are, and constructing the AATS.
- *epistemic reasoning*: determining the initial state in the structure formed at the previous stage.
- *choice of action*: developing the appropriate arguments and counter arguments, in terms of applications of the argument scheme and critical questions, and determining the status of the arguments with respect to other arguments and the value orderings.

These stages may be carried out sequentially, or they may iterate if the critical questioning leads to a reformulation of the problem.

We can now re-state argument scheme AS1 in terms of the extended AATS given in Section 4. This gives us:

AS2 The initial state $q_0 = q_x \in Q$,
Agent $i \in Ag$ should participate in joint action $j_n \in J_{Ag}$ where $j_n^i = \alpha_i$,
Such that $\tau(q_x, j_n)$ is q_y,
Such that $p_a \in \pi(q_y)$ and $p_a \notin \pi(q_x)$, or $p_a \notin \pi(q_y)$ and $p_a \in \pi(q_x)$,
Such that for some $v_u \in Av_i$, $\delta(q_x, q_y, v_u)$ is $+$.

Furthermore, we can now provide a precise definition of what it is to have an argument for an unfavourable answer for each of the critical questions associated with an instantiation of AS1, in terms of our extended definition of an AATS. However, before we present these descriptions, we will first make some remarks about the different categories under which each of the critical questions falls and state how these relate to the three stages of the practical reasoning process that we identified at the start of this section.

Firstly, eight of the critical questions uncover differences between agents, and the way in which they have represented the problem. CQ2 accepts that the pre-conditions for the performance of the action hold, but disputes the resultant state attained through execution of the action. CQ3 does not dispute the state reached through execution of the action but instead disputes the truth of some proposition (the goal) within this state. CQ4 reveals a difference between how individual agents value states of affairs; CQ12, CQ13, CQ14 and CQ15 all represent differences in language between individual agents (resulting, for example, in dispute over descriptions of the state of the world); and, CQ16 represents disagreement as to what counts as a value. Resolution of these eight critical questions falls under the remit of the *problem formulation* stage of the practical reasoning process.

In the *epistemic reasoning* stage of the process the agent determines which state of affairs it finds itself in within the structure, and so CQ1 is asked at this stage. CQ1 relates to whether the pre-conditions required for the performance of the desired action are satisfied. The other element of imperfect knowledge relates to the behaviour of other agents in the scenario. This is the concern of CQ17, which asks whether or not the other agent(s) will execute their part of the desired joint action. This part is epistemic since the uncertainty as to which state is reached derives from a lack of

knowledge of the other agent's choices, rather than from the consequences of particular actions, which are resolved in problem formulation.

The remaining critical questions comprise CQ5–CQ11 and these are all pertinent to the final stage of the practical reasoning process, the *choice of action* stage. CQ5, CQ6 and CQ7 all consider possible alternatives to the original action proposed with each of these critical questions considering the effects of any such alternative actions upon the consequences, goal and value, respectively. CQ8, CQ9 and CQ10 are all concerned with the side effects of the proposed action. CQ8 and CQ9 draw attention to possible negative side effects, whilst CQ10 can be seen as more of a supporting argument that identifies positive side effects of the action that endorse rather than dispute the performance of the action, and so CQ10 questions the *justification* for the action rather than the *performance* of the action itself. There now remains only one critical question, CQ11, which identifies a clash between the action proposed and some other desirable action. CQ11 arises when the goal state achieved by the proposed action is incompatible with the goal state of some other action that promotes a desirable value, so that only one of the actions can be executed. All of the critical questions discussed in this final stage propose alternative arguments of one type or another and so they can be viewed as arguments that are to be compared in order to choose an action to perform.

Given the above stages under which each of the critical questions falls, we now present the formal definitions of all the critical questions, grouped according to these categories. In our definitions below we use the subscripts in the instantiation given in AS2 above: thus q_x is the proposed initial state, q_y the state resulting from the performance of the action, etc. We begin with the critical questions from the *problem formulation* stage of the practical reasoning process:

CQ2: $\tau(q_x, j_n)$ is not q_y.
CQ3: $p_a \notin \pi(q_y)$.
CQ4: $\delta(q_x, q_y, v_u)$ is not $+$.
CQ12: $q_x \notin Q$.
CQ13: $j_n \notin J_{Ag}$.
CQ14: $\tau(q_x, j_n) \notin Q$.
CQ15: $p_a \notin \pi(q)$ for any $q \in Q$.
CQ16: $v_u \notin V$.

Next we define negative answers to the critical questions relevant to the second stage of the process, the *epistemic reasoning*:

CQ1: $q_0 \neq q_x$ and $q_0 \notin \rho(\alpha_i)$.
CQ17: $j_n^i = j_m^i$, $j_n \neq j_m$ and $\tau(q_x, j_n) \neq \tau(q_x, j_m)$.

Finally, we define negative answers to the critical questions pertinent to the choice of action stage of the process:

CQ5: Agent $i \in Ag$ can participate in joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_y.
CQ6: Agent $i \in Ag$ can participate in joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_y, such that $p_a \in \pi(q_y)$ and $p_a \notin \pi(q_x)$ or $p_a \notin \pi(q_y)$ and $p_a \in \pi(q_x)$.
CQ7: Agent $i \in Ag$ can participate in joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_z, such that $\delta(q_x, q_z, v_u)$ is $+$.
CQ8: In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y, such that $p_b \in \pi(q_y)$, where $p_a \neq p_b$, such that $\delta(q_x, q_y, v_u)$ is $-$.
CQ9: In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y, such that $\delta(q_x, q_y, v_w)$ is $-$, where $v_u \neq v_w$.
CQ10: In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y, such that $\delta(q_x, q_y, v_w)$ is $+$, where $v_u \neq v_w$.
CQ11: In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y and $\delta(q_x, q_y, v_u)$ is $+$. But, there is some other joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_z, such that $\delta(q_x, q_z, v_w)$ is $+$, where $v_u \neq v_w$.

Before we conclude our definition of the critical questions, we return briefly to the issue of variants of the critical questions, as discussed in Section 3. Recall that variants arise from the strength of the argument put forward: simply denying that an element of AS1 is as stated by a proponent is a weaker form of argument than the additional proposal of an alternative to the element denied. To illustrate this point we now provide a couple of definitions of such variant arguments. Consider first CQ2. The definition given above shows the minimalist attack that can be made by simply denying that the consequences of the action entail the goal state. If the purveyor of the attack were to offer a stronger attack, making use of CQ2 to both deny the consequences are as stated and suggest they are otherwise, then this is defined as follows:

CQ2b: $\tau(q_x, j_n)$ is not q_y and $\tau(q_x, j_n)$ is q_z.

For a second example, consider CQ4. Its minimalist attack states that a value is not promoted by a transition between two states. A variant on CQ4 would be the statement that not only does the transition not promote the value, but it actually demotes it:

CQ4b: $\delta(q_x, q_y, v_u)$ is not $+$ and $\delta(q_x, q_y, v_u)$ is $-$.

The above examples give a flavour of how variants of the critical questions can affect the strength of the attack put forward and although it would be a straightforward task to fully specify the full set of variants associated with the appropriate critical questions, we will not do so here. These variants become important when the reasoning is included in a dialogue framework (see e.g. [3]) since the choice of variant will have strategic implications, and affect the commitments of the dialogue participants.

This concludes our definition of negative answers to the critical questions associated with the argument scheme. In the next section we provide a detailed example to demonstrate how these definitions can be used in the context of a specific practical problem.

6. Example application

In this section we will apply our approach to a practical reasoning problem. We will use a classic AI problem, traditionally used to illustrate search problems, but here addressed using practical argumentation. We will consider the reasoning which the farmer, the main actor in the scenario, could use at each step of his problem in order to come to a view as to the best next step to take. In this way the problem will be solved without the need for look ahead, and produce satisfying justifications for each choice of action. Although simple, the example allows us to exercise the complete set of critical questions.

The situation is that a farmer is returning from market with a chicken (C), a bag of seeds (S) and his faithful dog (D). He needs to cross a river, and there is a boat (B) but it can only carry the farmer and one of his possessions. He cannot leave the chicken and seeds together because the chicken will eat the seeds. Similarly he cannot leave the dog and the chicken unattended together. His problem is how to organise his crossing.

We will represent the states by two lists, one for the items on the right bank, and one for items on the left. Thus [BCDS, _] will be selected from Q as the initial state, q_0. We will not list all the states at this point, but number them as they occur in our example.

We have ten joint actions:

j_0: All do nothing
j_1: Farmer rows alone, animals do nothing
j_2: Farmer rows alone, chicken eats seeds
j_3: Farmer rows alone, dog eats chicken
j_4: Farmer rows alone, chicken eats seeds, dog eats chicken
j_5: Farmer rows seeds, animals do nothing
j_6: Farmer rows seeds, dog eats chicken
j_7: Farmer rows dog, animals do nothing
j_8: Farmer rows dog, chicken eats seeds

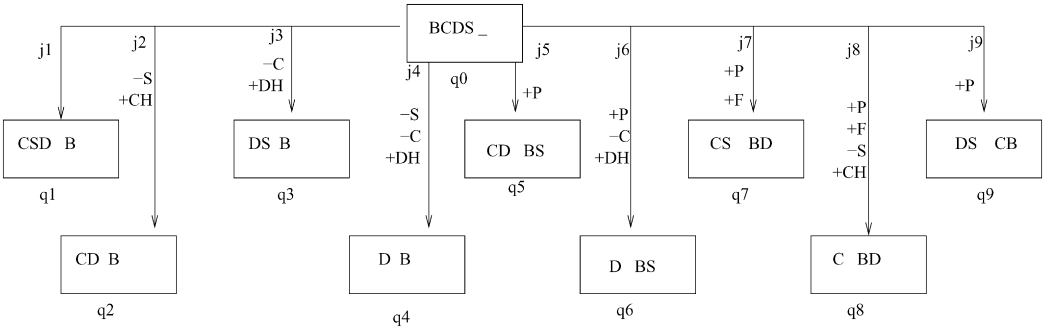

Fig. 1. Transitions from q_0.

j_9: Farmer rows chicken, animals do nothing
j_{10}: All continue their journey home.

We can also identify a number of possible values:[4]

P: Progress—promoted when farmer moves one of his possessions to the right side of the river, demoted when farmer moves one of his possessions to the left side of the river (through the goal of transportation), and demoted when a state is revisited (through the always undesirable "goal" of repetition).
S: Farmer has seeds—demoted when farmer loses seeds.
C: Farmer has chicken—demoted when farmer loses chicken.
F: Friendship— promoted when farmer travels with dog (it was for this companionship that he brought the dog with him).
DH: Dog is happy—promoted when it eats the chicken.
CH: Chicken is happy—promoted when it eats the seeds.

With these actions and values we can represent the transitions from the initial state of the problem as in Fig. 1.

In working through the example we will form a number of arguments and identify critical questions. We will not list all the critical questions that could be made against the arguments, but will give at least one example of each of the seventeen critical questions in the course of the example. To aid readability we will sometimes, where it is not relevant for the critical questions, omit the goal. The critical questions will also be expressed informally to aid readability, but each will be associated with the CQ number, allowing it to be related to the more precise descriptions given in Section 3 and the formal definitions given in Section 5.

Before we look at the arguments that an agent can generate on the basis of Fig. 1, we should note that several critical questions could be posed against arguments put forward by another agent using a different formulation of the problem. For example, if the other agent was to suggest that the farmer should take all his possessions together across a bridge, that would be open to the critical question that there was no such action (CQ13). If another agent was to refer to the dog being muzzled in the initial state or in a consequent state, that would be open to the objection that the representation contains no such proposition (CQ12 and CQ14 respectively). If the agent was to claim that one of the joint actions reached a state q_n where $n > 9$, it could be denied that this was the correct consequence of that action (CQ2). Finally if the other agent were to argue on the basis of an initial state other than [BCDS, _], CQ1 would apply.

We can now turn to the positive arguments that an agent can produce on the basis of Fig. 1. It can make a number of instantiations of AS1: one for each value promoted.

Arg1: Farmer should row alone, to reach q_2, in which the chicken is fed, which promotes CH.

This is subject to a number of critical questions, including:

[4] Some labels for values are the same as the propositions used in the state description. The context makes it clear which is intended.

Obj1a: The chicken won't eat the seeds and so CH won't be promoted (CQ17).
Obj1b: In q_2 the seeds are eaten and so S is demoted (CQ9)
Obj1c: The happiness of the chicken, CH, is not a value (CQ16)

Probably Obj1a will be rejected, since chickens lack restraint, but Obj1c is probably acceptable (farmers are unsentimental with respect to chickens), and Obj1b will be decisive (even farmers who care about their chickens care more about their seeds).

Arg2: Farmer should row alone, to reach q_3, in which the dog is fed, which promotes DH.

Again we can pose critical questions including:

Obj2a: The dog won't eat the chicken, but the chicken will eat the seeds reaching q_2 and demoting S (CQ17, CQ9).
Obj2b: The chicken will eat the seeds before the dog eats the chicken, reaching q_4 and demoting both S and C (CQ17, CQ9).

Against Obj2b it can be objected:

Obj2bi: Reaching q_4 promotes DH. So, even if, as Obj2b argues, q_4 is reached, the original value DH is still promoted.

Obj2c: Eating the chicken does not promote DH (perhaps because dogs often choke on chicken bones). (CQ4).
Obj2d: Reaching q_3 demotes C (CQ9).
Obj2e: The happiness of the dog, DH, is not a value (CQ16)
Obj2f: Rowing alone precludes rowing with the seeds which would put the seeds on the right bank and so promote P (CQ11) (note similar objections can also be raised for rowing with the dog and with the chicken).

Again we have enough here to reject Arg2: we could prefer C (Obj2d) or P (Obj2f) to DH (which is probable) or we can reject DH as a value altogether (Obj2e).

Moving on through the diagram,

Arg3: Farmer should row alone to reach q_4 promoting DH.

This is open to all the objections levelled against Arg2, except Obj2b, and the additional objection:

Obj3a: Reaching q_4 demotes S (CQ9).

We will thus reject Arg3. We now consider a different action for the farmer:

Arg4: Farmer should row the seeds to reach q_5, putting the seeds on the right bank which promotes P.

But this relies on the dog not molesting the chicken and so we get:

Obj4a: The dog will eat the chicken, reaching q_6 (CQ17).

The farmer may deny this, saying that he trusts the dog not to eat the chicken, but there are more alternative actions:

Obj4b: The farmer should row the chicken, putting the chicken on the right bank which also promotes P (CQ6).
Obj4c: If the farmer rows the seeds, he cannot row the dog to reach q_7 putting the dog on the right bank which would promote P and additionally would promote F (CQ11).

Reaching q_7 looks attractive: the farmer can get all the benefits of q_5 and more. This objection will reappear as Arg6, and we will consider counter arguments at that point.

Arg5: The farmer should row the seeds to reach q_6 where the seeds are on the right bank and the dog is fed promoting P and DH.

Again we can object that DH is not a value, or that we have a better action grounded on Arg6. Importantly also there is the objection:

Obj5a: Reaching q_6 demotes C (CQ9).

This threat of losing the chicken should be decisive. We next consider:

Arg6: The farmer should row the dog to reach q_7, promoting P and F.

This, however, is open to the critical question:

Obj6a: The chicken will eat the seeds, reaching q_8 and demoting S (CQ17, CQ9).

Assuming that the farmer values his seeds over the companionship of his dog, this will be decisive, unless he also accepts:

Obj6ai: The chicken won't eat the seeds (CQ2).

The above objection can be discounted, due to the presumed lack of restraint on the part of the chicken, and so q_7 will not be reached. Of course, there is a positive argument to reach q_8:

Arg7: The farmer should row the dog to reach q_8 where the dog is on the right bank and the chicken is fed promoting P, F and CH.

This is open to the objection Obj6ai, but more seriously open to:

Obj7a: Reaching q_8 will demote S (CQ9).

Finally we have one last argument:

Arg8: Farmer should row the chicken to reach q_9 putting the chicken on the right bank promoting P.

Note here the farmer's action determines the joint action since neither animal can do anything, but it is open to objections based on the preceding arguments which urge alternative actions on the farmer.

We have now generated a large number of arguments and objections. Reflecting on them we may well choose to reformulate our problem to make it simpler. For example we may accept Obj1c and Obj2e and decide that we do not wish to consider CH or DH as values. Further we may reject Obj1a and accept Obj2a and so decide that the chicken will always eat the seeds if it can and the dog will always eat the chicken if they are left unattended. Where seeds, chicken and dog are all left unattended, we may assume that the worst case will occur and that the chicken will eat the seeds and then the dog will eat the chicken. Making these decisions as to the factual aspects, the choices of the animals and the selection of acceptable values will reduce the situation of Fig. 1 to that shown in Fig. 2.

Now we have no argument for rowing alone, but Arg5, Arg7 and Arg8 are still possible. Arg5 will be rejected on the basis of Obj5a. Arg7 can also be rejected, on the basis of Obj7a (assuming S is preferred to F). Thus the accepted argument will be Arg8 and q_9 will be reached.

In q_9 we have rather fewer options, shown in Fig. 3. Doing nothing repeats state q_9, and so demotes progress, giving rise to the argument:

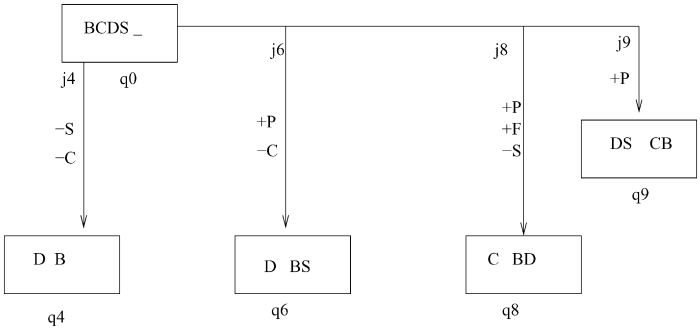

Fig. 2. Transitions from q_0 in reformulated problem.

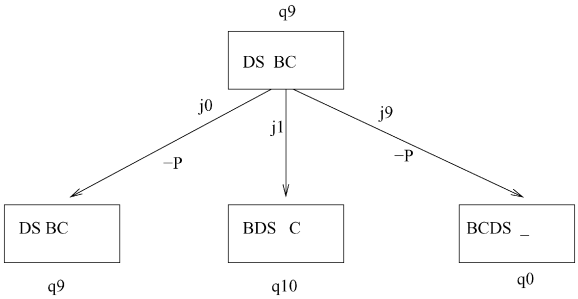

Fig. 3. Transitions from q_9.

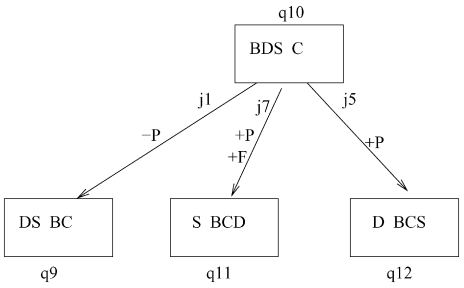

Fig. 4. Transitions from q_{10}.

Arg10: Farmer should not do nothing as this will reach a state already visited, demoting P.

Another agent might now use CQ3 to object that this did not achieve repetition, since it was staying in rather than repeating q_9. We can, however, reject this by choosing to count non-movement as repetition. Now we must consider the argument against rowing the chicken.

Arg11: Farmer should not row the chicken to reach q_0 putting the chicken on the left bank as this would demote P.

This leaves us with just one action which is not harmful, giving a default argument for rowing alone:

Arg12: Farmer should row alone to reach q_{10} which demotes nothing.

Note here the justification for the action is that it is the only non-harmful option. In some situations it may be that the "best" action is the "least bad". We now reach the state shown in Fig. 4.

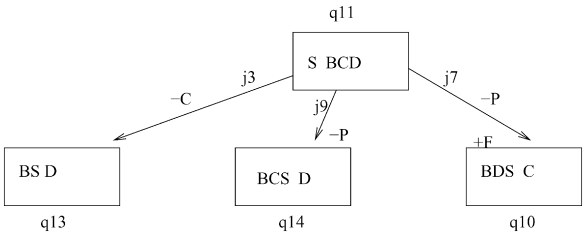

Fig. 5. Transitions from q_{11}.

Here we might argue for rowing the seeds to promote P, but this is open to the objection that rowing the dog is an alternative way of moving something to the right bank (CQ6), and this action, precluded by rowing the seeds, will promote F as well (CQ11). So we should row the dog here. This gives us the problem shown in Fig. 5.

Now the farmer has a reason to reject rowing alone, which will demote C, and so must choose to row the dog or the chicken.

Against rowing the chicken we have:

Arg13: Farmer should not row the chicken to reach q_{14} as this will put the chicken on the left bank which demotes P.

In favour of rowing the dog we have:

Arg14: Farmer should row the dog to reach q_{10}, which promotes F.

But this is open to the objections that it demotes P both by repeating q_{10} *and* by moving the dog to the left bank (CQ9).

Arg15: Farmer should not row the dog to reach q_{10}, which would put the dog on the left bank *and* so repeat q_{10}, *severely*[5] demoting P.

Rowing the dog has an additional degree of disadvantage for P to weigh against its promotion of F. Perhaps, however, the farmer will rank F as more important than severely impeding his progress (perhaps it is a pleasant day and he has plenty of time) and so will choose to row the dog. When he reaches q_{10} for the second time the context has changed: now, rowing the dog will demote progress by repeating q_{11}, and, when the farmer prefers P to F, he will choose to row the seeds rather than the dog. This gives the variant solution to the problem identified when the problem is considered as a search problem. Now as long as the farmer prefers F to P he will row the dog back and forth. Eventually, however, he will tire of this and progress will become urgent, and then he will prefer P over F. When he chooses to make progress he will either row the chicken, if in q_{11}, or the seeds, if in q_{10}. Let us, however, assume the farmer is in a hurry, and so prefers P to F from the outset and chooses to row the chicken as his first decision in q_{11}.

We now reach the situation shown in Fig. 6.

Rowing alone falls to the objection that the chicken will eat the seeds, demoting S (CQ9). Both rowing the seeds and rowing the chicken promote P, but rowing the chicken also demotes P, allowing CQ8. Thus the farmer will choose to row the seeds, and reach the situation shown in Fig. 7.

Here the farmer has two arguments to reject rowing the seeds, since it demotes P in two ways, by repetition and by putting the seeds back on the left bank. While he has no argument in favour of rowing alone, he has two arguments relating to rowing the dog, one for in that it promotes F and one against in that it demotes P (CQ9). The situation is similar to the one encountered in q_{11} (Fig. 5), and as there we will conclude that the farmer will, at some point, even if not immediately, tire of the dog's companionship sufficiently to value P more highly than F, leaving rowing alone as supported by the default argument that it is neutral as to all values.

[5] We use "severely" here to indicate the greater extent to which P is demoted by the realisation of two harmful goals. Here we need no more elaborate a mechanism for representing degrees of promotion and demotion: there is a discussion of more expressive quantitative methods in [3].

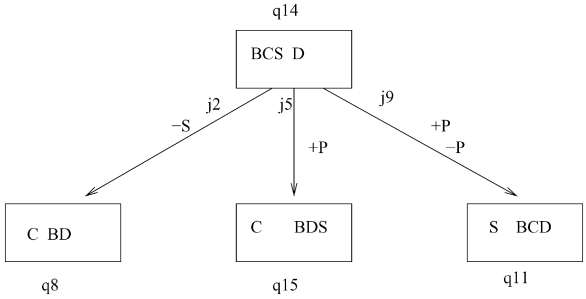

Fig. 6. Transitions from q_{14}.

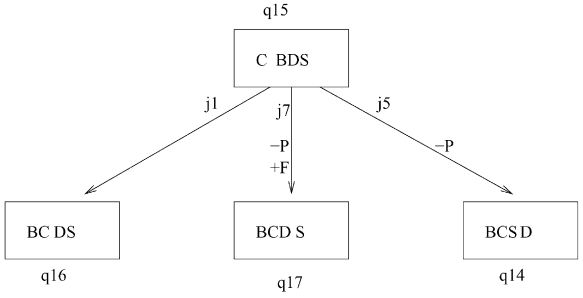

Fig. 7. Transitions from q_{15}.

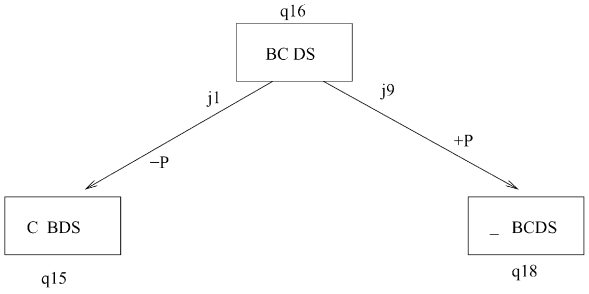

Fig. 8. Transitions from q_{16}.

At last we reach the situation shown in Fig. 8. Now the farmer will reject rowing alone because it demotes P, and the alternative of rowing the chicken promotes P. So he rows the chicken and now everything is on the right bank and so j_{10}, where all continue the journey home, is available. All the alternatives have arguments against them: rowing alone will demote C and S, rowing the chicken will demote P, rowing the dog will demote P and S, even though promoting F, and rowing the seeds will demote P and C. Thus going home will be selected and they will continue on their way.

In this example we have followed the farmer through a series of practical reasoning situations, as he solves the problem of how to transport his belongings across the river. The solution is, of course, essentially the same as would have been reached by traditional search methods, but it provides examples of the use of the proposed argument scheme, and the range of critical questions, both addressed to aspects of problem formulation, and to the candidate presumptive solutions to determine which would be the best action to perform according to the way in which the farmer ranks his values in particular contexts. Note that this permits the farmer to choose a less efficient solution, if it would serve his interests so to do. Note also that there is no need to appeal to an explicit ultimate goal: the problem is solved considering only what is best at each particular choice point. This is because the ultimate goal of going home is implicit in the value of progress being promoted by movement of his possessions to the right bank so as to satisfy the pre-conditions for j_{10}. We are not, of course, proposing this as a method of planning, since planning does require look ahead on occasion to avoid being stuck in a local optimum. This issue is examined in [13], where a variant of

CQ11 is used to take account of future actions that are precluded as well as current actions. The point here is just to illustrate the practical reasoning process using our argument scheme and critical questions.

We have discussed this simple example at length to illustrate the mechanics of our approach, as it provides scope for the instantiation of a number of competing arguments, and examples of all the critical questions. Additionally it shows the interaction of a number of values. In the next section we will compare our approach with an approach which uses social laws to co-ordinate multi-agent systems.

7. Comparison with social laws

In using an AATS to ground our account of practical reasoning we are using the same formalism as has been used to explore social laws as a means of co-ordinating multi-agent systems in [21]. Given this, there is opportunity to see whether we can gain insight by relating our work to this approach. We will use the running example of [21].

The example has two trains, one running eastwards and one running westwards. For most of the circuit each train has its own track, but this narrows to a single track shared by the trains where the track enters a narrow tunnel. If both trains enter the tunnel together, therefore, they will crash. The trains may be in one of three states, *away* from the tunnel, *waiting* to enter the tunnel, or *in* the tunnel. At each point they may move (away to waiting to in to away) or stay still. Two values are relevant: *safety*, which is demoted if there is a crash, and *progress*, which is promoted by moving and demoted by staying still. Initially they are both away. The transitions of the AATS for the example are shown in Table 1.

The social laws approach is to constrain the behaviour of agents in particular states, so as to achieve certain objectives, typically that some state is avoided, or that some states are eventually reached. A social law is said to be *effective* if compliance with the law ensures that the objectives are achieved. The objective of the example system is to ensure that there is no collision, that is, that state q8 is never reached. This state can only be reached from one of the three states q5, q6 and q7, and so constraints are needed to restrict the behaviour of the agents in these three cases. One such law proposed in [21], which we will call SL1, is as follows:

(1) when both trains are waiting (q5) the eastbound train should not move;
(2) when the westbound train is in the tunnel and the eastbound is waiting (q6) the eastbound train should not move;
(3) when the eastbound train is in the tunnel and the westbound is waiting (q7), then the westbound train should not move.

As is shown in [21] this social law is effective, in the sense that if it is obeyed, it will ensure that the trains do not collide. As noted there, however, this law is asymmetric, in that it favours the westbound train over the eastbound train (although a very similar social law could be made which favoured the eastbound train, by modifying the first condition).

The model of practical reasoning given in this paper enables us to evaluate the law in terms of what the agents would choose to do in the absence of constraints. So, let us view the situation as a practical reasoning problem. The AATS is given in [21], so no questions relating to problem formulation arise, and it is part of the assumptions that the agents are aware of which state they are in. In every state the agents will have an argument to move to promote

Table 1
Transitions/Pre-conditions/Interpretation

q/j	j_0	j_1	j_2	j_3	$\pi(q)$
q_0	q_0	q_1	q_3	q_5	$\{away_E, away_W\}$
q_1	q_1	q_2	q_5	q_6	$\{away_E, waiting_W\}$
q_2	q_2	q_0	q_6	q_3	$\{away_E, in_W\}$
q_3	q_3	q_5	q_4	q_7	$\{waiting_E, away_W\}$
q_4	q_4	q_7	q_0	q_1	$\{in_E, away_W\}$
q_5	q_5	q_6	q_7	q_8	$\{waiting_E, waiting_W\}$
q_6	q_6	q_3	q_8	q_4	$\{waiting_E, in_W\}$
q_7	q_7	q_8	q_1	q_2	$\{in_E, waiting_W\}$
q_8	q_8	–	–	–	$\{in_E, in_W\}$

progress. When in the tunnel, they will have the additional argument to move to avoid demoting safety. However, when waiting they will also have a counter-argument to moving, based on CQ17, since the action of the other agent may demote safety. There are thus, three states of interest, q5, q6 and q7, where CQ17 receives an unfavourable answer for one or both of the trains. Let us adopt the perspective of the eastbound train.

In q7 there will be an argument for the eastbound train to move, to promote progress. Note that this argument has no counter-arguments: whatever the westbound train does the eastbound train moves away and no collision occurs. Additionally there will be an argument to refrain from standing still, since this could demote safety if the westbound train chooses to move. Thus in q7 the eastbound train will choose to move. This action is permitted by SL1.

In q6, where the westbound train is in the tunnel, there is again an argument to move to promote progress, but this time there is also an argument not to move since this could lead to demotion of safety. Assuming the agent prefers safety to progress, this will cause it to remain stationary. Thus the agent will, if it does not consider the reasons why the westbound train will choose to move or stay still, choose the action prescribed by SL1.

In q5 there is similarly an argument to move, to promote progress, and an argument to stay still to avoid the possible demotion of safety. Once again, therefore, the rational agent will comply with the social law, even given freedom to act otherwise.

At this level of analysis, therefore it seems that the social law should, despite its asymmetry, be acceptable to the eastbound train, since it merely enjoins behaviour which would be chosen by a rational agent in any case.

There is, however, a problem with relying on rationality alone in that, in the absence of a social law, *both* trains will wait for one another in q5, since both will not dare to move themselves because of the danger that the other will move. The state will thus remain as q5 until the need for progress becomes so pressing that, for one or both of the trains, it comes to rank as more important than safety. If one train reaches this position before the other, the deadlock will be broken, and all will be well, but if both do so at the same time a collision will occur. Thus we cannot rely on rationality alone to ensure that q8 is not reached.

Here is where the social law gives an advantage: because SL1 does not allow the eastbound train to enter the tunnel, the westbound train has a counterargument to the argument that it must wait in case the eastbound train does enter the tunnel. If this assurance is sufficient to defeat the argument based on safety, the attack on the argument in favour of moving is defeated, and the train will choose to move.

Now the westbound train is able to move immediately on entering q5, while the eastbound train, to comply with the law, must idle for two turns. This will still, unless the eastbound train would become reckless very quickly, result in it entering the tunnel sooner that it would have done without the law. Thus both benefit from the law.

While the eastbound train will accept the constraint in q_5, it may have a problem with q_6, where it is not permitted to move because the westbound train is in the tunnel. In q_6 the westbound train will have a reason to leave the tunnel and none to stay there, and so, providing that the westbound train can be trusted to act in a rational manner, in the absence of a social law q_4 will be reached and no collision will occur. Thus the law that requires the eastbound train to wait will cause it to idle unnecessarily for a turn. Now just as the assumption that the eastbound train would comply with the social law that allowed the westbound train to enter the tunnel in q5, the assumption that the westbound train will act rationally would allow the argument from safety to be defeated in q6, so that the eastbound train could move immediately into the tunnel. In this situation, there would be a strong temptation to violate the social law, just as car drivers are often tempted to move off on an amber traffic light, relying on the cross traffic to have stopped in compliance with their signal. Similar considerations apply to the westbound train in q7.

For the both waiting situation, q5, there is a clear need for social regulation, and although one agent will benefit, the other will not suffer, since the enforced choice is the rational choice anyway and the existence of the social law will typically mean that progress can be made the sooner. With regard to waiting while the other train is in the tunnel, however, the social law can be seen to impose an unnecessary delay, compared to what would happen by rational choice alone, and it is possible that it will be disregarded by agents, in the belief that they can rely on the rationality of the other party. This raises the question of whether the other agent can rely on conformity with a social law by an agent who can see no reason, other than the law, to conform. Unless conformity can be assumed, the social law cannot allow it to choose an action which would lead to the undesired state.

The problem noted with the above formulation of the social law in [21], is different. There it is observed that since trains are always permitted to stay still under SL1, it could be that the system would come to a halt. In response the objective is refined by adding an additional objective that a waiting train will at some point be able to enter the tunnel. SL1 is not effective with respect to this objective. We would, however, be able to argue that this objective will in fact

be realised under SL1, since rational choice will cause a train in the tunnel to move away, allowing the other to enter when permitted to do so.

Since this consideration is not available to [21], another social law (SL2) is proposed, the same as the original, except that trains are forbidden from staying still the tunnel. This is effective, and should be acceptable to the agents, since they will choose to move in that situation in any case. Note, however, that the temptation to break constraints (2) and (3) of SL1 is increased, since both rationality and the assurance of the social law suggest that it is safe to move. Since removing these two clauses will not result in the law becoming ineffective, there seems to be every reason to remove these clauses from SL2.

Next we may consider the situation where the eastbound agent is not a train but a pedestrian. The reasoning of the eastbound agent will remain the same, but now the westbound agent, the train, will not have its safety jeopardised by a collision. Thus the train will have no argument based on its own safety to refrain from entering the tunnel in q5. In this case, assuming that the train considers only its own safety, there would be no need for a social law at all: we could simply rely on the rationality of the agents to avoid collisions. We might still, however, wish to make the social law to avoid deadlock where the train considered the safety of the pedestrian as a value, and rated this value over its own progress. In such a case it would be important to have the asymmetry so that the *train* is allowed to move in q5, since this reflects the rational situation, and to give priority to the pedestrian would risk the train violating the law since it was imposing behaviour which can be seen unnecessary from a rational point of view. This shows the importance of considering what is rational and framing the social law so as to reinforce rather than conflict with rationality. In this way agents can have the confidence that the other agents will conform to the law, which is essential if they are to choose their actions on this basis.

Social laws are intended to make the behaviour of other agents in the system more predictable. In our terms they provide an argument against an attack based on CQ17 to the effect that the other agent is not allowed to act so as to reach the undesired state. Note, however, that in situations such as that for the eastbound train in q7 there is also an argument against the attack based on CQ17 based on the fact that it is not in the other agent's interests to act so as to reach the undesired state. Agents may attempt to predict the behaviour of other agents both by awareness of social laws and by a consideration of what it is rational for these other agents to do. A good social law will be framed so that these considerations are in harmony.

This brief discussion suggests that considering practical reasoning in this manner gives a way of evaluating social laws additional to whether or not they are effective. A social law will be acceptable if it conforms to what would be rational behaviour, and desirable in that it permits desirable behaviour which might otherwise be avoided because of some associated risk. This acceptability remains even where there is a degree of asymmetry. Equally it is likely to be violated, and hence will have its effectiveness undermined, if it imposes behaviour not justifiable in rational terms.

8. Concluding remarks

Practical reasoning—reasoning about what it is best for a particular agent to do in a particular situation—exhibits a number of distinctive characteristics which make it quite different from reasoning about what is the case. Following Searle [20] we identify three characteristics in particular which any satisfactory account of practical reasoning must accommodate:

(1) The reasoning is inescapably defeasible and must always be evaluated in the context of the arguments that have been made in the particular situation. Additional arguments which can potentially lead to re-evaluation are always possible.
(2) The reasoning is inherently subjective, because it concerns *choice*. Different agents have different values, interests and aspirations, and this may lead them legitimately and rationally to accept different arguments.
(3) The preferences expressed by an agent should be capable of being a product of rather than an input to practical reasoning.

The approach described in Section 3 of this paper in which practical reasoning is seen as presumptive reasoning in the style of Walton [22] is capable of capturing these features. The defeasibility comes from posing an open ended set of critical questions to the *prime facie* justification represented by the instantiation of the argument scheme. The extension of Walton's original scheme to include the notion of *values* allows for the subjective nature of choice to be

represented. The third feature can also be included in this framework as is shown in [11]. The worth of the approach has been demonstrated in domains such as law [5,10], e-democracy [6] and medicine [8]. These previous applications of the approach, however, either used arguments taken from the domain and represented by hand, or were based on *ad hoc* implementations of BDI agents tailored for the purpose of the application. These applications generated a body of relevant argumentation which could be evaluated as a Value-Based Argumentation Framework [9], an extension to Dung's Argumentation Frameworks [14], able to reflect the different value orderings that different agents may wish to use. The approach described in this paper enables these frameworks to be generated on the basis of a well-defined structure.

The main contribution of this paper is to provide a well-defined structure enabling the precise specification of conditions under which an argument scheme and associated critical questions for practical reasoning can be instantiated. For this purpose we use an Action-based Alternating Transition System, extended to represent values and their promotion and demotion. This provides a rigorous foundation for further exploration of the approach, both theoretical and empirical. Some implementation work using a limited number of critical questions has already been done in [13]. Also worthy of investigation are properties of such systems: for example the conjecture in [4] that under certain constraints on value orderings an acceptable outcome from a situation of conflict of interest can be guaranteed. Representation in terms of an AATS will permit model checking techniques to be used to explore such questions, as has been done for social laws [21].

The promise of this approach to practical reasoning has been demonstrated in previous work. For this promise to be fully realised it is critical that some firm foundation for the approach be supplied: this paper supplies such a foundation.

Acknowledgements

We are grateful to Peter McBurney and Alison Chorley for their contributions to our earlier work which is developed into this paper.

References

[1] R. Alur, T.A. Henzinger, O. Kupferman, Alternating-time temporal logic, ACM Journal 49 (5) (2002) 672–713.
[2] L. Amgoud, S. Kaci, On the generation of bipolar goals in argumentation-based negotiation, in: I. Rahwan, P. Moraitis, C. Reed (Eds.), Argumentation in Multi-Agent Systems, in: Lecture Notes in Artificial Intelligence, vol. 3366, Springer, Berlin, 2004, pp. 192–207.
[3] K. Atkinson, What should we do?: Computational representation of persuasive argument in practical reasoning, PhD thesis, Department of Computer Science, University of Liverpool, Liverpool, UK, 2005.
[4] K. Atkinson, T. Bench-Capon, Addressing moral problems through practical reasoning, in: L. Goble, J.Ch. Meyer (Eds.), Deontic Logic and Artificial Normative Systems, in: Lecture Notes in Artificial Intelligence, vol. 4048, Springer, Berlin, 2006, pp. 8–23.
[5] K. Atkinson, T. Bench-Capon, P. McBurney, Arguing about cases as practical reasoning, in: Proceedings of the Tenth International Conference on Artificial Intelligence and Law (ICAIL 2005), ACM Press, 2005, pp. 35–44.
[6] K. Atkinson, T. Bench-Capon, P. McBurney, Multi-agent argumentation for eDemocracy, in: M.P. Gleizes, G. Kaminka, A. Nowé, S. Ossowski, K. Tuyls, K. Verbeeck (Eds.), Proceedings of the Third European Workshop on Multi-Agent Systems, 2005, pp. 35–46.
[7] K. Atkinson, T. Bench-Capon, P. McBurney, PARMENIDES: Facilitating deliberation in democracies, Artificial Intelligence and Law 14 (4) (2006) 261–275.
[8] K. Atkinson, T. Bench-Capon, S. Modgil, Argumentation for decision support, in: S. Bressan, J. Kung, R. Wagner (Eds.), Database and Expert Systems Applications, in: Lecture Notes in Computer Science, vol. 4080, Springer, Berlin, 2006, pp. 822–831.
[9] T. Bench-Capon, Persuasion in practical argument using value based argumentation frameworks, Journal of Logic and Computation 13 (3) (2003) 429–448.
[10] T. Bench-Capon, K. Atkinson, A. Chorley, Persuasion and value in legal argument, Journal of Logic and Computation 15 (2005) 1075–1097.
[11] T. Bench-Capon, S. Doutre, P.E. Dunne, Audiences in argumentation frameworks, Artificial Intelligence 171 (2006) 42–71.
[12] T. Bench-Capon, H. Prakken, Choosing what to do by accruing arguments, in: P.E. Dunne, T.J.M. Bench-Capon (Eds.), Computational Models of Natural Argument, Proceedings of COMMA 2006, in: Frontiers in Artificial Intelligence and Applications, vol. 144, IOS Press, 2006, pp. 247–258.
[13] A. Chorley, T. Bench-Capon, P. McBurney, Automating argumentation for deliberation in cases of conflict of interest, in: P.E. Dunne, T.J.M. Bench-Capon (Eds.), Computational Models of Natural Argument, Proceedings of COMMA 2006, in: Frontiers in Artificial Intelligence and Applications, vol. 144, IOS Press, 2006, pp. 279–290.
[14] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (1995) 321–357.
[15] A.J.P. Kenny, Practical reasoning and rational appetite, 1975. Reprinted in [19].

[16] S. Modgil, Value based argumentation in hierarchical argumentation frameworks, in: P.E. Dunne, T.J.M. Bench-Capon (Eds.), Computational Models of Natural Argument, Proceedings of COMMA 2006, in: Frontiers in Artificial Intelligence and Applications, vol. 144, IOS Press, 2006, pp. 297–308.
[17] Y. Moses, M. Tennenholtz, Artificial social systems, Computers and AI 14 (6) (1995) 533–562.
[18] I. Rahwan, L. Amgoud, An argumentation based approach for practical reasoning, in: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006), ACM Press, 2006, pp. 347–354.
[19] J. Raz (Ed.), Practical Reasoning, Oxford University Press, Oxford, UK, 1978.
[20] J.R. Searle, Rationality in Action, MIT Press, Cambridge, MA, 2001.
[21] W. van der Hoek, M. Roberts, M. Wooldridge, Social laws in alternating time: Effectiveness, feasibility and synthesis, Synthese 156 (1) (2007) 1–19.
[22] D.N. Walton, Argumentation Schemes for Presumptive Reasoning, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.