# Comparing decisions on the basis of a bipolar typology of arguments

**Leila Amgoud** and **Henri Prade**

Institut de Recherche en Informatique de Toulouse
118, route de Narbonne, 31062 Toulouse, France
{amgoud, prade}@irit.fr

**Abstract**

Arguments play two types of roles w.r.t. decision, namely helping to select an alternative, or to explain a choice. Until now, the various attempts at formalizing argument-based decision making have relied only on one type of arguments, in favor of or against alternatives.

The paper proposes a systematic typology that identifies eight types of arguments, some of them being weaker than others. First the setting emphasizes the *bipolar* nature of the evaluation of decision results by making an explicit distinction between prioritized *goals* to be pursued, and prioritized *rejections* that are stumbling blocks to be avoided. This is the basis for an argumentative framework for decision. Each decision is *supported* by arguments emphasizing its positive consequences in terms of goals certainly satisfied, goals possibly satisfied, rejections certainly avoided and rejections possibly avoided. A decision can also be *attacked* by arguments emphasizing its negative consequences in terms of certainly or possibly missed goals, or rejections certainly or possibly led to by that decision. The logical properties of this framework are studied. The richness of the proposed typology makes it possible to partition the set of alternatives into four classes, giving thus a *status* to decisions, which may be *recommended*, *discommended*, *controversial* or *neutral*. Each class may be refined into sub-classes taking advantage of the strengths of the different types of arguments. This typology is also helpful from an explanation point of view for being able to use the right type of arguments depending on the context.

The paper also presents a preliminary investigation on *decision principles* that can be used for comparing decisions. Three classes of principles can be considered: *unipolar*, *bipolar* or *non-polar* principles depending on whether i) only arguments pro or only arguments cons, or ii) both types, or iii) an aggregation of them into a meta-argument are used.

**Key words:** Decision making, Argumentation

## Introduction

Decision making, often viewed as a form of reasoning toward action, has raised the interest of many scholars including philosophers, economists, psychologists, and computer scientists for a long time. Any decision problem amounts to select the best option(s) among different alternatives.

The decision problem has been considered from different points of view. *Classical* decision theory, as developed by economists, has focused mainly on identifying *criteria* for comparing different alternatives. The inputs of this approach are a set of *feasible* actions, and a function that assesses the value of their consequences when the actions are performed in a given state. The output is a preference relation between actions. A decision criterion, such as the classical expected utility (Savage 1954), should then be justified on the basis of a set of postulates to which the preference relation between action should obey. Note that such an approach considers a group of candidate actions as a whole rather than focusing on a candidate action individually. Moreover, the candidate actions are supposed to be feasible.

More recently, some researchers in AI have advocated the need for a new approach in which the different aspects that may be involved in a decision problem (such as the goals of the agent, the feasibility of an action, its consequences, the conflicts between goals, the alternative plans for achieving the same goal, etc) can be handled. In (Bratman 1987; Bratman, Israel, & Pollack 1988), it has been argued that this can be done by representing the cognitive states, namely agent's beliefs, desires and intentions (thus the so-called *BDI* architecture). The decision problem is then to select among the conflicting desires a consistent and feasible subset that will constitute the intentions. The above line of research takes its inspiration in the work of philosophers who have advocated practical reasoning (Raz 1978). Practical reasoning mainly deals with the adoption, filling in, and reconsideration of intentions and plans. Moreover, it allows reasoning about individual actions using for instance the well-known *practical syllogism* (Walton 1996):

- G is a goal fir agent a
- Doing action A is sufficient for agent a to carry out goal G
- Then, agent a ought to do action A

In this setting, a candidate action may be rejected because it may lead to the violation of other important goals, or to other bad consequences, etc.

In this paper, we are concerned with an argumentative counterpart of classical decision theory. Humans use arguments for supporting, attacking or explaining choices. Indeed, each potential choice has usually pros and cons of various strengths. Adopting such an approach in a decision support system would have some obvious benefits. On the one hand, not only would the user be provided with a "good" choice, but also with the reasons underlying this recommendation, in a format that is easy to grasp. On the other hand, argumentation-based decision making is expected to be more akin with the way humans deliberate and finally make or understand a choice. This argumentative view of decision has not been much considered until recently even if the idea of basing decisions on arguments pros and cons is very old and was already somewhat formally stated by Benjamin Franklin (Franklin 1887) more than two hundred years ago.

Articulating decisions on the basis of arguments is relevant for different decision problems or approaches such as decision under uncertainty, multiple criteria decisions, or rule-based decisions. For instance, in medical domains, decisions are usually to be made under incomplete or uncertain information, and the potential results of candidate decisions may be evaluated from different criteria. Moreover, there may exist some expertise under the form of decision rules that associate possible decisions with given contexts. Thus, the different types of decision problems interfere, and consequently a unified argumentation-based model may be still more worth developing.

Whatever the decision problem is, the basic idea is that candidate decisions may lead to positively or negatively assessed results. This gives birth to arguments in favor of (pros) or against (cons) a decision in a given context. Different attempts at formalizing argument-based decision making can be found in the literature (Fox & Das 2000; Fox, Krause, & Ambler 1992; Bonet & Geffner 1996; Brewka & Gordon 1994; Amgoud & Prade 2004; Dubois & Fargier 2005; Amgoud, Bonnefon, & Prade 2005). These works do not much discuss the nature of arguments in a decision analysis, and usually rely on one type of argument that may be in favor of or against alternatives.

This paper emphasizes the *bipolar* nature of the evaluation of decision results, by making an explicit distinction between *goals* having a positive flavor, and rejections, with a negative flavor, that are stumbling blocks to be avoided. This, for instance, applies to criteria scales where the positive grades (associated with positive results) are separated from the negative ones (associated with negative results) by one or several neutral values.

The paper proposes a systematic typology that identifies eight types of arguments. Some of them are weaker than others, since they rather reflect the existence of examples or counter-examples as supporting or challenging possible choices. In the proposed framework, each decision is *supported* by arguments emphasizing its positive features in terms of goals certainly satisfied, goals possibly satisfied, rejections certainly avoided and rejections possibly avoided. The possibility that a goal may be reached or that a rejection may be avoided is assessed in practice by the existence of relevant known examples. A decision can also be *attacked* by arguments emphasizing its negative features in terms of certainly or possibly missed goals, or rejections certainly or possibly led to by that decision. The richness of the proposed typology makes it possible to partition the set of alternatives into four classes, giving thus a *status* to decisions, which may be *recommended*, *discommended*, *controversial* or *neutral*. Each class may be refined into sub-classes taking advantage of the strengths of the different types of arguments.

The aim of this paper is also to present a general discussion and a first study of the different classes of argument-based decision principles. In the following, we argue that three main classes of principles can be distinguished:

1. *Unipolar principles* that focus only on one type of arguments when comparing choices (either arguments pros or arguments cons)
2. *Bipolar principles* that take into account both types of arguments but still keeping the distinction between the two types
3. *Non-polar principles* that consist of aggregating the two types of arguments into a meta-argument and compare pairs of choices on the basis of their meta-arguments.

Note that, the use of suffix "polar" here refers to the dichotomy between arguments pros and arguments cons (and not to the bipolar structure induced by goals and rejections).

## A general framework for decision making

Solving a decision problem amounts to defining a pre-ordering, usually a complete one, on a set $\mathcal{D}$ of possible choices (or decisions), on the basis of the different consequences of each decision. Argumentation can be used for defining such a pre-ordering. An argumentation-based decision process can be decomposed into the following steps:

1. Constructing arguments in *favor*/*against* each decision in $\mathcal{D}$.
2. Evaluating the strength of each argument.
3. Comparing decisions on the basis of their arguments.
4. Defining a pre-ordering on $\mathcal{D}$.

In (Amgoud, Bonnefon, & Prade 2005), an argumentation-based decision framework is defined as follows:

**Definition 1 (Argumentation-based decision framework)** *An* argumentation-based decision framework *is a tuple* $<\mathcal{D}, \mathcal{A}, \succeq, \triangleright_{Princ}>$ *where:*

- $\mathcal{D}$ *is a set of all possible choices.*
- $\mathcal{A}$ *is a set of arguments.*
- $\succeq$ *is a (partial or complete) pre-ordering on* $\mathcal{A}$.
- $\triangleright_{Princ}$ *(for principle for comparing decisions), defines a (partial or complete) pre-ordering on* $\mathcal{D}$*, defined on the basis of* $\mathcal{A}$.

*The* output *of the framework is a (complete or partial) pre-ordering* $\triangleright_{Princ}$*, on* $\mathcal{D}$. $d_1 \triangleright_{Princ} d_2$ *means that the decision* $d_1$ *is* at least as preferred as *the decision* $d_2$ *w.r.t. the* principle $Princ$.

**Notation:** Let $A$, $B$ be two arguments of $\mathcal{A}$, and $\succeq$ be a preorder (maybe partial) on $\mathcal{A}$. $A \succeq B$ means that $A$ is at least as 'strong' as $B$.
$\succ$ and $\approx$ will denote respectively the strict ordering and the relation of equivalence associated with the preference between arguments, defined as follows:

- $A \succ B$ iff $A \succeq B$ and not ($B \succeq A$) (meaning that $A$ is strictly stronger than $B$),
- $A \approx B$ iff $A \succeq B$ and $B \succeq A$ (meaning that $A$ is as strong as $B$).

Different definitions of $\succeq$ or different definitions of $\rhd_{Princ}$ may lead to different decision frameworks that may not return the same results.

## Logical language

In what follows, let $\mathcal{L}$ be a propositional language. From $\mathcal{L}$ we can distinguish the four following sets:

1. The set $\mathcal{D}$ gathers all the possible *alternatives*, or *decisions*. These candidate actions are assumed to be feasible. Elements of $\mathcal{D}$ are supposed to be represented by literals.
2. The set $\mathcal{K}$ represents the *background knowledge* that is assumed to be consistent. Elements of $\mathcal{K}$ are formulas of $\mathcal{L}$.
3. The set $\mathcal{G}$ gathers the *goals* of an agent. A goal represents what the agent wants to achieve, and has thus a positive flavor. This base is assumed to be consistent too, i.e. an agent is not allowed to have contradictory goals. Note that a goal may be expressed in terms of a logical combination of constraints on criteria values, and does not necessarily refer to one criterion. Elements of $\mathcal{G}$ are supposed to be literals.
4. The set $\mathcal{R}$ gathers the *rejections* of an agent. A rejection represents what the agent wants to avoid. Clearly rejections express negative preferences. The set $\{\neg r | r \in \mathcal{R}\}$ is assumed to be consistent since acceptable alternatives should satisfy $\neg r$ due to the rejection of $r$. However, note that if $r$ is a rejection, this does not necessarily mean that $\neg r$ is a goal. For instance, in case of choosing a medical drug, one may have as a goal the immediate availability of the drug, and as a rejection its availability only after at least two days. As it can be guessed on this example, if $g$ is a goal only $r$ such that $r \vdash \neg g$ can be a rejection, and conversely. This means that rejection can be more specific than the negation of goals. Moreover, recent cognitive psychology studies (Cacioppo, Gardner, & Bernston 1997) have confirmed the cognitive validity of this distinction between goals and rejections. Elements of $\mathcal{R}$ are supposed to be literals.

**Definition 2** *A* decision problem *is a tuple* $\mathcal{T} = \langle\mathcal{D}, \mathcal{K}, \mathcal{G}, \mathcal{R}\rangle$.

## A new typology of arguments

When solving a decision problem, there may exist several alternative solutions. Each alternative may have arguments in its favor (called PROS), and arguments against it (called CONS). In the following, an argument is associated with an alternative, and always either refers to a goal or to a rejection.

Arguments PROS point out the existence of good consequences or the absence of bad consequences for a given alternative. More precisely, we can distinguish between two types of good consequences, namely the guaranteed satisfaction of a goal when $\mathcal{K} \cup \{d\} \vdash g$, and the possible satisfaction of a goal when $\mathcal{K} \cup \{d\} \nvdash \neg g$, with $d \in \mathcal{D}$ and $g \in \mathcal{G}$. Note that this latter situation corresponds to the existence of an interpretation that satisfies $\mathcal{K}$, $d$, and $g$. This leads to the following definition:

**Definition 3 (Types of positive arguments PRO)** *Let* $\mathcal{T}$ *be a decision problem. A* positively expressed argument in favor of *an alternative* $d$ *is a pair* $A = \langle d, g\rangle$ *such that:*

1. $d \in \mathcal{D}$, $g \in \mathcal{G}$, $\mathcal{K} \cup \{d\}$ *is consistent*
2. - $\mathcal{K} \cup \{d\} \vdash g$ *(arguments of Type SPP), or*
   - $\mathcal{K} \cup \{d\} \nvdash \neg g$ *(arguments of Type WPP)*

The consistency of $\mathcal{K} \cup \{d\}$ means that $d$ is applicable in the context $\mathcal{K}$, in other words that we cannot prove from $\mathcal{K}$ that $d$ is impossible. This means that impossible alternatives w.r.t. $\mathcal{K}$ have been already taken out when defining the set $\mathcal{D}$.
Note that SPP arguments are stronger than WPP ones since ($\mathcal{K} \cup \{d\} \vdash g$) entails ($\mathcal{K} \cup \{d\} \nvdash \neg g$). SPP stands for "Strong Positive PROS", whereas WPP is short for "Weak Positive PROS". Clearly, WPP arguments have interest only when the corresponding SPP do not exist. From a practical point of view, SPP arguments will be generally expressed by only exhibiting a minimal subset $S$ of $\mathcal{K}$ such that $S \cup \{d\} \vdash g$, while a WPP argument corresponds to the existence of (at least) one known case where $g$ was satisfied while $d$ was applied. This means that together with $\mathcal{K}$, the agent stores a case memory of already experienced choices. Let $Arg_{SPP}(d)$ (resp. $Arg_{WPP}(d)$) be the set of all arguments of type SPP (resp. WPP) in favor of $d$ in the sense of the above definition. The following inclusion holds:

**Property 1** *Let* $d \in \mathcal{D}$. *Then* $Arg_{SPP}(d) \subseteq Arg_{WPP}(d)$.

Similarly, there are two forms of absence of bad consequences that lead to arguments PROS: the first one amounts to avoid a rejection for sure, i.e. $\mathcal{K} \cup \{d\} \vdash \neg r$, and the second form corresponds only to the possibility of avoiding a rejection ($\mathcal{K} \cup \{d\} \nvdash r$), which can be testified in practice by the existence of a case counter-example assuring that $\mathcal{K} \wedge d \wedge \neg r$ is consistent (with $r \in \mathcal{R}$). This leads to the following definition:

**Definition 4 (Types of negative arguments PROS)** *Let* $\mathcal{T}$ *be a decision problem. A* negatively expressed argument in favor of *an alternative is a pair* $A = \langle d, r\rangle$ *such that:*

1. $d \in \mathcal{D}$, $r \in \mathcal{R}$, $\mathcal{K} \cup \{d\}$ *is consistent*
2. - $\mathcal{K} \cup \{d\} \vdash \neg r$ *(arguments of Type SNP), or*
   - $\mathcal{K} \cup \{d\} \nvdash r$ *(arguments of Type WNP)*

Here again, SNP arguments are stronger than WNP ones since $\mathcal{K} \cup \{d\} \vdash \neg r$ entails $\mathcal{K} \cup \{d\} \nvdash r$. SNP stands for

"Strong Negative PROS", while WNP means "Weak Negative PROS". Let $Arg_{SNP}(d)$ (resp. $Arg_{WNP}(d)$) be the set of all arguments of type SNP (resp. WNP) in favor of $d$.

**Property 2** *Let $d \in \mathcal{D}$. Then $Arg_{SNP}(d) \subseteq Arg_{WNP}(d)$.*

Arguments CONS highlight the existence of bad consequences for a given alternative, or the absence of good ones. As in the case of arguments PROS, there are a strong form and a weak form of both situations. Namely, negatively expressed arguments CONS are defined either by exhibiting a rejection that is necessarily satisfied, or a rejection that is possibly satisfied. Formally:

**Definition 5 (Types of negative arguments CONS)** *Let $\mathcal{T}$ be a decision problem. A* negatively expressed argument against *an alternative $d$ is a pair $A = \langle d, g \rangle$ such that:*

1. *$d \in \mathcal{D}$, $r \in \mathcal{R}$, $\mathcal{K} \cup \{d\}$ is consistent*
2. - *$\mathcal{K} \cup \{d\} \vdash r$ (arguments of Type SNC), or*
   - *$\mathcal{K} \cup \{d\} \nvdash \neg r$ (arguments of Type WNC)*

Let $Arg_{SNC}(d)$ (resp. $Arg_{WNC}(d)$) be the set of all arguments of type SNC (resp. WNC) against $d$, where C stands for Cons.

**Property 3** *Let $d \in \mathcal{D}$. Then $Arg_{SNC}(d) \subseteq Arg_{WNC}(d)$.*

Lastly, the absence of positive consequences can also be seen as an argument against (CONS) an alternative. A strong form and a weak form of positively expressed arguments against an alternative can be defined as follows:

**Definition 6 (Types of arguments CONS)** *Let $\mathcal{T}$ be a decision problem. A* positively expressed argument against *an alternative is a pair $A = \langle d, g \rangle$ such that:*

1. *$d \in \mathcal{D}$, $g \in \mathcal{G}$, $\mathcal{K} \cup \{d\}$ is consistent*
2. - *$\mathcal{K} \cup \{d\} \vdash \neg g$ (arguments of Type SPC), or*
   - *$\mathcal{K} \cup \{d\} \nvdash g$ (arguments of Type WPC)*

Let $Arg_{SPC}(d)$ (resp. $Arg_{WPC}(d)$) be the set of all arguments of type SPC (resp. WPC) against $d$.

**Property 4** *Let $d \in \mathcal{D}$. Then $Arg_{SPC}(d) \subseteq Arg_{WPC}(d)$.*

Let us consider positively expressed arguments for instance. Observe that for a given alternative $d$ and a fixed goal $g$, all the types of arguments cannot take place at the same time. Formally,

**Property 5** *SPP and WPC (resp. SPC and WPP, and SPP and SPC) arguments are mutually exclusive.*

In the first two cases, this is due to the opposite characteristic conditions of the definitions. The last exclusion is due to the consistency of $\mathcal{K} \cup \{d\}$, and thus $g$ and $\neg g$ cannot be obtained simultaneously. Taking into account the subsumptions between weak and strong forms of arguments, the following result holds:

**Property 6** *Let $d \in \mathcal{D}$ and $g \in \mathcal{G}$. There are only three possible situations w.r.t a positively expressed argument linking $d$ and $g$, namely: i) there is an SPP argument, ii) there is an SPC argument, iii) there are both an WPP and an WPC arguments.*

The above property reflects the three possible epistemic statuses of a knowledge base ($\mathcal{K} \cup \{d\}$) w.r.t a proposition (here $g$), which may be true, false, or having an unknown truth status. In the latter case, emphasizing either a WPP argument, or a WPC argument is a matter of optimism vs pessimism to which we come back later. Since WPP and WPC arguments are somewhat neutralizing each other, we will not consider them in the decision status classification that we introduce now.

## Decision status

In the previous section, we have shown that each decision may be supported by two types of strong arguments, and attacked by two other types of strong arguments. In summary, given a decision $d \in \mathcal{D}$, we will have the following sets of arguments:

- $Arg_{SPP}(d)$ = those arguments which capture the goals that are reached when applying $d$ in context $\mathcal{C}$
- $Arg_{SNP}(d)$ = those arguments which capture the rejections that are avoided when applying $d$ in context $\mathcal{C}$
- $Arg_{SNC}(d)$ = those arguments which capture the rejections that are not avoided when applying $d$ in context $\mathcal{C}$
- $Arg_{SPC}(d)$ = those arguments which capture the goals that are missed when applying $d$ in context $\mathcal{C}$

Note that when a given set is empty, for instance $Arg_{SPP}(d)$ = $\emptyset$, this does not mean at all that decision $d$ cannot lead to any goal, but rather we cannot be certain that a goal is reached as some information is missing. The above types of arguments supporting or attacking a choice $d$ give birth to four main different *statuses* for that decision: *recommended*, *discommended*, *neutral* and *controversial* (see Table 1 for the straightforward formal definitions). Recommended choices are those choices that have only arguments in favor of them and no arguments against them (whatever their type). Discommended choices are those choices that have no arguments in favor of them and only arguments against them. Regarding neutral choices, they have neither arguments in favor of them, nor arguments against. Choices that have at the same time arguments in favor of them and arguments against are said controversial. As shown in Table 1, there are 9 situations in which a choice is controversial.

**Property 7** *Let $d \in \mathcal{D}$. Then $d$ is either fully recommended, or fully discommended, or controversial or neutral.*

Note that one may give priority to SPP and SNC arguments, which directly state that a goal is reached, or a rejection is not missed respectively. SNP and SPC arguments that are in favor of or against a choice only indirectly can be used for refining the two first types of arguments. For instance, in Table 1, a choice that has both SPP and SNP arguments in favor of it is strongly recommended, whereas a choice with only SPP arguments in favor of it is only recommended.

This classification of choices is of interest from a persuasion or explanation perspective, since e.g recommended choices are more easily arguable than controversial ones. This does not mean that recommended choices are always better than any other as we shall see in the next section.

| Status | Sub-status | Combination |
|---|---|---|
| Recommended | Strongly recom. | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) \neq \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) = \emptyset \rangle$ |
| | Recom. | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) = \emptyset \rangle$ |
| | Weakly recom. | $\langle Arg_{SPP}(d) = \emptyset, Arg_{SNP}(d) \neq \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) = \emptyset \rangle$ |
| Discommended | Strongly discom. | $\langle Arg_{SPP}(d) = \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) \neq \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |
| | Discom. | $\langle Arg_{SPP}(d) = \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) \neq \emptyset, Arg_{SPC}(d) = \emptyset \rangle$ |
| | Weakly discom. | $\langle Arg_{SPP}(d) = \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |
| Neutral | | $\langle Arg_{SPP}(d) = \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) = \emptyset \rangle$ |
| Controversial | | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) \neq \emptyset, Arg_{SNC}(d) \neq \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |
| | | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) \neq \emptyset, Arg_{SNC}(d) \neq \emptyset, Arg_{SPC}(d) = \emptyset \rangle$ |
| | | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) \neq \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |
| | | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) \neq \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |
| | | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) \neq \emptyset, Arg_{SPC}(d) = \emptyset \rangle$ |
| | | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |
| | | $\langle Arg_{SPP}(d) = \emptyset, Arg_{SNP}(d) \neq \emptyset, Arg_{SNC}(d) \neq \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |
| | | $\langle Arg_{SPP}(d) \neq \emptyset, Arg_{SNP}(d) = \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |
| | | $\langle Arg_{SPP}(d) = \emptyset, Arg_{SNP}(d) \neq \emptyset, Arg_{SNC}(d) = \emptyset, Arg_{SPC}(d) \neq \emptyset \rangle$ |

Table 1: Decision status

## Principles for comparing decisions

In the following we are interested in discussing possible choices for $\rhd_{Princ}$. The relation $\succeq$ is assumed to be given. It may be either a partial or a complete preorder. This preorder may account for the certainty of the pieces of knowledge involved in the argument and/or to the importance of the goal to which the argument pertains. However, this will not be detailed in the following.

Comparing choices on the basis of the sets of PROS or CONS arguments that are associated with them is a key step in an argumentative decision process. Depending on what sets are considered and how they are handled, one can roughly distinguish between three categories of principles:

**Unipolar principles:** are those that only refer to either the arguments PROS or the arguments CONS.

**Bipolar principles:** are those that reason on both types of arguments at the same time.

**Non-polar principles:** are those where arguments PROS and arguments CONS a given choice are aggregated into a unique *meta-argument*. It results that the negative and positive polarities disappear in the aggregation.

Below we present the main principles that can be thought of for each category. In what follows, $Arg_{Pro}(d)$ = $Arg_{SPP}(d) \cup Arg_{SNP}(d)$, and $Arg_{Cons}(d)$ = $Arg_{SNC}(d)$ $\cup$ $Arg_{SPC}(d)$. Moreover, the function `Result` returns for a given set of arguments, all the goals/rejections involved in those arguments.

## Unipolar principles

In this section we present basic criteria for comparing decisions on the basis of only arguments PROS. Note that similar ideas apply to arguments CONS.

We start by presenting those criteria that do not involve the strength of arguments, then their respective refinements when strength is taken into account.

A first natural criterion consists of preferring the decision $d_1$ over $d_2$ if for each argument $<d_2, g>$, there exists an argument $<d_1, g>$, while the reverse is not true. Formally:

**Definition 7** *Let* $d_1$, $d_2 \in \mathcal{D}$.
$d_1 \rhd d_2$ *iff* $\texttt{Result}(Arg_{Pro}(d_2)) \subseteq \texttt{Result}(Arg_{Pro}(d_1))$.

This *partial* preorder is refined by the following *complete* preorder in terms of cardinality, i.e preferring the decision that has more arguments PROS.

**Definition 8 (Counting arguments PROS)** *Let* $d_1$, $d_2 \in \mathcal{D}$.
$d_1 \rhd d_2$ *iff* $|Arg_{Pro}(d_1)| \geq |Arg_{Pro}(d_2)|$.

When the strength of arguments is taken into account in the decision process, one may think of preferring a choice that has a dominant argument, i.e. an argument PROS that is preferred to any argument PROS the other choices.

**Definition 9** *Let* $d_1$, $d_2 \in \mathcal{D}$.
$d_1 \rhd d_2$ *iff* $\exists\ A \in Arg_{Pro}(d_1)$ *such that* $\forall\ B \in Arg_{Pro}(d_2)$, $A \succeq B$.

The above definition relies heavily on the relation $\succeq$ that compares arguments. Thus, the properties of this criterion depends on those of $\succeq$. Namely, it can be checked that the above criterion works properly only if $\succeq$ is a complete preorder.

**Property 8** *If the relation* $\succeq$ *is a complete preorder, then* $\rhd$ *is also a complete preorder.*

Note that the above relation may be found to be too restrictive, since when the strongest arguments in favor of $d_1$ and $d_2$ have equivalent strengths (in the sense of $\approx$), $d_1$ and $d_2$ are also seen as equivalent. However, we can refine the above definition by ignoring the strongest arguments with equal strengths, by means of the following *strict preorder*.

**Definition 10** *Let $d_1$, $d_2 \in \mathcal{D}$, and $\succeq$ a complete preorder. Let ($P_1$, ..., $P_r$), ($P'_1$, ..., $P'_s$) be the vectors of arguments PROS the decisions $d_1$ and $d_2$ respectively. Each of these vectors is assumed to be decreasingly ordered w.r.t $\succeq$ (e.g. $P_1 \succeq \ldots \succeq P_r$). Let $v$ = min(r, s).*
*$d_1 \rhd d_2$ iff:*

- *$P_1 \succ P'_1$, or*
- *$\exists\, k \leq v$ such that $P_k \succ P'_k$ and $\forall\, j < k$, $P_j \approx P'_j$, or*
- *$r > v$ and $\forall\, j \leq v$, $P_j \approx P'_j$.*

Note that in all the above criteria, the two types of arguments PROS are considered as having the same importance. Thus, reasoning with goals is as important as reasoning with rejections. However, this may be debatable, since one may prefer arguments ensuring that a goal is reached to an argument that shows that a rejection is avoided, since the latter is the least thing that can be expected. On the basis of this new source of preference between arguments, the above criteria can be further reformulated by processing separately the two sets of arguments. More precisely, we can apply the above definitions only for the set of arguments of type SPP, and only in case of ties to apply again the same definitions on SNP arguments. We may even a different criterion on the set SNP of arguments.

Another point that is worth discussing is the impact of possible dependencies between goals (or rejections) on the decision criteria. Namely, assume that two goals are, for instance, redundant, i.e they are logically equivalent giving $\mathcal{K}$. Applying the cardinality-based criterion may lead to privilege decisions reaching redundant goals. This maybe debatable although allowing for redundancy is clearly a way of stressing the importance of a goal (or a rejection). Note that not all the above criteria are sensible to redundancy, for instance the first one (Definition 7).

Till now, we have only discussed decision criteria based on arguments PROS. However, the counterpart criteria when arguments CONS are considered can also be defined. Thus, the counterpart criterion of the one defined in Definition 7 is the following partial preorder:

**Definition 11** *Let $d_1$, $d_2 \in \mathcal{D}$. $d_1 \rhd d_2$ iff $\texttt{Result}(Arg_{Cons}(d_1)) \subseteq \texttt{Result}(Arg_{Cons}(d_2))$.*

Similarly, it refinement in terms of cardinality is given by the following complete preorder:

**Definition 12 (Counting arguments CONS)** *Let $d_1$, $d_2 \in \mathcal{D}$.*
*$d_1 \rhd d_2$ iff $|Arg_{Cons}(d_1)| \leq |Arg_{Cons}(d_2)|$.*

The criteria that take into account the strengths of arguments have also their counterparts when handling arguments CONS.

**Definition 13** *Let $d_1$, $d_2 \in \mathcal{D}$.*
*$d_1 \rhd d_2$ iff $\exists\ B \in Arg_{Cons}(d_2)$ such that $\forall\ A \in Arg_{Cons}(d_1)$, $B \succeq A$.*

As in the case of arguments PROS, when the relation $\succeq$ is a complete preorder, the above relation is also a complete preorder, and can be refined into the following strict one.

**Definition 14** *Let $d_1$, $d_2 \in \mathcal{D}$.*
*Let ($C_1$, ..., $C_r$), ($C'_1$, ..., $C'_s$) be the vectors of arguments CONS the decisions $d_1$ and $d_2$. Each of these vectors is assumed to be decreasingly ordered w.r.t $\succeq$ (e.g. $C_1 \succeq \ldots \succeq C_r$). Let $v$ = min(r, s).*
*$d_1 \succ d_2$ iff:*

- *$C'_1 \succ C_1$, or*
- *$\exists\, k \leq v$ such that $C'_k \succ C_k$ and $\forall\, j < k$, $C_j \approx C'_j$, or*
- *$v < s$ and $\forall\, j \leq v$, $C_j \approx C'_j$.*

Finally, it may be also worth distinguishing between SNC and SPC arguments, and to privilege those which are SNC since they are the most striking ones. Similar ideas given in the case of arguments PROS apply.

### Bipolar principles

Let's now define some principles where both types of arguments (PROS and CONS) are taken in account when comparing decisions. Generally speaking, we can conjunctively combine the criteria dealing with arguments PROS with their counterpart handling arguments CONS. For instance, the criterion given in Definition 8 can be combined with that given in Definition 12 into the following one:

**Definition 15** *Let $d_1$, $d_2 \in \mathcal{D}$.*
*$d_1 \rhd d_2$ iff*

1. *$|Arg_{Pro}(d_1)| \geq |Arg_{Pro}(d_2)|$, and*
2. *$|Arg_{Cons}(d_1)| \leq |Arg_{Cons}(d_2)|$.*

However, note that unfortunately this is no longer a complete preorder. Similarly, the criteria given respectively in Definition 9 and Definition 13 can be combined into the following one:

**Definition 16** *Let $d_1$, $d_2 \in \mathcal{D}$.*
*$d_1 \rhd d_2$ iff:*

- *$\exists\ A \in Arg_{pros}(d_1)$ such that $\forall\ B \in Arg_{pros}(d_2)$, $A \succeq B$.*
- *$\nexists$ A' $\in Arg_{Cons}(d_1)$ such that $\forall$ B' $\in Arg_{Cons}(d_2)$, A $\succeq B$.*

This means that one prefers a decision which has at least one supporting argument which is better than any supporting argument of the other decision, and also which has not a very strong argument against it.
Note that the above definition can be also refined in the same spirit as Definitions 10 and 14.

Another family of bipolar decision criteria applies the *Franklin principle* which is a natural extension to the bipolar case of the idea underlying Definition 10. This criterion consists, when comparing pros and CONS a decision, of ignoring pairs of arguments pros and CONS which have the same strength. After such a simplification, one can apply any of the above bipolar principles. In what follows, we will define formally the Franklin simplification.

**Definition 17 (Franklin simplification)** *Let $d \in \mathcal{D}$.*
*Let $P$ = ($P_1$, ..., $P_r$), ($C$ = $C_1$, ..., $C_m$) be the vectors of the arguments PROS and CONS the decision $d$. Each of these vectors is assumed to be decreasingly ordered w.r.t $\succeq$*

*(e.g.* $P_1 \succeq \ldots \succeq P_r$*).*
*The* result of the simplification *is* $P'$ = ($P_{j+1}, \ldots, P_r$), $C'$ = ($C_{j+1}, \ldots, C_m$) *such that:*

- $\forall\, 1 \leq i \leq j$, $P_i \approx C_i$ *and* ($P_{j+1} \succ C_{j+1}$ *or* $C_{j+1} \succ P_{j+1}$)
- *If* $j = r$ *(resp.* $j = m$*), then* $P' = \emptyset$ *(resp.* $C' = \emptyset$*).*

### Non-polar principles

In some applications, the arguments in favor of and against a decision are aggregated into a unique *meta-argument* having a unique strength. Thus, comparing two decisions amounts to compare the resulting meta-arguments. Such a view is well in agreement with current practice in multiple principles decision making, where each decision is evaluated according to different principles using the same scale (with a positive and a negative part), and an aggregation function is used to obtain a global evaluation of each decision.

**Definition 18 (Aggregation criterion)** *Let* $d_1, d_2 \in \mathcal{D}$*. Let* $<P_1, \ldots, P_n>$ *and* $<C_1, \ldots, C_m>$ *(resp.* $<P_1', \ldots, P_l'>$ *and* $<C_1', \ldots, C_k'>$*) the vectors of the arguments PROS and CONS the decision* $d_1$ *(resp.* $d_2$*).*
$d_1 \rhd d_2$ *iff* $h(P_1, \ldots, P_n, C_1, \ldots, C_m) \succeq h(P_1', \ldots, P_l', C_1', \ldots, C_k')$*, where* $h$ *is an* aggregation function.

A simple example of this aggregation attitude is computing the difference of the number of arguments PROS and CONS.

**Definition 19** *Let* $d_1, d_2 \in \mathcal{D}$*.*
$d_1 \rhd d_2$ *iff* $|Arg_{Pros}(d_1)| - |Arg_{Cons}(d_1)| \geq |Arg_{Pros}(d_2)| - |Arg_{Cons}(d_2)|$.

This has the advantage to be again a complete preorder, while taking into account both PROS and CONS arguments.

## Conclusion

The paper has proposed an argumentation-based framework for decision making. The framework emphasizes clearly the bipolar nature of the consequences of choices by distinguishing goals to be pursued from rejections to be avoided. This bipolar setting gives birth to two kinds of arguments for each choice: arguments in favor of that choice and arguments against it. Moreover, we have shown that there are four types of arguments PROS a choice (resp. against a choice), and some of them are stronger than others. The different types of arguments allow us to give a unique status to each choice (recommended, discommended, neutral or controversial). We have also proposed different criteria for comparing pairs of choices. The proposed approach is very general and includes as particular cases already studied argumentation-based decision principles (Fox & Das 2000; Amgoud & Prade 2004; Dubois & Fargier 2005). Besides, the richness of the different possible behaviors when arguing a decision in this framework should be compared to the actual practice of humans as studied in cognitive psychology.

## Acknowledgments

This work was supported by the Commission of the European Communities under contract IST-2004-002307, ASPIC project "Argumentation Service Platform with Integrated Components".

## References

Amgoud, L., and Prade, H. 2004. Using arguments for making decisions. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 10–17.

Amgoud, L.; Bonnefon, J.-F.; and Prade, H. 2005. An argumentation-based approach to multiple criteria decision. In *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'2005*, 269–280.

Bonet, B., and Geffner, H. 1996. Arguing for decisions: A qualitative model of decision making. In E. Horwitz, F. J. e., ed., *Proc. 12th Con. on Uncertainty in Artificial Intelligence (UAI'96)*, 98–105.

Bratman, M.; Israel, D.; and Pollack, M. 1988. *Plans and resource bounded reasoning.*, volume 4. Computational Intelligence.

Bratman, M. 1987. *Intentions, plans, and practical reason.* Harvard University Press, Massachusetts.

Brewka, G., and Gordon, T. 1994. How to buy a Porsche: An approach to defeasible decision making. In *Working Notes of the AAAI-94 Workshop on Computational Dialectics*, 28–38.

Cacioppo, J.; Gardner, W.; and Bernston, G. 1997. Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review* 1, 1:3–25.

Dubois, D., and Fargier, H. 2005. On the qualitative comparison of sets of positive and negative affects. In *Proceedings of ECSQARU'05.*

Fox, J., and Das, S. 2000. *Safe and Sound. Artificial Intelligence in Hazardous Applications*. AAAI Press, The MIT Press.

Fox, J.; Krause, P.; and Ambler, S. 1992. Arguments, contradictions and practical reasoning. In *Proceedings of the 10th European Conference on AI, ECAI'92.*

Franklin, B. 1887. *Letter to J. B. Priestley, 1772, in The Complete Works, J. Bigelow, ed.* New York: Putnam.

Raz, J. 1978. Practical reasoning. *Oxford, Oxford University Press*.

Savage, L. 1954. *The Foundations of Statistics*. New York: Dover. Reprinted by Dover, 1972.

Walton, D. 1996. *Argument schemes for presumptive reasoning*, volume 29. Lawrence Erlbaum Associates, Mahwah, NJ, USA.