Preprint version. Official version may differ.

Argumentation-Based Decision Making and Structural Models of Personality

Pietro Baroni and **Federico Cerutti** and **Massimiliano Giacomin** and **Giovanni Guida**
Dipartimento di Ingegneria dell'Informazione, Università di Brescia,
Via Branze 38, I-25123 Brescia, Italy

Abstract

This paper extends a previous proposal of an argumentation-based approach to modelling articulated decision making contexts. The proposed approach encompasses a variety of argument and attack schemes aimed at representing basic knowledge and reasoning patterns for decision support. The paper draws a preliminary map between computational argumentation and structural models of personality using as a basic reference Freud's three-entities theory. Formal backing to this approach is provided by the $AFRA$ formalism, a recently proposed extension of Dung's argumentation framework. An example concerning a decision problem about medical treatments is adopted to illustrate the approach.

Introduction

Decision making is not just a matter of identifying a decision to be suggested. Good decision making outcomes, similarly to good human advices, should involve explanation and interaction with decision makers (Girle et al. 2003):

1. the advice should be presented in a form which can be readily understood by decision makers;
2. there should be ready access to both information and reasoning underpinning the advice;
3. if the decision making involves details which are unusual to the decision maker, it is of primary importance that s/he can discuss these details with his advisor.

In particular, the second point requires transparency of the reasoning leading to the proposed advice about what to do. Reasoning about what to do is often called "practical reasoning", an important investigation subject in the "Argumentation in AI" research community. In this context, two main questions can be identified: on the one hand, appropriate schemes for the representation of knowledge and reasoning patterns have to be defined, on the other hand mechanisms to compute reasoning outcomes have to be identified.

Concerning the former question (representation), the influential work of (Walton 1996) introduces the concept of "argument scheme" intended as the statement of a presumption in favor of a given conclusion, or goal. Whether this presumption stands or falls depends on the positive or negative answers to a set of "critical questions" associated with the scheme. This approach was further developed in (Atkinson, Bench-Capon, and McBurney 2006) where a refined argument scheme for practical reasoning has been proposed, encompassing the distinction between goals, which are the desired effects of an action, and values, which represent the actual underlying reasons for an agent to achieve a goal.

As to the latter question (computation), all the approaches based on argument schemes mentioned above seem to assume the existence of different reasoning levels; for example in (Atkinson, Bench-Capon, and McBurney 2006) the levels of decision outcomes, goals, and values. In this perspective the Value-based Argumentation Framework (VAF) (Bench-Capon 2003) extends Dung's argumentation framework (AF) (Dung 1995) by introducing meta-level reasoning about values. Another extension of Dung's AF, called Extended Argumentation Framework (EAF), proposed in (Modgil 2007; 2009), advocates the existence of attacks to attacks, rather than just to arguments. Relationships between these approaches are discussed in (Bench-Capon and Modgil 2008) where it is shown that any instance of VAF can be put in correspondence with an instance of EAF. Recently a formalism called Argumentation Framework with Recursive Attacks ($AFRA$) has been introduced in (Baroni et al. 2009b), where a more general notion of attack to attack than in EAF is considered. In fact, $AFRA$ encompasses attacks to attacks recursively without constraints, while in EAF attacks to attacks can not be attacked in turn.

Although these approaches are well-studied from the formal and computational viewpoint, they lack a mapping with a model of human personality. In this paper we want to draw a preliminary proposal for this mapping taking into account Freud's work (Freud 1933; 1923; 1940) where a three-entities (Id, Ego and Superego) model of human personality is provided. According to (Freud 1933), the Id is the innate part of our personality and is based on the pleasure principle. In other words, the Id wants whatever feels good at the time, with no consideration for the actual situation. The Ego, on the contrary, acts according to the reality principle, and "is that part of the Id which has been modified by the direct influence of the external world [...] The Ego represents what may be called reason and common sense, in contrast to the Id, which contains the passions" (Freud 1923). Finally, the Superego works in contradiction to the Id: indeed the Superego strives to act in a socially appropriate manner, whereas the Id just wants instant self-gratification. Therefore, the Ego has to conciliate the innate instincts driven by

Preprint version. Official version may differ.

pleasure principle and the social constraints.

In this paper we describe a model of human decision making in the context of an argumentation-based approach, introduced in (Baroni et al. 2009a), and propose a correspondence between this model and Freud's three-entities model of personality. This allows us to carry out the task of computing the decision(s) to be proposed relying on a sound argumentation framework, thus achieving transparency of the computation process and a mapping to the personality entities involved in the decision. In order to achieve this goal, we proceed in three steps: first we introduce an articulated set of concepts for representing a Decision Support Problem; then, we focus on the computation of the decision to be suggested, relying on the AFRA formalism (Baroni et al. 2009b; 2010). Finally we draw a mapping between computational notions and Freud's entities. An example is used throughout the paper to illustrate the proposed approach.

An Example

We will illustrate the decision process as described in (Baroni et al. 2009a) by referring to an example concerning medical treatment.

John, a volunteer of a Non-Governmental Organisation who is about to leave for a work-trip in the centre of Amazon Rainforest, suffers from pain due to spinal disc hernia. Therefore, the action to be chosen concerns the treatment for John disease where the relevant goal is to reduce disc herniation. It is assumed that in the available knowledge base there are only two actions to achieve that goal: "do a discectomy surgery" (**A2**) and "have a long non-invasive treatment" (**A3**). Both choices achieve the goal of reducing disc herniation which will promote the value of safety (**V2**) and require John to stay in Europe and this is in contrast with John's unselfishness. Therefore, John has another possible action, namely "take analgesics" (**A1**) achieving the goal of letting him free to go to Brazil and then promoting the value of "charity" (**V1**). Since we assume that only one decision has to be adopted, the alternative choices are mutually incompatible.

Furthermore, John has a history of anaesthesia allergy; thus surgery is not appropriate (**A4**) because it gives rise to risk of shock which will demote the value of safety. Clearly, **A4** and **A2** are in conflict, moreover from the reasons underlying **A4** we have learned that **A2** demotes the value of safety. Therefore the evidence supporting **A4** also provides information against the relation between **A2** and the value of safety. However, there is a kind of anaesthesia to which he is not allergic (**A5**): in this case **A4** looses its support and **A2** is reinstated.

Then suppose that between a long non-invasive treatment and surgery John prefers surgery (**P1**) because in few days he could get over his illness. Such a preference is not in contrast with **A3**: it states only that in this particular situation, if John has to choose between **A2** and **A3**, he would prefer **A2**.

However, John is really frightened by surgery (**E1**) and although he rationally considers surgery as a preferable action, he emotionally cannot accept it. Therefore this emotional argument undermines **A2** without affecting the rational preference **P1**. John can now determine the ultimate decision outcome by considering values. In order to achieve the goal of reducing disc herniation, which promotes the value of safety, the action of having a long non-invasive treatment has to prevail on the action "take analgesics". To this purpose we have to consider the value of safety in a strong form stating that "we must promote the value of safety". Then, the final decision outcome, if John chooses to definitely promote the value of safety, will be {**A3**}, namely John will have a long non-invasive treatment, even if this requires him not to go to Brazil. This is a rational decision that consider also John's feeling about surgeries. Furthermore, if John finds a reason (rational or emotional) that undermines his fear of surgeries or, at least, his fear of this particular surgery, then his rational preference **P1** allows him to choose the discectomy rather the long non-invasive treatment. Otherwise, if he considers charity more important than safety, the final advice will be "take analgesics".

Representing a decision problem

A decision problem may be formalized adopting an argument-based approach where two basic notions, namely arguments and attacks, are encompassed. For a suitable representation, both notions need to be specialized. Arguments of different sorts can be identified in relation with different reasoning levels (e.g. about goals rather than about values). This involves in turn different kinds of attack relations (possibly encompassing attacks to attacks, as it will be discussed later). Accordingly, the modeling approach we propose is based on an articulated set of concepts:

- the notion of practical argument scheme **PAS**, derived from (Atkinson, Bench-Capon, and McBurney 2006);
- the concept of practical attack scheme **PAtS**, that defines the conditions for an attack relation to hold between two instances of **PAS**;
- the concept of factual argument scheme **FAS**, defining the circumstances holding in a certain step of the reasoning process, and the related concept of factual attack scheme **FAtS** between an instance of **FAS** and an instance of **PAS** or **FAS**;
- the concept of value argument scheme **VAS**, asserting that a given value is in force, and the related concepts of: (1) value attack scheme **VAtS**, involving incompatible values, (2) value defence scheme **VDeS**, involving attacks from a value scheme to other attacks, and (3) value-argument attack scheme **VAAtS** involving argument schemes related through a value defence scheme;
- a preference argument scheme **PRAS**, to define a preference ordering between instances of other schemes, and the related preference attack scheme **PRAtS**, covering the cases where a preference undermines an attack;
- an emotional argument scheme **EAS**, to define a personal feeling about a course of action, and the related attack scheme **EAtS** aimed at defeating or supporting (accordingly with the feeling) the argument concerning the specific course of action at hand;

- a must argument scheme **MAS**, which acts as referee that accommodates personal feelings about practical arguments and values by stating a value to be promoted above all others, and the related must attack schemes **MAtS**.

We will proceed in our presentation according to the above plan: each scheme will be introduced as a tuple of entities, whose meaning will be commented case by case taking into account the example described in the previous section.

First, the following definition provides a modified version of the Practical Argument Scheme proposed in (Atkinson, Bench-Capon, and McBurney 2006) (in particular, we omit the future circumstances which are caused by an action and consider only the goals it achieves).

PAS: `circumstance:` C,
`action:` A,
`goal:` G,
`value:` V,
`sign:` +/-.

The scheme means that "in the circumstances C, the suggested action is A, which achieves the goal G, which, depending on sign, promotes or demotes the value V". We assume that the suggested action may be also a "negative" action $\neg$ A, with the meaning "it is suggested not to perform action A". We assume also that $\neg\neg A = A$.

From the example concerning John's decision, there are four instances of **PAS**, namely:

A1: `circumstance:` in my situation,
`action:` take analgesics,
`goal:` being free to go to Brazil,
`value:` charity,
`sign:` +.

A2: `circumstance:` in my situation,
`action:` discectomy surgery,
`goal:` reducing disc herniation,
`value:` safety,
`sign:` +.

A3: `circumstance:` in my situation,
`action:` long non-invasive treatment,
`goal:` reducing disc herniation,
`value:` safety,
`sign:` +.

A4: `circumstance:` alternative anaesthesia not available for discectomy,
`action:` not discectomy surgery,
`goal:` risk of shock,
`value:` safety,
`sign:` −.

According to (Atkinson, Bench-Capon, and McBurney 2006), **A4** should be read as "In the circumstances where alternative anaesthesia is not available for discectomy, John should not do discectomy surgery to avoid the risk of shock which will demote the value of safety".

We can then introduce two attack schemes between instances of **PAS**. Any attack scheme has a source, namely an instance of a scheme with the role of attacker, and a target, namely an instance of a scheme which is attacked, and specifies the conditions under which the attack takes place.

PAtS1: `source:` an instance of **PAS**,
`target:` an instance of **PAS**,
`conditions:` `source.action` $\neq$ `target.action` and both `source.action` and `target.action` are positive.

PAtS2: `source:` an instance of **PAS**,
`target:` an instance of **PAS**,
`conditions:` `source.action` $=$ $\neg$ `target.action`.

PAtS1 corresponds to the case where distinct positive actions are incompatible and therefore we have to choose exactly one action. **PAtS2** corresponds to the case where the source **PAS** suggests not to perform the action supported by the target **PAS** and vice versa.

In the example, we assume that only one action among those supported by **A1**, **A2**, and **A3** can be chosen. Therefore **A1**, **A2**, and **A3** attack each other according to the **PAtS1** scheme. Moreover **A4** supports the negation of the action supported by **A2**. Hence **A2** and **A4** attack each other according to the **PAtS2** scheme.

We now introduce a simple factual argument scheme, corresponding to the assertion that the circumstances C hold.

FAS: `circumstances:` C.

We assume that it is possible to negate that certain circumstances hold, using again $\neg$ as negation symbol. We can then introduce a factual attack scheme: a factual assertion attacks a **PAS** or another **FAS** by negating its circumstances.

FAtS1: `source:` an instance of **FAS**,
`target:` an instance of **PAS** or **FAS**,
`conditions:`
`source.circumstances` $=$ $\neg$ `target.circumstances`.

In the example, the only instance of **FAS** is:

A5: `circumstances:` alternative anaesthesia is available for discectomy.

Furthermore, **A5** negates the circumstance of **A4**, thus attacking it according to the **FAtS** scheme.

We introduce also a simple value argument scheme, representing the fact that the value V is in force.

VAS: `value:` V.

In the example two distinct values are considered:

V1: `value:` charity.

V2: `value:` safety.

Several attack schemes involve instances of **VAS**. First, we assume that a symmetric incompatibility relation, denoted as I, is defined among values: incompatible values attack each other, giving rise to a value attack scheme. Let us notice that, in the illustrating example, **V1** and **V2** are assumed not to be incompatible *per se*. Therefore there are no instances of the following attack scheme:

VAtS: `source:` an instance of **VAS**,
`target:` an instance of **VAS**,
`conditions:`
(`source.value`, `target.value`) $\in I$.

Then we consider the defence of practical arguments based on values, following an idea proposed in (Modgil 2007; Bench-Capon and Modgil 2008). In words, a **VAS**

argument, let say **V1**, defends an instance of **PAS**, let say **Pa1**, against an instance of attack **PAtS1** or **PAtS2** whose target is **Pa1**.

The defence takes place if the value of **V1** coincides with that of **Pa1** and is different from the value of the source of the attack against **Pa1**. This is represented by the following scheme.

VDeS1: `source:` an instance of **VAS**,
`target:` an instance of **PAtS1** or **PAtS2**,
`conditions: target.target.value` $=$ `source.value` and
`target.source.value` $\neq$ `source.value`.

For instance, in John's dilemma **V1** attacks the attacks directed to **A1** whose source promotes a different value, and so does **V2** with respect to attacks against **A2** and **A3**.

As values may defend arguments, they may analogously defend attacks. In fact, if the source of an attack is a practical argument, then it promotes or demotes a particular value. Such an attack is in turn strictly related to such value, therefore the value must defend this attack against the attacks it may receive from other values. To exemplify, an instance of **VDeS1**, let say **Vd1**, may be attacked by an instance of **VAS**, let say **V2**, if the value of **V2** is different from that of the source of **Vd1** and coincides with the value of the source of the instance of **PAtS1** or **PAtS2** which is the target of the attack **Vd1**. This is expressed by the following scheme.

VDeS2: `source:` an instance of **VAS**,
`target:` an instance of **VDeS1**,
`conditions: target.source` $\neq$ `source` and
`target.target.source.value` $=$ `source.value`.

In the illustrating example, **V1** attacks the instances of **VDeS1** based on a different value (actually **V2**) and whose target is an attack whose source promotes **V1** and a dual consideration applies to **V2**.

Since both **VDeS1** and **VDeS2** schemes represent a defence originated by a value in favor of a practical argument, we can see them as specializations of an abstract defence scheme that we call **VDefence**.

VDefence: `defending:` an instance of **VAS**,
`defended:` an instance of **PAS**,
`conditions: defending.value` $=$ `defended.value` and `defended.sign` $= +$.

In particular, letting **X** be an instance of **VDeS1**, we have **X**`.defending` $=$ **X**`.source` and **X**`.defended` $=$ **X**`.target.target`. On the other hand, letting **Y** be an instance of **VDeS2** we have: **Y**`.defending` $=$ **Y**`.source` and **Y**`.defended` $=$ **Y**`.target.target.source`. In the example, **V2** defends **A2** and **A3**, rather **V1** defends **A1**.

We can now introduce the last attack scheme concerning practical arguments which is strictly related with an issue pointed out in (Atkinson, Bench-Capon, and McBurney 2006). This concerns the case where a practical argument suggests not to perform an action A, since it demotes a value.

VAAtS: `source:` an instance of **PAS**,
`target:` an instance of **VDefence**,
`conditions: source.action` $=$ $\neg$ `target.defended.action`
and `source.value` $=$ `target.defended.value`
and `source.sign` $= -$
and `target.defended.sign` $= +$.

This scheme applies in cases where an instance **Pa1** of **PAS** (the source) has a strict relation with an instance **Pa2** of **PAS** defended by an instance of **VDefence** (**Pa2** corresponds to `target.defended`). **Pa1** and **Pa2** are related to the same value but with different signs (**Pa1** demotes it, while **Pa2** promotes it), and **Pa1** suggests not to execute the action supported by **Pa2**. In a word, **Pa1** tells that in order to promote the value of **Pa2** one has actually not to perform the action suggested by **Pa2**, i.e. in the current circumstances **Pa2** would demote, instead of promoting, its value. Consequently, any instance of **VDefence**, which defends **Pa2** on the basis of the value it promotes, is attacked in turn by **Pa1** (note that, given the above assumptions, it also holds that **Pa1** attacks **Pa2** according to the **PAtS2** scheme). In the example, since **A4** gives us a presumption that **A2** demotes the value of safety, then **A4** attacks the defences of **A2** grounded on the same value argument.

Let us turn now to the representation of preferences. A preference argument scheme simply corresponds to stating that an argument is preferred to another one.

PRAS: `preferred:` P,
`notpreferred:` nP.

Following (Modgil 2007), we can then define an attack scheme based on preferences: an instance of **PRAS**, let say **Pref1**, attacks an attack, let say **Att1**, if **Pref1** states that the target of **Att1** is preferred to the source of **Att1**:

PRAtS: `source:` an instance of **PRAS**,
`target:` an instance of **PAtS1**, or **PAtS2**, or **VAtS**,
`conditions: source.preferred` $=$ `target.target`
and `source.notpreferred` $=$ `target.source`.

Let us consider again John's dilemma; he has a personal preference for surgery rather the other treatment, and this can be formalised as an instance of **PRAS**:

P1: `preferred:` **A2**,
`notpreferred:` **A3**.

Moreover, given the preference for **A2** over **A3**, **P1** attacks the attack from **A3** to **A2** according to the **PRAtS** scheme.

Concerning emotions, they are often fundamental parts in human decision making (Girle et al. 2003; Nawwab, Dunne, and Bench-Capon 2010). A comprehensive review of this topic is beyond the scope of this preliminary work. For our purpose, we consider the relationships between emotional arguments and practical arguments that support a "favourable" (respectively "unfavourable") action w.r.t. the emotional feeling. Indeed an emotional argument should attack the unfavourable actions and defend the favourable

ones. The emotional argument contains the description of the personal emotion:

EAS: `emotion:` E.

Before discussing the relevant attack schemes involving emotional arguments, we define an emotional relationship between them and practical arguments that suggest a favourable or unfavourable course of action.

Definition 1. *Let $\mathcal{A}_{EAS}$ be the set of instances of **EAS** and $\mathcal{A}_{PAS}$ the set of instances of **PAS**, we define the emotional relationship as the partial function $e : \mathcal{A}_{EAS} \times \mathcal{A}_{PAS} \mapsto \{+, -\}$*

Then we define the following two emotional attack schemes. The first one considers the case where a practical argument suggests to perform an action considered as unfavourable by an emotional argument.

EAtS1: `source:` an instance of **EAS**,
`target:` an instance of **PAS**,
`conditions:` $e($`source`, `target`$) = -$.

Indeed, John's emotions suggest him not to choose surgery:

E1: `emotion:` surgery frightening.

with $e(\mathbf{E1}, \mathbf{A2}) = -$ and, consequently, **E1** attacks the argument **A2** according to **EAtS1**.

The following emotional abstract attack scheme considers the case where a practical argument is defended by an emotional argument.

EDefence: `defending:` an instance of **EAS**,
`defended:` an instance of **PAS**,
`conditions:`
$e($`defending`, `defended`$) = +$.

The specific schemes instantiating **EDefence** follow the main reasoning line underlying **VDeS1**, **VDeS2** and their relationship with **VDefence** with obvious modifications.

This analysis of the relationships between practical and emotional arguments is still preliminary. Indeed, we do not consider in this paper how emotional arguments and the attacks grounded on them can be undermined, although we believe that they too should be subjected to evaluation by other arguments.

In order to accommodate personal attitudes about practical arguments and values, we define a simple "must" argument scheme concerning a single value which must be promoted over all others.

MAS: `value:` V.

A **MAS** argument tries to mediate between the necessity to reach a decision, and the moral requirements that can affect it. In other words, the **MAS** argument acts as a referee that accommodates personal attitudes about values and the reality that requires a decision. In this preliminary work we consider that a **MAS** argument can only attack the defences of some values in favour of some practical arguments, and indirectly affect the practical arguments. In the example, John considers a **MAS** argument concerning safety.

MUST V2: `value:` safety.

In relation with **MAS** we introduce two attack schemes.

MAtS1: `source:` an instance of **MAS**,
`target:` an instance of **VAtS**,
`conditions:` `source.value` $=$ `target.target.value`
and `source.value` $\neq$ `target.source.value`.

MAtS2: `source:` an instance of **MAS**,
`target:` an instance of **VDeS2**,
`conditions:` `source.value` $=$ `target.target.source.value`
and `source.value` $\neq$ `target.source.value`.

In both schemes an instance of **MAS**, call it **M1**, defends an instance of **VAS** with the same value. In particular, in the **MAtS1** scheme, **M1** attacks an instance of **VAtS** attack (let say **Va1**) such that the target of **Va1** has the same value as **M1** and the value of the source of **Va1** is different from the one of **M1**. In the **MAtS2** scheme, **M1** attacks an instance of **VDeS2** attack (let say **Vd2**) since **Vd2** attacks another attack whose source is based on the same value as **M1**, while the source of **Vd2** is a value different from the one of **M1**.

Summing up, we may define an Argumentation Framework representing a Decision Support Problem as a tuple including instances of all schemes introduced above.

Definition 2 ($AFDSP$)**.** *An Argumentation Framework for Decision Support Problem ($AFDSP$) is a 10-ple $\langle \mathcal{A}_{PAS}, \mathcal{A}_{PRAS}, \mathcal{A}_{EAS}, \mathcal{A}_{VAS}, \mathcal{A}_{FAS}, \mathcal{A}_{MAS}, \mathcal{R}_{PAS}, \mathcal{R}_{PRAS}, \mathcal{R}_{EAS}, \mathcal{R}_{VAS}, \mathcal{R}_{FAS}, \mathcal{R}_{MAS} \rangle$ s.t.:*

- *$\mathcal{A}_{PAS}$ is a set of instances of **PAS**;*
- *$\mathcal{A}_{PRAS}$ is a set of instances of **PRAS**;*
- *$\mathcal{A}_{EAS}$ is the set of instances of **EAS**;*
- *$\mathcal{A}_{VAS}$ is a set of instances of **VAS**;*
- *$\mathcal{A}_{FAS}$ is a set of instances of **FAS**;*
- *$\mathcal{A}_{MAS}$ is a set of instances of **MAS**;*
- *$\mathcal{R}_{PAS}$ is a set of instances of **PAtS1** and **PAtS2**;*
- *$\mathcal{R}_{PRAS}$ is a set of instances of **PRAtS**;*
- *$\mathcal{R}_{EAS}$ is a set of instances of **EAtS1** and **EDefence**;*
- *$\mathcal{R}_{VAS}$ is a set of instances of **VAtS**, **VDeS1**, **VDeS2**, **VAAtS**;*
- *$\mathcal{R}_{FAS}$ is a set of instances of **FAtS**;*
- *$\mathcal{R}_{MAS}$ is a set of instances of **MAtS1**, and **MAtS2**.*

Referring to John's dilemma, we can build an Argumentation Framework for Decision Support Problem $\widehat{\Phi}$ where: $\mathcal{A}_{PAS} = \{\mathbf{A1}, \mathbf{A2}, \mathbf{A3}, \mathbf{A4}\}$, $\mathcal{A}_{PRAS} = \{\mathbf{P1}\}$, $\mathcal{A}_{EAS} = \{\mathbf{E1}\}$, $\mathcal{A}_{FAS} = \{\mathbf{A5}\}$, $\mathcal{A}_{VAS} = \{\mathbf{V1}, \mathbf{V2}\}$ and $\mathcal{A}_{MAS} = \{\mathbf{MUST\ V2}\}$.

Let us examine now the attack relations that arise in the example. It is straightforward to see that: $\mathcal{R}_{PAS} = \{(\mathbf{A1}, \mathbf{A2}), (\mathbf{A2}, \mathbf{A1}), (\mathbf{A3}, \mathbf{A1}), (\mathbf{A1}, \mathbf{A3}), (\mathbf{A2}, \mathbf{A3}), (\mathbf{A3}, \mathbf{A2}), (\mathbf{A4}, \mathbf{A2}), (\mathbf{A2}, \mathbf{A4})\}$; $\mathcal{R}_{FAS} = \{(\mathbf{A5}, \mathbf{A4})\}$; $\mathcal{R}_{PRAS} = \{(\mathbf{P1}, (\mathbf{A3}, \mathbf{A2}))\}$; and $\mathcal{R}_{EAS} = \{(\mathbf{E1}, \mathbf{A2})\}$.

The attack schemes **VDeS1** and **VDeS2** give rise, respectively, to the following sets of attacks: $\mathcal{R}^1_{VAS} = \{(\mathbf{V1}, (\mathbf{A3}, \mathbf{A1})), (\mathbf{V1}, (\mathbf{A2}, \mathbf{A1})), (\mathbf{V2}, (\mathbf{A1}, \mathbf{A2})), (\mathbf{V2},$

Preprint version. Official version may differ.

$(\mathbf{A1}, \mathbf{A3}))\}$ and $\mathcal{R}^2_{VAS} = \{(\mathbf{V1}, (\mathbf{V2}, (\mathbf{A1}, \mathbf{A2}))), (\mathbf{V1}, (\mathbf{V2}, (\mathbf{A1}, \mathbf{A3}))), (\mathbf{V2}, (\mathbf{V1}, (\mathbf{A3}, \mathbf{A1}))), (\mathbf{V2}, (\mathbf{V1}, (\mathbf{A2}, \mathbf{A1})))\}$. To conclude the discussion concerning attacks related to values, let us take into account the **VAAtS** schema, which gives rise to $\mathcal{R}^3_{VAS} = \{(\mathbf{A4}, (\mathbf{V2}, (\mathbf{A1}, \mathbf{A2}))), (\mathbf{A4}, (\mathbf{V2}, (\mathbf{V1}, (\mathbf{A2}, \mathbf{A1}))))\}$. In summary, $\mathcal{R}_{VAS} = \mathcal{R}^1_{VAS} \cup \mathcal{R}^2_{VAS} \cup \mathcal{R}^3_{VAS}$.

Finally, according to the **MAtS2** scheme two further attacks arise: $\mathcal{R}_{MAS} = \{(\mathbf{MUST\ V2}, (\mathbf{V1}, (\mathbf{V2}, (\mathbf{A1}, \mathbf{A2})))), (\mathbf{MUST\ V2}, (\mathbf{V1}, (\mathbf{V2}, (\mathbf{A1}, \mathbf{A3}))))\}$.

Therefore, $\widehat{\Phi} = \langle \mathcal{A}_{PAS}, \mathcal{A}_{PRAS}, \mathcal{A}_{EAS}, \mathcal{A}_{VAS}, \mathcal{A}_{FAS}, \mathcal{A}_{MAS}, \mathcal{R}_{PAS}, \mathcal{R}_{PRAS}, \mathcal{R}_{EAS}, \mathcal{R}_{VAS}, \mathcal{R}_{FAS}, \mathcal{R}_{MAS} \rangle$. $\widehat{\Phi}$ is shown in Fig. 1, where arrows represent instances of **PAtS**, **FAtS**, **PRAtS**, and **VAAtS**; dotted arrows represent instances of **VAtS** and **VDeS**; single dotted-dashed arrows represent instances of **EAtS**; and double dotted-dashed arrows represent instances of **MAtS**.

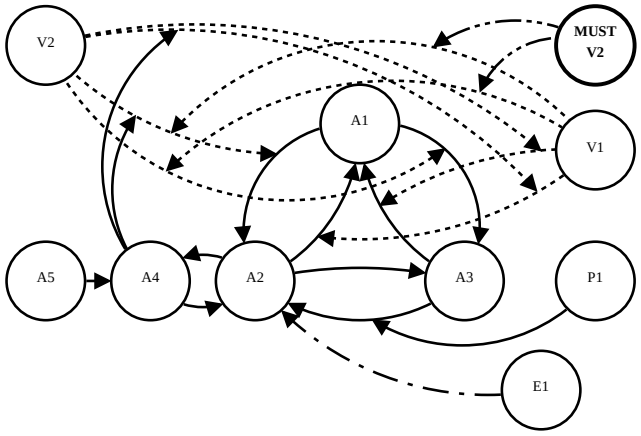

Figure 1: A graphical representation of $\widehat{\Phi}$

We turn now to the problem of providing a formal backing to the proposed representation in order to support the computation of the decision to be suggested.

Computing the outcome of the decision process

According to our approach, once a Decision Support Problem has been represented using the concepts defined in the previous section, we can compute the relevant decision outcomes relying on the formal notion of extensions of an argumentation framework. To this purpose, we rely on a new formalism called $AFRA$ (Baroni et al. 2009b; 2010) which extends Dung's AF by allowing attacks to attacks in a recursive way. We briefly recall in the following the main notions of AFRA.

Definition 3 ($AFRA$)**.** *An Argumentation Framework with Recursive Attacks ($AFRA$) is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$ where $\mathcal{A}$ is a set of arguments and $\mathcal{R}$ is a set of attacks, namely pairs $(A, \mathscr{X})$ s.t. $A \in \mathcal{A}$ and ($\mathscr{X} \in \mathcal{R}$ or $\mathscr{X} \in \mathcal{A}$).*
Given an attack $\alpha = (A, \mathscr{X}) \in \mathcal{R}$, we will say that A is the source of α, denoted as $src(\alpha) = A$ and $\mathscr{X}$ is the target of α, denoted as $trg(\alpha) = \mathscr{X}$.

We start substantiating the role played by attacks by introducing a notion of defeat which regards attacks, rather than their source arguments, as the subjects able to defeat arguments or other attacks, as encompassed by Definition 4.

Definition 4 (Direct Defeat)**.** *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an $AFRA$, $\mathscr{V} \in \mathcal{R}$, $\mathscr{W} \in \mathcal{A} \cup \mathcal{R}$, then $\mathscr{V}$ directly defeats $\mathscr{W}$ iff $\mathscr{W} = trg(\mathscr{V})$.*

Moreover, since we are interested also in how attacks are affected by other attacks, we introduce a notion of indirect defeat for an attack, corresponding to the situation where its source receives a direct defeat.

Definition 5 (Indirect Defeat)**.** *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an $AFRA$, $\mathscr{V} \in \mathcal{R}$, $\mathscr{W} \in \mathcal{A}$, if $\mathscr{V}$ directly defeats $\mathscr{W}$ then $\forall \alpha \in \mathcal{R}$ s.t. $src(\alpha) = \mathscr{W}$, $\mathscr{V}$ indirectly defeats α.*

A defeat is a direct or indirect defeat.

Definition 6 (Defeat)**.** *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an $AFRA$, $\mathscr{V} \in \mathcal{R}$, $\mathscr{W} \in \mathcal{A} \cup \mathcal{R}$, then $\mathscr{V}$ defeats $\mathscr{W}$, denoted as $\mathscr{V} \rightarrow_R \mathscr{W}$, iff $\mathscr{V}$ directly or indirectly defeats $\mathscr{W}$.*

The definition of conflict-free set follows directly.

Definition 7 (Conflict–free)**.** *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an $AFRA$, $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$ is conflict–free iff $\nexists \mathscr{V}, \mathscr{W} \in \mathcal{S}$ s.t. $\mathscr{V} \rightarrow_R \mathscr{W}$.*

The definition of acceptability is similar to the traditional one, but involves both arguments and attacks.

Definition 8 (Acceptability)**.** *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an $AFRA$, $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$, $\mathscr{W} \in \mathcal{A} \cup \mathcal{R}$, $\mathscr{W}$ is acceptable w.r.t. $\mathcal{S}$ iff $\forall \mathscr{Z} \in \mathcal{R}$ s.t. $\mathscr{Z} \rightarrow_R \mathscr{W}$ $\exists \mathscr{V} \in \mathcal{S}$ s.t. $\mathscr{V} \rightarrow_R \mathscr{Z}$.*

On this basis, also the definitions of *admissible set* and *preferred extension* are analogous to the traditional ones.

Definition 9 (Admissible set - Preferred Extension)**.** *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an $AFRA$, $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{R}$ is admissible iff it is conflict–free and each element of $\mathcal{S}$ is acceptable w.r.t. $\mathcal{S}$. A preferred extension is a maximal (w.r.t. set inclusion) admissible set.*

We propose a natural correspondence from an $AFDSP$ to an $AFRA$: the instances of argument schemes in $AFDSP$ compose the set of arguments in $AFRA$ and the instances of attack schemes in $AFDSP$ give rise to the attack relation in $AFRA$. Formally, letting $\Phi = \langle \mathcal{A}_{PAS}, \mathcal{A}_{PRAS}, \mathcal{A}_{EAS}, \mathcal{A}_{VAS}, \mathcal{A}_{FAS}, \mathcal{A}_{MAS}, \mathcal{R}_{PAS}, \mathcal{R}_{PRAS}, \mathcal{R}_{EAS}, \mathcal{R}_{VAS}, \mathcal{R}_{FAS}, \mathcal{R}_{MAS} \rangle$ be an $AFDSP$, the corresponding $AFRA$ is defined as $\Gamma = \langle \mathcal{A}, \mathcal{R} \rangle$ s.t. $\mathcal{A} = \mathcal{A}_{PAS} \cup \mathcal{A}_{PRAS} \cup \mathcal{A}_{EAS} \cup \mathcal{A}_{VAS} \cup \mathcal{A}_{FAS} \cup \mathcal{A}_{MAS}$; and $\mathcal{R} = \mathcal{R}_{PAS} \cup \mathcal{R}_{PRAS} \cup \mathcal{R}_{EAS} \cup \mathcal{R}_{VAS} \cup \mathcal{R}_{FAS} \cup \mathcal{R}_{MAS}$.

In order to keep a correspondence between our approach and the one by (Atkinson, Bench-Capon, and McBurney 2006), we exploit the notion of preferred extension to define the outcomes of the decision process. In general, adopting a multiple status semantics is compatible with the idea that several alternative courses of actions may be considered and that the decision maker is encouraged to evaluate and criticize the advice provided by the system. More precisely, every preferred extension of an $AFRA$ is a set of arguments and attacks which can be regarded altogether as a reasonable and defendable position. The instance(s) of practical arguments included in a preferred extension correspond to the

suggested action(s). In general, several distinct preferred extensions may exist, corresponding to alternative and equally defendable courses of actions.

To conclude the discussion concerning John's dilemma, given $\widehat{\Phi}$, we directly obtain its corresponding $AFRA\,\widehat{\Gamma}$. The arguments included in its (unique) preferred extension are: {**MUST V2**, **V2**, **V1**, **P1**, **E1**, **A5**, **A3**}. This corresponds to suggesting the actions of having a long non-invasive treatment, as expected. We omit the relevant formal derivation due to space limitations. As a final remark concerning a different scenario, note that if we would have assumed a different **MAS** argument concerning charity, we would have obtained a different outcome, with **A1** accepted and the "take analgesics" action supported.

Arguments and Personality

The described computational model of decision making may have a conceptual counterpart in the context of Freud's model of personality. Indeed we modelled human reasoning as common arguments used in everyday discourse, as well as in special contexts like medical argumentation, and we structured them in form of argumentation schemes. In (Baroni et al. 2009a) we proposed several argument schemes (extended in this work), which, taking into account the goal of modelling personality, can be gathered into three main sets:

1. the set of arguments that take into account values;
2. the set of arguments whose bases are not rational (e.g. emotions);
3. the set of arguments that deal with both reality and values.

The first set includes **VAS** arguments. In the example, two values are considered: the value of human health (safety) and the value of charity.

EAS arguments belongs to the second set since they aim at representing how human emotions can affect practical reasoning by favouring or not some particular actions. In John's problem, he is frightened by surgeries, therefore he does not agree to undergo surgery even if he rationally believes that this course of action is the most preferable for achieving the goal of reducing disc herniation.

The last set considers each argument having to deal with:

- the reality as it is perceived, and relevant reasoning aimed to derive some consequences from observation, or to suggest an action on the basis of its effects in the real world;
- the values, by considering carefully the pros and cons of each issue and how it affects other aspects of the reasoning.

Considering the example, elements of this set are the **FAS**, **PAS**, **PRAS** and **MAS** arguments. In particular **FAS** arguments correspond to our perception of the world; **PAS** arguments deal with actions to be taken concerning the real world situation; **PRAS**s represent rational preference, therefore they still refer to considerations about reality (in the particular case of the example, the effect of practical decisions on reality). Finally, **MAS**s represent a balancing between values and reality requirements.

From these considerations, we can classify these sets as:

1. the set of Superego arguments (that take into account values);
2. the set of Id arguments (without rational basis);
3. the set of Ego arguments (that deal with both reality and values);

Figure 2 evidences these classes of arguments for the example.

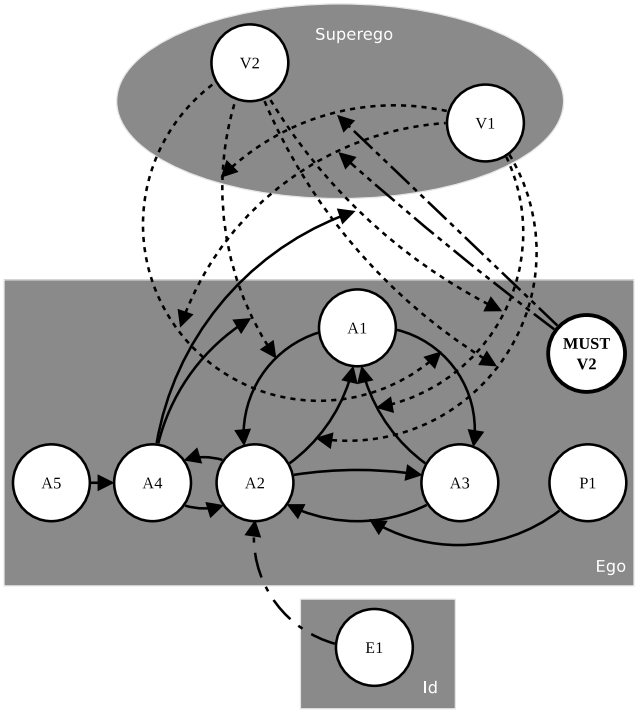

Figure 2: $\widehat{\Gamma}$ as a modelling tool for personality components.

Finally, in our formalisation, we considered not only arguments but also attacks as fundamental components of the argumentation process. Indeed, in (Baroni et al. 2009a) we proposed the notion of Attack Scheme as a way to formalise the reasoning leading to a conclusion of the form "The argument A attacks the argument B". This allows to explain the reasons for a conflict that other approaches may not explicitly consider because they refer to a notion of inconsistency at the syntactic level of arguments' underlying logic. Moreover, considering attacks as fundamental components of the argumentation process leads to associate a justification status to attacks too. In fact, as shown in the example, in our formalism attacks can be attacked by other attacks, and attacks on attacks are, in turn, subjected to justification evaluation.

Considering the classification into three sets, attacks refer to Superego and Ego. Indeed, conflicts between some values or other issues may be represented as attacks and pertain to Superego. The remaining kinds of attacks represent the result of a rational process which pertains to Ego (which evaluates if a factual circumstance is actually conflicting with the required condition in a practical argument).

Preprint version. Official version may differ.

Discussion and conclusions

In this paper we extended the proposal by (Baroni et al. 2009a) concerning the representation of decision making problems through an argumentation-based approach by taking into account human emotions in the decision making process. The main contribution of this paper is the preliminary description of a mapping between an argumentation-based decision making process and Freud's three-entities model of personality. In particular we identified some classes of arguments and attacks generated during the decision making process, and put them in correspondence with Ego, Superego and Id entities.

As to related works, (Atkinson, Bench-Capon, and McBurney 2006) is focused on an approach to practical reasoning, regarded as presumptive justification of a course of action, based on the answers to a set of critical questions. An interaction protocol (PARMA) for agent dialogue about the proposed action is also provided. This approach adopts an argument scheme, along with relevant critical questions, where relations among practical arguments, goals and values are represented in a structured way. Encompassing attacks to attacks is however not considered in (Atkinson, Bench-Capon, and McBurney 2006), which adopts the Value-based Argumentation Framework. Considering emotions in decision making contexts is a relatively new issue in argumentation theory. Recently (Nawwab, Dunne, and Bench-Capon 2010) describes how emotions can trigger replanning of a scheduled course of action through a change in value ordering.

The present work is at an early stage of development and is far from being unquestionable and complete. First of all, this paper considers attacks schemes, showing the role that they may play (an issue which, as to our knowledge, has received so far only limited attention in the literature) and demonstrating that different levels of attacks to attacks are useful to capture some intuitive patterns of practical reasoning thus emphasizing the role of the $AFRA$ formalism. The three-sets approach, based on Freud's model, still lacks details concerning Superego and Id. Future works should fill this gap enhancing the relevant argument schemes and might consider also other more articulated personality models.

References

Atkinson, K.; Bench-Capon, T. J. M.; and McBurney, P. 2006. Computational representation of practical argument. *Synthese* 152(2):157–206.

Baroni, P.; Cerutti, F.; Giacomin, M.; and Guida, G. 2009a. An argumentation-based approach to modeling decision support contexts with what-if capabilities. In *AAAI Fall Symposium. Technical Report SS-09-06*, 2–7. AAAI Press.

Baroni, P.; Cerutti, F.; Giacomin, M.; and Guida, G. 2009b. Encompassing attacks to attacks in abstract argumentation frameworks. In *Proc. of ECSQARU 2009*, 83–94.

Baroni, P.; Cerutti, F.; Giacomin, M.; and Giovanni, G. 2010. AFRA: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning* In Press.

Bench-Capon, T. J. M., and Modgil, S. 2008. Integrating object and meta-level value based argumentation. In *Proc. of COMMA 2008*, 240–251.

Bench-Capon, T. J. M. 2003. Persuasion in practical argument using value based argumentation frameworks. *J. of Logic and Computation* 13(3):429–448.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. *AIJ* 77(2):321–357.

Freud, S. 1923. *Das Ich und das Es*.

Freud, S. 1933. *New Introductory Lectures on Psycho-Analysis*.

Freud, S. 1940. *Abriß der Psychoanalyse*.

Girle, R.; Hitchcock, D. L.; McBurney, P.; and Verheij, B. 2003. Decision support for practical reasoning: A theoretical and computational perspective. In Reed, C., and Norman, T. J., eds., *Argumentation Machines. New Frontiers in Argument and Computation*, 55–84. Kluwer.

Modgil, S. 2007. An abstract theory of argumentation that accommodates defeasible reasoning about preferences. In *Proc. of ECSQARU 2007*, 648–659.

Modgil, S. 2009. Reasoning about preferences in argumentation frameworks. *AIJ* 173(9-10):901–934.

Nawwab, F. S.; Dunne, P. E.; and Bench-Capon, T. 2010. Exploring the role of emotions in rational decision making. In Baroni, P.; Cerutti, F.; Giacomin, M.; and Simari, G., eds., *Computational Models of Argument - Proceedings of COMMA 2010*. IOS Press.

Walton, D. N. 1996. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates.