Book Title
Book Editors
IOS Press, 2003

Justifying Actions by Accruing Arguments

Trevor J.M. Bench-Capon [a], Henry Prakken [b]
[a] *Department of Computer Science, University of Liverpool, Liverpool, UK*
[b] *Department of Information and Computing Sciences, Universiteit Utrecht & Faculty of Law, University of Groningen, The Netherlands*

Abstract. This paper offers a logical formalisation of an argument-based account of reasoning about action, taking seriously the abductive nature of this form of reasoning. The particular question addressed is *what is the best way to achieve a specified goal?* Given a set of final goals and a set of rules on the effects of actions, the formation of subgoals for a goal is formalised as the application of an inference rule corresponding to the practical syllogism well-known from practical philosophy. Positive and negative applications of the practical syllogism are then accrued as a way to capture the positive and negative side effects of an action. Positive accruals can be attacked by negative accruals and by arguments for alternative ways to achieve the same goal. Defeat relations between accrued action arguments are determined in terms of the values promoted and demoted by the actions considered in the arguments. Applying preferred semantics to the result then yields the admissible ways to achieve the desired goal.

Keywords. Practical reasoning, argumentation, choice, goals, values

1. Introduction

In this paper we will address the problem of practical reasoning, which embraces questions such as: what is the best way to achieve a given purpose? how can an action be justified? and what should be done in a given situation? Here we will focus the first two of the questions, and discuss why this approach does not answer the third.

In philosophy the centre of discussion has been the practical syllogism, originally proposed by Aristotle [1]. Modern formulations take a form such as:

PS1: Agent P wishes to realise goal G
If P performs action A, G will be realised
Therefore, P should perform A

Problems with the practical syllogism as noted by, e.g. Kenny [2] include its abductive nature, and the need to consider alternatives and negative side effects before applying it. Walton [3] treats the practical syllogism as an argument scheme: instantiating the scheme supplies a presumptive reason for A, but this instantiation is then subject to a characteristic set of critical questions, which must be answered satisfactorily if the argument is to stand and the presumption upheld. These critical questions relate to the difficulties noted by Kenny. Atkinson [4] elaborated Walton's argument scheme by distinguishing the goal

into three elements: the state of affairs brought about by the action; the features of that state of affairs which are desired; and the social end or value which make those features desirable for the agent. These distinctions extended the critical questions from Walton's four to sixteen.

In this paper we aim to develop a logical formalisation of Atkinson's account within a logic for defeasible argumentation. We aim in particular to take the abductive nature of the practical syllogism seriously; its defeasible nature will be captured by stating negative answers to critical questions as counterarguments. A key ingredient in our formalisation is the use of [5]'s accrual mechanism for arguments to deal with side effects of an action. More precisely, given a set of final goals and a set of rules on the effects of actions, the formation of subgoals is formalised as the application of an inference rule expressing a positive or negative version of the scheme PS1. Both the positive and the negative applications are then accrued to capture the positive and negative side effects of an action. Positive accruals can be attacked by negative accruals and by arguments for alternative ways to achieve the same goal. Defeat relations between accrued arguments for actions are determined in terms of the values promoted and demoted by the actions advocated by the arguments. The admissible arguments are then computed within the logic using preferred semantics: if alternative ways to achieve the same goal are admissible, an ultimate choice has to be made outside the logic.

The remainder of the paper will proceed as follows. In Section 2 we recall Atkinson's account of PS1 and identify the aspects we will formalise. In Section 3 we will give some logical preliminaries, after which we present our main contribution in Section 4. Section 5 illustrates our approach with an example of a judge who must choose an appropriate sentence in order to punish a guilty person and we end in Sections 6 and 7 with a discussion of related research and some concluding remarks.

2. Atkinson's analysis of the practical syllogism

In this section we recall Atkinson's systemization of the practical syllogism and its sixteen critical questions, and we indicate which of these critical questions will be formalised in this paper. Atkinson's version of the practical syllogism is: *in the current circumstances, action A should be performed to bring about circumstances in which goal G is achieved, as this promotes value V*. The sixteen critical questions which can be posed against this argument scheme are:

CQ1: Are the believed circumstances true?
CQ2: Assuming the circumstances, does the action have the statedconsequences?
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 someother 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?

Addressing all these questions is beyond the scope of this paper. Five of the questions cater for differences between agents: in language (CQ12, CQ14 and CQ15); in the evaluation of states of affairs (CQ4); and in what counts as a value (CQ16). We will consider only a single agent, and so these questions do not arise.

CQ1 and CQ13 relate to the state of affairs in which the agent finds itself: CQ13 representing preconditions of the action and CQ1 preconditions for the action to have the desired effect. CQ2 on the other hand represents an undercutter of the defeasible rule that the action will achieve the goal if these preconditions are satisfied. These questions are internal to the argument deriving from the practical syllogism and can be considered answered if there is a (defeasible) proof. By embedding the practical syllogism in a general formalism for defeasible argumentation, we address these questions. In contrast, CQs5-11 all involve a separate argument, which attacks or reinforces the original argument, and so require a means of comparing arguments.

CQs5-7 concern alternatives to the proposed action. We will not consider further the distinction between state and goal: this is important only if a distinction between observable and inferred states is important. Although we will distinguish between goal and value, in the limiting case where there is a one-to-one correspondence between goals and values CQ6 and CQ7 collapse. On these assumptions, only CQ6 need be considered.

CQs8-10 all concern side effects. CQ8 and CQ9 refer to adverse side effects: for this we will require a negative form of the practical syllogism, so that we can conclude that we should refrain from an action. CQ10 refers to positive side effects and the existence of an argument here will encourage the performance of the action. CQ11 is different again in that it arises when the performance of an action achieves a goal which is incompatible with the goal which motivates some other action, thus preventing the simultaneous performance of both actions.

Questions relating to side effects (CQ8-10), positive and negative, all provide extra reasons for and against performing the action. To determine the net effect of these arguments we need to accrue them, and the use of a mechanism to allow this is a main idea of this paper. Before considering alternatives we need first to establish that the action provides a net benefit, which will determine the strength of the case for performing the action. Once the beneficial actions have been identified, the best should be chosen, and now alternatives must be considered, both alternative ways of achieving a goal (CQ6) and alternative goals (CQ11). Values are used in both comparisons. We will now present our formalisation of the argument scheme and the selected critical questions.

3. Logical preliminaries

The formalism used in this paper is based on Dung's [6] abstract approach to defeasible argumentation instantiated with a familiar tree-style approach to the structure of arguments [7, 8] and incorporating an accrual mechanism of arguments [5]. Here only the main definitions of these formalisms will be given; for the full technical details the reader is referred to the original sources.

The abstract framework of [6] assumes as input a set of unstructured arguments ordered with a binary defeat relation and defines various semantics for argument-based inference, all designating one or more conflict-free sets of arguments as so-called argument extensions. Two often-used semantics are grounded semantics, which always produces a unique extension, and preferred semantics, which produces more than one extension when a conflict between arguments cannot be resolved. In this paper we will adopt preferred semantics, since reasoning about action often involves an ultimate choice between various admissible courses of action. The basic notions of [6] that we need are defined as follows.

Definition 3.1 An *argument system* is a pair $\mathcal{H} = (\mathcal{A}, \mathcal{D})$, in which $\mathcal{A}$ is a set of *arguments* and $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$ is the *defeat* relationship for $\mathcal{H}$. When $(a, b) \in \mathcal{D}$ we say that a *defeats* b. For $S \subseteq \mathcal{A}$ we say that

1. $a \in \mathcal{A}$ is *acceptable w.r.t* S if for every $b \in \mathcal{A}$ that defeats a there is some $c \in S$ that defeats b.
2. S is *conflict-free* if no argument in S is defeated by an argument in S.
3. S is *admissible* if iS is conflict-free and every argument in S is acceptable w.r.t S.
4. S is a *preferred extension* of $\mathcal{H}$ if it is a $\subseteq$-maximal admissible subset of $\mathcal{A}$.
5. An argument is *justified w.r.t* $\mathcal{H}$ if it is in every preferred extension of $\mathcal{H}$.
6. An argument is *defensible w.r.t.* $\mathcal{H}$ if it is in some but not all preferred extensions of $\mathcal{H}$.

As for the *structure of arguments*, we assume they have a tree-structure where applications of *strict* and *defeasible* inference rules are chained into trees. Support relations between arguments are thus captured in the internal structure of arguments. Strict inference rules will be those of a monotonic propositional modal logic (see Section 4 below), while defeasible inference rules will be a modus ponens rule for defeasible conditionals, a rule for accrual of arguments and positive and negative versions of the practical syllogism. As for notation, all knowledge is expressed in a *logical language* $\mathcal{L}$. Strict inference rules are written as $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and defeasible rules as $\varphi_1, \ldots, \varphi_n \rightsquigarrow \varphi$ (where each φ and φ_i is a formula of $\mathcal{L}$). For any rule r its premises and conclusion are denoted, respectively, by $prem(r)$ and $conc(r)$. Each defeasible rule $\varphi_1, \ldots, \varphi_n \rightsquigarrow \varphi$ has a possibly empty set of *undercutters*, which are inference rules with conclusion $\neg\lceil\varphi_1, \ldots, \varphi_n \rightsquigarrow \varphi\rceil$. (For any term φ from the informal metalanguage of $\mathcal{L}$ the expression $\lceil\varphi\rceil$ denotes the object-level translation of φ in $\mathcal{L}$; cf. [7].)

The logical language $\mathcal{L}$ is divided into two sublanguages $\mathcal{L}_0$ and $\mathcal{L}_1$, where $\mathcal{L}_0$ is the language of a propositional modal logic to be defined in more detail in Section 4 and $\mathcal{L}_1$ is a rule language defined on top of $\mathcal{L}_0$. More specifically, $\mathcal{L}_1$ consists of so-called *defeasible conditionals*, or *defaults* for short, of the form $\varphi \Rightarrow \psi$, where ψ is a propo-

sitional literal and φ a conjunction of propositional literals of $\mathcal{L}_0$. (Note that defeasible conditionals, which express domain-specific knowledge in the object language, are not the same as defeasible inference rules, which express domain-independent inference principles in the metalanguage.) Defaults are assumed to satisfy the following inference rule of *defeasible modus ponens*:

DMP: $\varphi, \varphi \Rightarrow \psi \rightsquigarrow \psi$

Reasoning operates on a *theory* $T = (F, D)$ where F, a consistent set of $\mathcal{L}_0$ formulas, is a set of *facts* and D is a set of defaults. *Arguments* chain inference rules into AND trees, starting with a theory (F, D). For any argument A, its *formulas*, $form(A)$, are all the nodes in A, its *premises*, $prem(A)$, are all the leaf nodes of A, its *conclusion*, $conc(A)$, is its root and its *top rule*, $top(A)$, is the rule connecting the root of A with its children. An argument A is a *subargument* of an argument B if both have the same premises and $conc(A)$ is a node in B. An argument is *strict* if all its rules are strict, otherwise it is *defeasible*. A partial preorder $\leq_{\mathcal{A}}$ on the set $\mathcal{A}$ of arguments is assumed, where $A \leq_{\mathcal{A}} B$ means that B is at least as preferred as A. Related symbols are defined and subscripts omitted as usual. The preorder is assumed to satisfy the basic requirement that whenever A is strict and B defeasible then $A > B$.

As for *conflicts between arguments*, we include Pollock's [7] two ways of defeating defeasible arguments: they can be *rebutted* with an argument for the opposite conclusion and they can be *undercut* with an argument whose conclusion is that the defeasible reason applied in the attacked argument does not apply in the given circumstances. In Section 4 we will define a third form of attack for practical arguments, to deal with alternative ways to achieve the same goal. Non-undercutting conflicts between arguments will be adjudicated in terms of preference relation on arguments that takes into account the goals and values promoted and frustrated by an action.

Our formal definition of *defeat* follows common definitions in the literature.

Definition 3.2 (Defeat) Let A be an argument and B a defeasible argument.

- A *rebuts* B if $conc(A) = \neg conc(B)$ or vice versa, and $A \not< B$
- A *undercuts* B if $conc(A) = \neg\lceil top(B)\rceil$
- A *defeats* B if A rebuts or undercuts a subargument of B.

The following useful observation holds:

Observation 3.3 For all arguments A and B and preferred extensions E:

1. if A defeats a subargument of B then A defeats B;
2. if A is a subargument of B and $A \notin E$ then $B \notin E$.

Finally, as for *accrual of arguments*, [5] explains why it is worthwhile formalising this as an inference principle. Here we just recall its formalisation. The idea is that conclusions of defeasible arguments are labelled with their premises and that various defeasible arguments for the same conclusion are accrued by a defeasible inference rule that 'delabels' their conclusions. So, for instance, defeasible modus ponens now has the following form:

DMP: $\varphi, \varphi \Rightarrow \psi \rightsquigarrow \psi^{\{\varphi, \varphi \Rightarrow \psi\}}$

In the examples below the labels will for readability often be abbreviated with, possibly indexed, letters.

Next the definitions of conflicts between arguments are adjusted such that for rebutting the opposite conclusions must either be both unlabelled or have the same labels and that undercutting attack requires that the attacking arguments have unlabelled conclusions. Then a new accrual inference rule is added to the system, of the following form (in fact, the rule is a scheme for any natural number i such that $1 \leq i \leq n$):

$\varphi^{l_1}, \ldots, \varphi^{l_n} \rightsquigarrow \varphi$ (Accrual)

This inference rule and its undercutter below are the only ones that apply to labelled formulas; all other inference rules only apply to unlabelled formulas. Also, arguments are now required to have subset-minimal sets of premises to infer their conclusion, otherwise many irrelevant arguments would enter an accrual. Finally, to ensure that all relevant reasons for a conclusion are always accrued, the following undercutter scheme is formulated for any i such that $1 \leq i \leq n$.

$\varphi^{l_1}, \ldots, \varphi^{l_n} \rightsquigarrow \neg\lceil\varphi^{l_1}, \ldots, \varphi^{l_{n-i}} \rightsquigarrow \varphi\rceil$ (Accrual-undercutter)

The latter says that when a set of reasons accrues, no proper subset accrues. This undercutter is not needed if accruing arguments cannot weaken the case for the conclusion but this does not hold for all domains. For counterexamples see [5].

4. Arguments, conflict and defeat in practical reasoning

In this section we present our main contribution, a formalisation of reasoning with the practical syllogism. First we complete the definition of the logical language and its logic. The language $\mathcal{L}_0$ is a propositional modal logic with a single modality D standing for *desire*. Occurrences of D cannot be nested. To keep things simple, we abstract from the distinctions actions vs. states, procedural vs. declarative goals and achievement vs. maintenance goals: we only assume that the propositional part of $\mathcal{L}_0$ can be divided into *controllable* and *uncontrollable* formulas. Intuitively, the truth of controllable formulas is within an agent's control, but that of uncontrollable formulas (e.g. that it is raining) is not, so that only controllable formulas can be the subject of desires. The logic of D is assumed to be of type KD. Most importantly, this means that it validates $\neg(\mathrm{D}\varphi \wedge \mathrm{D}\neg\varphi)$, so that an argument for $\mathrm{D}\varphi$ can be extended by strict reasoning into an argument for $\neg\mathrm{D}\neg\varphi$.

Again for simplicity, we impose some further syntactic restrictions. Firstly, defaults cannot contain the modality D, and the only formulas in F that may contain D are of the form $\mathrm{D}\varphi$ where φ is a propositional literal from $\mathcal{L}_0$. We call the set of all such formulas in F the *goal base* G. Note that since it is a subset of F, it is assumed consistent. At first sight this would seem to prevent conflicting desires but as we will see below, we will allow for desires that turn out to be conflicting given the course that the world has taken; such 'contingent' conflicts between desires will then be subjected to our defeasible-reasoning mechanism. Contingent desire conflicts are inevitable and so our model must account for them, but it seems irrational to have desires that conflict no matter what happens.

Secondly, defaults now take one of the following forms, where all of a, r, r' and p are propositional literals and a is a controllable formula, r and r' are uncontrollable formulas and p is any propositional literal:

(i) $a \wedge r \Rightarrow p$
(ii) $a \Rightarrow p$
(iii) $r \Rightarrow r'$

Formulas of type (i) express that realising a in circumstance r achieves p, formulas of type (ii) say the same without referring to a circumstance, and formulas of type (iii) express that one circumstance typically implies another circumstance. In (i) and (ii), if p represents a state then the conditional is a causal rule, while if p represents an action the conditional is an action abstraction rule or 'counts as rule' [9], as in 'raising one's arm at an auction counts as making a bid'. Finally, formulas of type (iii) express defeasible kowledge about the world.

Next we formulate two defeasible inference rules for practical reasoning, viz. a positive and negative instance of the practical syllogism. Informally, if an agent who desires p and believes r also believes that realising a in circumstance r realises p, then this is a reason for desiring a, while if the agent believes that realising a in circumstance r instead realises $\neg p$, then this is a reason not to desire a. Note that thus practical and epistemic reasoning are interleaved, since r must be derived by epistemic reasoning. The new inference rules have the following form:

PPS: $a \wedge r \Rightarrow p, \mathrm{D}p, r \rightsquigarrow \mathrm{D}a$
NPS: $a \wedge r \Rightarrow \neg p, \mathrm{D}p, r \rightsquigarrow \neg \mathrm{D}a$

Applications of PPS can be rebutted as usual, for instance, by applications of NPS, but they can also be attacked by alternative applications of PPS to the same goal. In fact, the definition of alternatives attack is more complex than this, to deal with accrual of PPS applications to different goals.

Definition 4.1 Let A and B be two arguments.

1. A is an *alternative to* B if
   (a) $conc(A) = \mathrm{D}\varphi$ and $conc(B) = \mathrm{D}\psi$ $(\varphi \neq \psi)$; and
   (b) the last inferences in A, respectively, B apply the accrual inference rule to formulas $\mathrm{D}\varphi^{l_1}, \ldots, \mathrm{D}\varphi^{l_j}$, respectively, $\mathrm{D}\psi^{l_k}, \ldots, \mathrm{D}\psi^{l_n}$, such that:
      i. each such formula is the conclusion of a PPS application; and
      ii. at least one such PPS application in A shares a premise $\mathrm{D}\chi$ with at least one such a PPS application in B.
2. Argument A is a *sufficient alternative* to argument B if A is an alternative to B and $A \not\preceq B$.
3. A *defeats* B if A rebuts, undercuts or is a sufficient alternative to a subargument of B.

In this paper we assume for simplicity that goals are neither already achieved nor already prevented. This assumption could be relaxed by providing undercutters of PPS and NPS in terms of what can be concluded about whether p and a hold.

The next thing to address is the preference ordering on arguments. Following [10] we first formally define the notion of a value promoted by a goal.

Definition 4.2 Let V be a set of *values* ordered by a partial preorder $\leq_V$. The function v assigns to each formula $\mathrm{D}\varphi$ a, possibly empty, subset of V of values *promoted* by $\mathrm{D}\varphi$.

We allow that $\mathrm{D}\varphi \in V$ so that as a limiting case each goal just promotes itself. Note that this ranking of values may not only differ from agent to agent, but will also be dependent on the context in which the agent is reasoning. That this is how it should be is clear from [9], where it is persuasively argued that orderings of values often emerges from the reasoning process rather than being an input to it. In particular, when considering the question of the best way to achieve a particular goal, the value promoted by that goal must be given overriding importance, since the context presupposes that the decision to achieve that goal has already been taken, and that goal must be achieved if the question is to be answered. In other contexts, when considering how best to promote other goals, the values promoted by those other goals will take on greater importance. Now the idea is that the preference relation between conflicting practical arguments is determined by the sets of values promoted and demoted by the actions considered in the arguments, where an action demotes a value if it prevents the achievement of a goal promoting it. Alternative arguments will be compared by comparing pairs of sets: for each argument the pair contains the sets of values promoted, respectively demoted, by the argument.

As for notation, for any argument A and $\mathcal{L}_0$ formula φ, the *epistemic closure* $e(A, \varphi)$ of A under φ is the set of all propositional formulas that can be derived from $prem(A) \cup \{\varphi\}$ with epistemic reasoning, i.e., by using only strict inference rules and Defeasible Modus Ponens.

Definition 4.3 For any argument A with conclusion $\mathrm{D}\varphi$ or $\neg\mathrm{D}\varphi$ the pair $v(A) = (p_A, d_A)$ of *values promoted and demoted by* A is defined as follows.

1. If *conc*(A) = $\mathrm{D}\varphi$ then
   (a) $p_A = \{v \in V \mid v \in v(\mathrm{D}\psi)$ for some $\mathrm{D}\psi \in form(A)$ such that $\psi \in e(A, \varphi)\}$
   (b) $d_A = \{v \in V \mid v \in v(\mathrm{D}\psi)$ for some $\mathrm{D}\psi$ such that there exists an argument B with conclusion $\neg\mathrm{D}\varphi$ and $\mathrm{D}\psi \in form(B)$ and $\neg\psi \in e(B, \varphi)\}$
2. If *conc*(A) = $\neg\mathrm{D}\varphi$ then if $A_1, \ldots, A_n$ are all maximal proper subarguments of A for which $v(A_i)$ is defined $(1 \leq i \leq n)$ then
   (a) $p_A = p_{A_1} \cup \ldots p_{A_n}$
   (b) $d_A = d_{A_1} \cup \ldots d_{A_n}$

Let E be the set of all pairs (p_A, d_A) thus defined. Then $\leq_E$ is a partial preorder on E.

In clause (1), the function p_A simply collects A's initial goal and the goals derived from it using PPS, while d_A collects all initial and derivable goals that are prevented if A's final desire is carried out. To find these prevented goals, d_A looks at all rebuttals of A and computes their epistemic closures under A's final desire. The rationale of clause (2) is that in our setting the only ways to derive a conclusion of the form $\neg\mathrm{D}\varphi$ are to derive it from a positive desire by either NPS or $\mathrm{D}\neg\varphi \rightarrow \neg\mathrm{D}\varphi$. In other words, a negative desire always 'protects' a positive desire so that it seems reasonable that they have the same sets of promoted and demoted values.

We now impose the following constraint on the argument ordering $\leq_{\mathcal{A}}$. Let A be a defeasible argument with conclusion $\mathrm{D}\varphi$ and B a defeasible argument with conclusion $\mathrm{D}\psi$ or $\neg\mathrm{D}\varphi$. Then:

- $A \leq_{\mathcal{A}} B$ iff $v(A) \leq_E v(B)$

The idea now is that $\leq_E$ is defined in terms of $\leq_V$. Clearly, many reasonable definitions are possible and a discussion of them is beyond the scope of this paper; see [11] for some related definitions.

5. An example

In this section we illustrate our formalism with an example of a judge who must determine the best way to punish (pu) a criminal found guilty. He has three options: imprisonment (pr), a fine (fi) and community service (cs). Besides punishment there are three more goals at stake, deterring the general public (de), rehabilitating the offender (re) and protecting society from crime (pt). The judge must ensure that the offender is punished, and so pu will be the most important goal, but the method of punishment chosen will depend on the other goals that can be achieved by the various methods of punishing the offender. The judge believes that imprisonment promotes both deterrence and protection of society, while it demotes rehabilitation of the offender. He believes that a fine promotes deterrence but has no effect on rehabilitation or the protection of society since the offender would remain free, and he believes that community service has a positive effect on rehabilitation of the offender but a negative effect on deterrence since this punishment is not feared. This gives (with all $\mathcal{L}_0$ formulas controllable):

$pr \Rightarrow pu$	$pr \Rightarrow de$	$fi \Rightarrow de$	$cs \Rightarrow \neg\, de$
$fi \Rightarrow pu$	$pr \Rightarrow pt$		
$cs \Rightarrow pu$	$pr \Rightarrow \neg\, re$		$cs \Rightarrow re$

Finally, the judge's goal base $G = \{\mathrm{D}pu, \mathrm{D}pt, \mathrm{D}de, \mathrm{D}re\}$. These goals just promote themselves while no other goal promotes anything: in other words, the three possible sentences are purely instrumental in achieving goals in G.

The relevant arguments are depicted in Figures 1, 2 and 3. Assuming an equality

$$\dfrac{\dfrac{\dfrac{}{pr \Rightarrow pu} \quad \dfrac{}{\mathrm{D}pu}}{\mathrm{D}pr^{l_1}} \qquad \dfrac{\dfrac{}{pr \Rightarrow de} \quad \dfrac{}{\mathrm{D}de}}{\mathrm{D}pr^{l_2}} \qquad \dfrac{\dfrac{}{pr \Rightarrow pt} \quad \dfrac{}{\mathrm{D}pt}}{\mathrm{D}pr^{l_3}}}{\mathrm{D}pr} \qquad\qquad \dfrac{\dfrac{\dfrac{}{pr \Rightarrow \neg re} \quad \dfrac{}{\mathrm{D}re}}{\neg \mathrm{D}pr^{l_4}}}{\neg \mathrm{D}pr}$$

Figure 1. Accruals concerning imprisonment

$$\dfrac{\dfrac{\dfrac{}{fi \Rightarrow pu} \quad \dfrac{}{\mathrm{D}pu}}{\mathrm{D}fi^{l_5}} \qquad \dfrac{\dfrac{}{fi \Rightarrow de} \quad \dfrac{}{\mathrm{D}de}}{\mathrm{D}fi^{l_6}}}{\mathrm{D}fi}$$

Figure 2. Accrual concerning fining

$$
\begin{array}{ll}
Cs^+: & \\
\dfrac{\dfrac{cs \Rightarrow pu \quad \mathrm{D}pu}{\mathrm{D}cs^{l_7}} \quad \dfrac{cs \Rightarrow re \quad \mathrm{D}re}{\mathrm{D}cs^{l_8}}}{\mathrm{D}cs} & \qquad
Cs^-: \\
 & \dfrac{\dfrac{cs \Rightarrow \neg de \quad \mathrm{D}de}{\neg \mathrm{D}cs^{l_9}}}{\neg \mathrm{D}cs}
\end{array}
$$

Figure 3. Accruals concerning community service

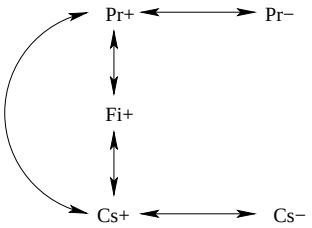

Figure 4. Partial defeat graph

argument ordering for the moment and ignoring subarguments this induces the defeat graph of Figure 4:

To adjudicate these conflicts, we must consider the values promoted and demoted by these arguments. We have that

$$
\begin{array}{llll}
v(Pr^+) = & (\{pu, de, pt\}, \{re\}) & v(Pr^-) = & (\{re\}, \emptyset) \\
v(Fi^+) = & (\{pu, de\}, \emptyset) & & \\
v(Cs^+) = & (\{pu, re\}, \{de\}) & v(Cs^-) = & (\{de\}, \emptyset)
\end{array}
$$

Recall that our question is *what is the best way to punish the offender?* We make pu an essential value, able to defeat any combination of other values, since no action that does not promote it can be an answer. This is enough to ensure that Pr^+ defeats Pr^- and Cs^+ defeats Cs^-. This leaves us with three ways to achieve our goal. Suppose that next to punishment we desire rehabilitation, and that promoting this is considered to be more important than deterrence and protection put together. Now Cs^+ will defeat Pr^+. Next we must consider whether promoting rehabilitation while demoting deterrence is preferable to promoting deterrence. If we think it is, we will accept Cs^+: if not we will accept Fi^+; and if we have no preference we will have two preferred extensions, and the choice of action must be made outside of this reasoning system. Suppose we in fact choose promoting rehabilitation while demoting deterrence over promoting deterrence: that will m ean that community service is our best way to achieve punishment. The justification for our choice will then be that given that we must punish the offender, we choose to do so in a way which will aid his rehabilitation.

We cannot now, however, go on to pose the question of what is our best *set* of actions in the situation. The problem is that both the actions of sending to prison and levying a fine have had the argument for them defeated because they are (given our preference for rehabilitation) inferior alternatives to community service with respect to punishment. But if these actions were compatible with community service we might wish to perform them for the sake of their other effects. We do not, however, have any undefeated arguments to justify this. We could, of course, develop a fresh set of arguments relating to the situation where community service is performed and its goals achieved, and use this

new framework to find the best way to achieve some other goal. Such a process would, however, be dependent on the order in which goals were focussed on, and so would not provide a good answer to this question. This identifies a limitation in our approach, which we will need to address in future work.

Finally, we briefly illustrate the interleaving in our approach of practical and epistemic reasoning. Consider a refinement of the rule that community service achieves rehabilitation with a noncontrollable condition that the offender is motivated:

$$cs \wedge mo \Rightarrow re$$

The condition *mo* must now hold to make PPS applicable; this gives rise to epistemic defeasible reasoning, where the new argument for $\mathrm{D}cs$ may be defeated because the subargument for *mo* is rebutted or undercut.

6. Related work

Because of space constraints we can only briefly discuss related work.

Thomason [12], Broersen et al. [13] and van Riemsdijk et al. [14] formalise defeasible reasoning about action using default logic as a way to deal with conflicting desires. They do not formalise abductive goal generation.

Pollock [15] argues that epistemic reasoning and planning should be interleaved and models this in his OSCAR system, adopting an abductive notion of goal regression. While we especially focus on choosing an action to achieve a particular goal, Pollock's focus is more on reasoning about plans for carrying out certain actions.

Most closely related to our work is Amgoud [11], who presents a unified model of argument-based inference and decision making within the same general framework adopted by us. Her counterpart to our positive and negative form of subgoal generation is a division of the goal base into goals to achieve and goals to avoid. Abductive goal generation is allowed but cannot be chained. Also, conflicts between alternatives do not arise in the logic but are subject to a separate decision-making process in which the logically justified action arguments are further compared. Amgoud's approach also allows for 'modus ponens' generation of subgoals applied to conditional desires. Since we allow for arbitrary chains of abductive goal generation, introducing conditional desires is not trivial in our case, for which reason we leave this for future study.

7. Conclusion

In this paper we have formalised a philosophically plausible approach to practical reasoning as defeasible argumentation, to address the question of what is the best way to achieve some particular goal. We have especially focussed on the abductive nature of reasoning about desires on the basis of beliefs and goals and on the accrual of positive and negative side effects of actions. Having said this, much future work remains. The restriction to contexts in which a goal to achieve has already been selected needs to relaxed. We also need to study extension to conjunctive desires, as well as refinements of the logical language to distinguish between actions and states, declarative and procedural goals, and achievement and maintenance goals. Finally, we should explore the various ways in which value orderings influence the comparison of arguments.

Acknowledgements

This work was partially supported by the EU under IST-FP6-002307 (ASPIC). Discussions within the ASPIC project have been very helpful, especially with Leila Amgoud, Peter McBurney and Sanjay Modgil.

References

[1] Aristotle. *Topics*. Clarendon Press, Oxford, 1997. Translated by R. Smith.

[2] A.J.P. Kenny. Practical reasoning and rational appetite. In J. Raz, editor, *Practical Reasoning*, pages 63–80. Oxford University Press, Oxford, 1978.

[3] D.N. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[4] 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.

[5] H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 85–94, New York, 2005. ACM Press.

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

[7] J.L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, Cambridge, MA, 1995.

[8] G.A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[9] J.R. Searle. *Rationality in Action*. MIT Press, Cambridge, MA, 2001.

[10] T.J.M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13:429–448, 2003.

[11] L. Amgoud. A unified setting for inference and decision: an argumentation-based approach. In *Proceedings of the IJCAI-2005 Workshop on Computational Models of Natural Argument*, pages 40–43, 2005.

[12] R.H. Thomason. Desires and defaults: a framework for planning with inferred goals. In *Proceedings of the 7th International Conference on Principles of Knowledge Representation and Reasoning*, pages 702–713, San Fransisco, CA, 2000. Morgan Kaufmann Publishers.

[13] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly Journal*, 2:428–447, 2002.

[14] M.B. van Riemsdijk, M. Dastani, and J.-J. Ch. Meyer. Semantics of declarative goals in agent programming. In *Proceedings of the Fourth International Conference on Autonomous Agents and Multiagent Systems (AAMAS-05)*, pages 133–140, 2005.

[15] J.L. Pollock. The logical foundations of goal-regression planning in autonomous agents. *Artificial Intelligence*, 106:267–335, 1998.