Paper:

Zhong, Q., Fan, X., Luo, X. & Toni, F. (2019). An explainable multi-attribute decision model based on argumentation. *Expert Systems with Applications, 117*, 42-61.
http://dx.doi.org/10.1016/j.eswa.2018.09.038

An Explainable Multi-Attribute Decision Model based on Argumentation

Qiaoting Zhong

Center for Studies of Hong Kong, Macao and Pearl River Delta, Institute of Guangdong Hong Kong and Macao Development Studies, Sun Yat-Sen University, Guangzhou 510275, China.

Email: i.qt.zhong@gmail.com

Xiuyi Fan

Department of Computer Science, Swansea University, Swansea SA2 8PP, United Kingdom.

Email: xiuyi.fan@swansea.ac.uk

Xudong Luo

Department of of Information and Management Science, Guangxi Normal University, Guilin 541004, China.

Email: luoxd@mailbox.gxnu.edu.cn

Francesca Toni

Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom.

Email: ft@imperial.ac.uk

Abstract

We present a multi-attribute decision model and a method for explaining the decisions it recommends based on an argumentative reformulation of the model. Specifically, (i) we define a notion of best (*i.e.*, minimally redundant) decisions amounting to achieving as many goals as possible and exhibiting as few redundant attributes as possible, and (ii) we generate explanations for why a decision is best or better than or as good as another, using a mapping between the given decision model and an argumentation framework, such that best decisions correspond to admissible sets of arguments. Concretely, natural language explanations are generated automatically from dispute trees sanctioning the admissibility of arguments. Throughout, we illustrate the power of our approach within a legal reasoning setting, where best decisions amount to past cases that are most similar to a given new, open case. Finally, we conduct an empirical evaluation of our method with legal practitioners, confirming that our method is effective for the choice of most similar past cases and helpful to understand automatically generated recommendations.

Keywords: multi-attribute decision-making, explainable artificial intelligence, computational argumentation, natural language generation

1. Introduction

Numerous intelligent systems with (semi-)automated decision-making capabilities have been developed recently, in settings as diverse as loan assessments (Niu et al., 2017), financial investment (Sul et al., 2017), risk assessment (Song et al., 2017), self-driving car (Huval et al., 2015), automated negotiation (Luo et al., 2003a; Cao et al., 2015) and healthcare (Williams et al., 2015). Several of these systems lack interpretable internal components and are as a consequence black-boxes. This is widely perceived as an issue to be addressed (Dong et al., 2017a,b; Ding et al., 2017; Wachter et al., 2017), as without a clear understanding of how a recommended decision is generated, it is hard to ensure human trust, debug and improve these systems. In particular, when mission-critical tasks are carried out (*e.g.* by self-driving cars), it is important that human users understand the rationale behind decisions (Dong et al., 2017a) so that they may trust the system enough to delegate their task to it. Actually, trust can be seen as the glue that holds human users and systems together and the lubricant for tasks to be carried out smoothly, with systems' failures, such as Google autonomous car's crash (Mathur, 2015) and Tesla autopilot fatal car's crash (Banks et al., 2017), generating mistrust. Human trust can be strengthened by an understanding of why an intelligent system acts in a certain way, or comes to reach a certain conclusion, and explanations are means to facilitate this understanding. In the case of decision-making models, it is very important for humans to understand and trust the decisions recommended, especially when applications require the engagement of human users (Lacave and Díez, 2002; Teach and Shortliffe, 1984; Tintarev and Masthoff, 2012, 2015). For example, when an intelligent decision-support system in healthcare suggests to a doctor to carry out an aggressive medical treatment on a patient, in order to make sure it is proper the doctor and/or the patient needs to understand why it is necessary. As another example, when a judge needs to make her judgement according to the sentence of a previous case that is most similar to the current case, a decision-aid system could indicate one such a case for the judge, but to avoid making a wrong judgement, the judge needs the explanation of the recommendation by the system.

Argumentation has been seen as an effective means to facilitate many aspects of decision-making and decision-support systems (Fox et al., 2010; Nawwab et al., 2008), especially when decisions recommended by such systems need to be explained. However, with few exceptions (notably Amgoud and Prade (2009); Matt et al. (2009); Fan and Toni (2013); Visser et al. (2013)), it is unclear whether or not the output decisions of argumentation-based methods can be deemed to be rational in some decision-theoretic sense. Furthermore, current works in argumentation-based decision-making either fail to automatically generate explanations (*e.g.* see Atkinson et al. (2004)) or consider the outputs of argumentation engines as a form of explanation directly (*e.g.* see Kakas and Moraitis (2003)), even though these may be obscure to domain experts who are non-experts in argumentation. In this work, we present an argumentation-based decision-making model that can produce automatically generated natural language argumentative explanations obtained from an argumentation engine, for rational decisions, and evaluate it empirically with domain experts.

More specifically, we study multi-attribute decision-making where goals may not be fulfilled by some decisions, depending on whether or not these decisions exhibit attributes capable of achieving those goals. We give *rational* decisions in the sense that they achieve most goals (namely they are strongly / weakly dominant in the sense of Fan and Toni (2013), see Section 2) and also are *minimally redundant*, by having as few redundant attributes (*i.e.*, attributes that do not contribute to achieving the goals) as possible. Then, in the spirit of Fan and Toni (2013), we develop a process that maps the type of decision framework we study to a concrete instance of Assumption Based Argumentation (ABA) (Bondarenko et al., 1997; Dung et al., 2009; Toni, 2014), so that rational (strongly/weakly dominant and minimally redundant) decisions correspond to arguments belonging to admissible sets. Then, we can use the argumentation process to produce natural language explanations of the reason why some decisions are considered better (more rational) than or as good (analogously rational) as others. Our method generates explanations automatically, using template-based algorithms in the spirit of Winograd (1972); Reiter and Dale (2000), from dispute trees computed as argumentative explanations as standard in ABA. Moreover, we conduct an empirical evaluation of our method with legal practitioners, pointing to potential for impact of our method in practice.

In our model, best decisions are defined as minimally redundant (strongly/weakly dominant) decisions because, in application domains, redundant attributes could be *harmful*. Take an example of a physical education teacher who needs to choose amongst two pupils to finalise the composition of the school's water polo team. Suppose that both pupils are experienced swimmers and are keen to join the team, and one of the two, but not the other, is in the rugby team. This latter attribute is redundant (for playing water polo) and could be a distraction/possible source of injuries. As a result, the pupil not in the rugby team may be deemed to be a better choice (as minimally redundant). Even though it is possible that attributes that look redundant may turn out to be beneficial, given that people tend to avoid ambiguity (Ellsberg, 1961; Treich, 2010), decisions with a minimal number of redundant attributes make sense.

We present a detailed literature review in Section 8, discussing argumentation-based decision making, multi-attribute decision-making and case-based reasoning. Table 1 provides a summary of existing works on argumentation-based decision making that are most relevant to ours. From this table, it can be noted that many of these works are not based on well-defined decision criteria and none of them produces natural language explanations systematically for results of decision making. Thus the two main contribution of this work are:

1. Development of the minimally redundant decision criterion.

2. Introduction of natural language explanations systematically generated from argumentation-based reformulations of decision-making.

The usefulness of these two contributions is validated through an empirical evaluation (Section 7).

The remainder of the paper is organised as follows. Section 2 recaps necessary background knowledge of ABA and the decision framework that we use as our starting point, including notions of strongly/weakly dominant decisions. Section 3 introduces the concept of minimally redundant decisions. Section 4 discusses how to compare

Table 1: Summary of existing argumentation-based decision making work.

Literature	Argumentation Framework	Decision Criteria	Decision Explanation
Amgoud and Prade (2009)	Abstract Argumentation	Argument Counting	N/A
Matt et al. (2009)	ABA	Decision Dominance	N/A
Müller and Hunter (2012)	ASPIC+	Decision Dominance	N/A
Visser et al. (2012b,a, 2013)	Abstract Argumentation	Decision Preference	N/A
Fan and Toni (2013)	ABA	Decision Dominance	N/A
Fan et al. (2013)	ABA	Decision Preference	N/A
Heras et al. (2013)	Value-based Argumentation	Argumentation Semantics	Dialogue Graphs
Ferretti et al. (2017)	Abstract Argumentation	Pair-wise Comparison	N/A

decisions. Section 5 provides our argumentative counterpart of rational decisions. Section 6 gives the algorithmic machinery for generating natural language explanations for the reasons why certain decisions are better than, or as good as, others. Section 7 presents an empirical evaluation of our method with legal practitioners. Section 8 discusses related work. Finally, Section 9 summarises the paper and indicates possible directions for future work.

The paper is a significantly improved and extended version of Zhong et al. (2014). In particular, Section 6 is a heavily revised version of the same section in Zhong et al. (2014) by strengthening the algorithmic natural language explanatory part of our contribution, the empirical evaluation in Section 7 is completely new and Section 8 is a significant extension of the related work section of Zhong et al. (2014).

2. Preliminaries

Our method is based upon Assumption-Based Argumentation (Bondarenko et al., 1997; Dung et al., 2009; Toni, 2014) and the decision framework of Fan and Toni (2013). We recap them in this section.

2.1. Assumption-Based Argumentation

An *Assumption-Based Argumentation (ABA) framework* is a tuple $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$, where:

- $\langle \mathcal{L}, \mathcal{R} \rangle$ is a deductive system with *language* $\mathcal{L}$ and rule set $\mathcal{R} \subseteq \{s_0 \leftarrow s_1 \wedge \cdots \wedge s_m \mid s_0, \cdots, s_m \in \mathcal{L}, m \geq 0\}$;[1]
- $\mathcal{A} \subseteq \mathcal{L}$ is a (non-empty) set, referred to as the *assumptions*; and
- $\mathcal{C}$ is a total mapping from $\mathcal{A}$ to $2^{\mathcal{L}} \backslash \{\emptyset\}$; each element of $\mathcal{C}(\alpha)$ is referred to as a *contrary* of α.

For a *rule* $\rho = s_0 \leftarrow s_1 \wedge \ldots \wedge s_m$, s_0 is the *head* (denoted *Head*$(\rho) = s_0$) and $s_1 \wedge \ldots \wedge s_m$ constitutes the *body* (denoted *Body*$(\rho) = \{s_1, \ldots, s_m\}$). If $m = 0$, ρ is represented as $s_0 \leftarrow$ and $Body(\rho) = \emptyset$. If no assumption occurs in the head of rules in an ABA framework, then this is *flat*. All ABA frameworks in this paper are flat.

In ABA, *arguments* are deductions of claims using rules and supported by assumptions, and *attacks* are directed at assumptions. Concretely, given an ABA framework $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$,

- *an argument for claim $c \in \mathcal{L}$ supported by $S \subseteq \mathcal{A}$* (denoted $S \vdash c$) is a (finite) tree with nodes labelled by sentences in $\mathcal{L}$ or by the symbol τ,[2] such that the root is labelled by c, leaves are either τ or assumptions in S, and a non-leaf s has as many children as the elements in the body of a rule with head s, in a one-to-one correspondence with the elements of this body; and

[1] Here $\wedge$ is simply used as a separator and does not have any semantic connotation. In particular, it is different from any symbol for conjunction occurring in $\mathcal{L}$. Note that in all ABA frameworks in this paper, sentences in $\mathcal{L}$ are simply atoms.

[2] $\tau \notin \mathcal{L}$ stands for "true" and is used to represent the empty body of rules (Dung et al., 2009).

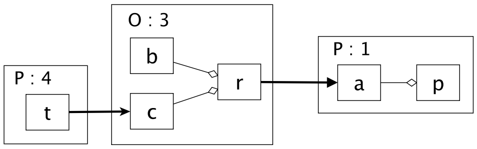

Figure 1: An admissible dispute tree for Example 1.

- an argument $S_1 \vdash c_1$ *attacks* an argument $S_2 \vdash c_2$ if and only if $c_1 \in \mathcal{C}(\alpha)$ for some $\alpha \in S_2$.

With arguments and attacks given, several semantics determining sets of arguments as "acceptable" can be defined. In this paper, we are concerned with the notion of admissibility, where a set of arguments is *admissible* if and only if it does not attack any argument it contains but attacks all arguments attacking it.[3] An admissible set of arguments can be characterised in terms of *(abstract) dispute trees* as defined by Dung et al. (2006), *i.e.*, trees with *proponent (P)* and *opponent (O)* nodes, labelled by arguments, which attack the argument in their parent node. Each P-node has all arguments attacking the argument labelling it as its children, and each O-node has one child only. If no argument in a dispute tree labels a P-node as well as an O-node, then the dispute tree is *admissible* and the set of all arguments labelling P-nodes (called *defence set*) is admissible (Dung et al., 2006).

Argumentation computation engines, including proxdd,[4] can compute dispute trees and sets of arguments that are admissible. More specifically, proxdd receives in input an ABA framework and a sentence, then determines whether or not there is an argument with that sentence as its claim and, in the set-up for the admissibility semantics, an admissible set of arguments to which that argument belongs, and finally outputs this admissible set of arguments as well as, in the case of specific choices of parameters (notably a patient selection function), a dispute tree whose defence set is that admissible set. The proxdd system is a faithful Prolog implementation of the X-dispute derivations of Toni (2013) (supporting also computation under other ABA semantics). We briefly illustrate its output (under the admissibility semantics and a patient selection function) as well as all the ABA notions with the following simple example adapted from Toni (2014).

Example 1. *Let $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$ be the ABA framework with:*[5]

- $\mathcal{R} = \{p \leftarrow q \wedge a, q \leftarrow, r \leftarrow b \wedge c, t \leftarrow\}$*;*
- $\mathcal{A} = \{a, b, c\}$*; and*

[3] An argument set As attacks an argument B if and only if some $A \in As$ attacks B, and an argument A attacks an argument set Bs if and only if A attacks some $B \in Bs$.

[4] robertcraven.org/proarg/

[5] In this and all other ABA frameworks in the paper we omit to indicate explicitly the $\mathcal{L}$ component. Implicitly, this is always the set of all the sentences occurring in the other components. Here, for example, $\mathcal{L} = \{a, b, c, p, q, r, s, t\}$.

- $\mathcal{C} : \mathcal{A} \rightarrow 2^{\mathcal{L}} \setminus \{\emptyset\}$ *is given by:* $\mathcal{C}(a) = \{r\}$, $\mathcal{C}(b) = \{s\}$, *and* $\mathcal{C}(c) = \{t\}$.

Some examples of arguments for this ABA framework are

$$\{a\} \vdash p, \quad \{b, c\} \vdash r, \quad \emptyset \vdash t.$$

The set of arguments $A = \{\emptyset \vdash t, \{a\} \vdash p\}\}$ *is admissible (there are several other admissible sets,* e.g. $\emptyset$*). This admissible set* A *is the defence set of the dispute tree shown in Figure 1, presented in a format similar to the one obtained as output from* `proxdd`*. In the figure, each argument is represented as a rectangle, the argument claim is denoted by the right inner rectangle and assumptions are denoted by the left inner rectangle, connected by a diamond-pointed arrow. Arguments supported by the empty set* $\emptyset$ *of assumptions are denoted by rectangles with a single inner rectangle. Attacks are denoted by arrows between (claims and assumptions of) outer rectangles (namely the arguments). Arguments are* P *or* O *and are numbered for ease of reference.*

2.2. Decision Frameworks

Decision frameworks (Fan and Toni, 2013) capture information used in decision-making. Specifically, a decision framework is composed of goals, attributes and (alternative) decisions such that decisions *have* attributes and goals are *achieved by* attributes. These *decisions-have-attributes* and *goals-are-achieved-by-attributes* relations are specified by two tables. Formally,

- a *decision framework* is a tuple $\langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$, with a set of decisions $\mathtt{D} = \{d_1, \cdots, d_n\}$ $(n > 0)$, a set of attributes $\mathtt{A} = \{a_1, \cdots, a_m\}$ $(m > 0)$, a set of goals $\mathtt{G} = \{g_1, \cdots, g_l\}$ $(l > 0)$, and two tables $\mathtt{DA}$ (of size $n \times m$) and $\mathtt{GA}$ (of size $l \times m$) (we use $\mathtt{X}_{i,j}$ to represent the cell in row i and column j in $\mathtt{X} \in \{\mathtt{DA}, \mathtt{GA}\}$) such that: (i) every $\mathtt{DA}_{i,j}$ is either 1, representing that alternative decision d_i *has* attributes a_j, or 0, otherwise; and (ii) every $\mathtt{GA}_{i,j}$ is either 1, representing that goal g_i is *achieved* by attribute a_j, or 0, otherwise.

$\mathtt{DA}$ and $\mathtt{GA}$ share the same column orders. The row numbers of decisions and goals in $\mathtt{DA}$ and $\mathtt{GA}$ are the *indices* of decisions and goals, respectively, whereas the indices of attributes are defined as the column numbers in both $\mathtt{DA}$ and $\mathtt{GA}$. We use $\mathcal{DEC}$ and $\mathcal{DF}$ to represent the set of all possible decisions and the set of all possible decision frameworks, respectively. Note that, in decision frameworks, goals are achieved by single attributes, but the relationship between goals and attributes is not one-to-one, since different attributes may achieve the same goal and different goals may be achieved by the same attribute.

Given a decision framework, the relation between decisions and goals can be obtained from $\mathtt{DA}$ and $\mathtt{GA}$ as follows:

- a decision $d \in \mathtt{D}$ with row index i in $\mathtt{DA}$ *achieves* a goal $g \in \mathtt{G}$ with row index j in $\mathtt{GA}$ if and only if there is an attribute $a \in \mathtt{A}$ with column index k in both $\mathtt{DA}$ and $\mathtt{GA}$, such that $\mathtt{DA}_{i,k} = 1$ and $\mathtt{GA}_{j,k} = 1$.

Table 2: DA (left) and GA (right) for Example 2, with $V_1, V_2 \in \{0, 1\}$

	a_1	a_2
d_1	1	0
d_2	V_1	1

	a_1	a_2
g_1	1	0
g_2	V_2	1

In the remainder of the paper, we will use $\gamma(d)$ to denote the set of goals achieved by d (obviously $\gamma(d) \subseteq \mathtt{G}$). Moreover, unless otherwise specified, $df = \langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle \in \mathcal{DF}$ is a generic decision framework.

Decision criteria are used to select the *best* decisions in a decision framework according to some criteria. Our work is based upon two decision criteria: *strong dominance* and *weak dominance*. Strongly dominant decisions achieve all goals, and the set of goals achieved by a weakly dominant decision is not a proper subset of the set of goals achieved by any other decision. Formally,

- $d \in \mathtt{D}$ is *strongly dominant* if and only if $\gamma(d) = \mathtt{G}$; and
- $d \in \mathtt{D}$ is *weakly dominant* if and only if $\nexists d' \in \mathtt{D} \backslash \{d\}$ such that $\gamma(d) \subset \gamma(d')$.

Note that as jointly shown by Proposition 2 and 3 in Fan and Toni (2013), a strongly dominant decision is also weakly dominant. This is easy to see as for any strongly dominant decision d, since d achieves all goals, there is no d' such that d' achieves more goals than d. Therefore, d is also weakly dominant.

Now we illustrates decision frameworks and the two notions of dominant decisions as follows:

Example 2. *Consider the decision framework* $\langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$ *with decision set* $\mathtt{D} = \{d_1, d_2\}$*, attribute set* $\mathtt{A} = \{a_1, a_2\}$*, goal set* $\mathtt{G} = \{g_1, g_2\}$*, and tables* DA *and* GA *as given in Table 2. Trivially, if* $V_1 = V_2 = 1$ *then both* d_1 *and* d_2 *are strongly as well as weakly dominant (they are strongly dominant as they meet both goals; they are weakly dominant as there is no decision meeting more goals than either of the two). If instead* $V_1 = V_2 = 0$ *then neither decision is strongly dominant but both are weakly dominant.*

3. Minimally Redundant Decisions

In this section, we will take one more factor into consideration when choosing best decisions: the presence/absence of *redundant* attributes, and define *minimally redundant* decisions as (strongly or weakly) dominant decisions with as few redundant attributes as possible.

3.1. Motivation

Achieving goals is central to rational decision-making, however, goal achievement alone does not always allow to discriminate amongst decision alternatives, and, futhermore, it ignores the presence of *redundant* attributes, which is an important factor in some decision-making, as illustrated below.

Table 3: $\mathtt{DA}$ (left) and $\mathtt{GA}$ (right) for Example 3

	a_1	a_2	a_3
d_1	1	1	0
d_2	1	1	1

	a_1	a_2	a_3
g_1	1	0	0
g_2	0	1	0

Example 3. *Consider the decision framework* $\langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$ *with* $\mathtt{D} = \{d_1, d_2\}$*,* $\mathtt{A} = \{a_1, a_2, a_3\}$*,* $\mathtt{G} = \{g_1, g_2\}$*, and* $\mathtt{DA}$ *and* $\mathtt{GA}$ *as given in Table 3. Let* d_1 *and* d_2 *be two player candidates for a water polo team, with* a_1 *and* a_2 *representing, respectively, that a candidate belongs to a swimming club and has asked to join the team, and* a_3 *representing that a candidate also actively plays rugby. Let* g_1 *and* g_2 *represent, respectively, strong swimming skills and willingness to participate in water polo. Then,* d_1 *and* d_2 *are (strongly and weakly) dominant, but* d_1 *may be deemed a better decision, since* a_3 *is* redundant *(and potentially harmful) towards the achievement of the goals.*

It is worth noting that the concept of redundant attributes is specific to pre-defined goals in the sense that the existence of some attributes can be harmful when considering decisions achieving those goals. For instance, with respect to a set of goals G_1, some attributes A are redundant; yet there may exist a different set of goals G_2 such that with respect to G_2, the same attributes A are not redundant. Moreover, a decision framework can be viewed as an information store, such that it is constructed to include as much relevant information as possible. Thus, it is the task of a decision criterion to select the "best" decisions from a given decision framework, and best decisions may change if new goals are considered and re-writing the information store is needed.

The concept of redundant attributes is also useful in case-based reasoning, because the most relevant cases to a new case can be viewed as the ones with "fewest" redundant attributes (*i.e.*, as accurate as possible) as illustrated by the following medical literature search example, adapted from Fan et al. (2013).

Example 4. *In medical research, one sometimes faces the problem of choosing which medical studies to base a diagnosis or a treatment on, for a given patient. This can be viewed as a decision making problem and our techniques can be used to solve it. Assume that we have identified 4 randomised clinical trials on the treatment of brain metastases.*[6] *The decisions (*$\mathtt{D}$*) of our model are choices to use a given paper in a diagnosis or treatment. Each literature paper describes a two-arm trial. We extract a list of representative trial design criteria and patient characteristics from these papers. These criteria and characteristics can be considered as* attributes *(*$\mathtt{A}$*) of decisions.*

Assume that the relations between papers and trial criteria / characteristics are as given in Table 4 ($\mathtt{DA}$*). Here, a "1" in row* k*, column* i *should be interpreted as "the trial reported in paper* p_k *has criterion / characteristics* i*". A blank means that the corresponding criterion / characteristics is either not reported or not met by the particular paper. The list of characteristics are: over 18 years old, with 1 metastases, with 2 metastases, with more than 2 metastases, having endocardial cushion defect (ECD), with performance status 0 or 1 (PS1), and with performance status 2 (PS2).*

[6] These are real published medical literature. The PMID of each paper can be found in Fan et al. (2013).

For instance, the first row should be read as: the trial reported in paper p_1 includes patients over 18 years old, with 1, 2 or more brain metastases, with performance status either 0 or 1 (PS1).

Table 4: Paper / Trial Characteristics (DA)

	> 18	$1m$	$2m$	$> 2m$	ECD	$PS\,0,1$	$PS\,2$
p_1	1	1	1	1		1	
p_2					1	1	1
p_3	1	1	1	1		1	1
p_4	1	1	1	1			
Goal	1			1		1	

Since the aim is to find medical papers for a particular patient, we view properties of the given patient as goals (G). In this setting, "good" decisions are medical papers that better match with the particular patient's properties. We present relations between patient's properties and trial characteristics in Table 5 (GA). The occurrence of a "1" in the table represents trial characteristics meeting patient properties. For instance, the sample patient shown in Table 5 has four properties: he is 64 years old, has three metastases, has no ECD, and has a performance status of 1.

Amongst the four papers $p_1, \ldots, p_4$, we observe that:

- *Paper p_2 does not match with the patient's properties as it does not contains studies for adult patients, it does not report the number of metastases its patients' have, and its patients have ECD.*

- *Paper p_4 does not match with the patient's properties either, as its patients have performance statuses other than 0 to 2.*

- *Thus, we need to choose between p_1 and p_3 for the best match. We argue that p_1 is a better choice for the specific patient as its study is solely focused on patients with performance status 0 and 1 and does not include patients with more severe conditions such as performance status 2. The "tighter" fit of p_1 makes it more suitable to our target patient than p_3, i.e., p_3 having the attribute $PS2$ makes it less desirable, as $PS2$ can be deemed redundant (with respect to this specific goal patient).*

The concept of redundant attributes is generic, as further shown in the following legal example, which we will use throughout the paper to illustrate our work.

Table 5: Patient Properties / Trial Characteristics (GA)

	> 18	$1m$	$2m$	$> 2m$	ECD	$PS\,0,1$	$PS\,2$
$Age64$	1						
$3m$				1			
$PS\,1$						1	
$Lung$							

Table 6: A fragment of the Past Cases Characteristics

Index	No.	Attr. a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	Sentence
d_1	245	0	1	0	1	0	1	0	0	0	1y+¥1k
d_2	97	1	0	0	0	0	0	1	0	0	10y+3y+¥10k
d_3	420	1	0	0	1	1	0	0	1	0	6m+¥1k
d_4	96	1	0	0	0	0	1	0	1	0	3y+¥3k
d_5	48	1	0	0	0	1	0	0	1	0	6m+¥1k
d_6	751	1	0	0	0	0	1	0	1	0	5y+¥5k
d_7	801	1	0	0	0	1	0	0	1	0	6m+¥1k
d_8	1962	1	0	0	0	0	0	1	0	1	7y+¥7k
d_9	389	1	0	0	1	1	0	0	0	0	4m+¥1k
d_{10}	686	1	0	1	0	1	0	0	1	0	6m+¥1k
Goal case	355	1	0	0	0	0	0	1	1	0	10y+3y+¥10k

Example 5. *In the legal domain, when jury members, judges, lawyers or other involved parties review a new law case, it is common practice that they search for and examine similar past cases so that they can use the (court) sentence of the past cases as a basis for estimating the outcome of the new case (Kronman, 1990). Identifying most similar past cases can be formulated in decision-making terms, where past cases are seen as alternative decisions. The suitability of each past case can be computed based on the sentence of the new case: the closer the sentence to the ones for the past cases, the more rational the choice of these past cases as similar. Table 6 summarises the eleven real cases, all concerning theft, from the Nanhai District People's Court in the city of Foshan, Guangdong Province, China. Each case has:*

- *a number of attributes,* i.e.*, $a_1, \cdots, a_9$, standing for "older than 18", "age between 16 and 18", "burglary", "repeatedly", (value of goods) "large amount", "huge amount", "extremely huge amount", "goods found" and "accessory", respectively; and*
- *a sentence,* e.g. *"1 year of imprisonment with ¥1,000 fine" (represented in Table 6 as "1y+¥1,000") or "10 years of imprisonment, 3 years deprivation of political right with ¥10,000 fine" (represented in Table 6 as "10y+3y+¥10,000").*

For each case, the value of an attribute can either be 1 or 0, representing the case having the attribute or not, respectively. For example, the row indexed d_1 in Table 6 represents case No. 245, where:

> the defendant, aged between 16 and 18 (a_2), stole repeatedly (a_4), and the value of the stolen goods was huge (a_6); and the resulting sentence was 1 year imprisonment with ¥1,000 fine.

From this point forward, we view case No. 355 (indexed **Goal case** *in Table 6) as new (thus ignoring its sentence). In this case,*

> the defendant was older than 18 (a_1), the value of the stolen goods was extremely huge (a_7), and the stolen goods were found (a_8).

Table 7: GA table for Example 6.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
g_1	1	0	0	0	0	0	0	0	0
g_2	0	0	0	0	0	0	1	0	0
g_3	0	0	0	0	0	0	0	1	0

Case No. 97, where

> the defendant was older than 18 (a_1), the value of the stolen goods was extremely huge (a_7), and the resulting sentence was 10 years of imprisonment, 3 years deprivation of political right plus ¥10,000 fine,

can be deemed to be the past case most similar to case No. 355 as case No. 97 is the case matching most attributes of case No. 355 while also exhibiting the fewest redundant attributes (as case No. 97 only deviates from case No. 355 on a_8). Such conclusion is validated by the fact that both case No. 355 and case No. 97 have the same sentence.

We can formalise the above example as a decision framework as follows:

Example 6 (Example 5 as a decision framework)**.** *The decision problem of determining which one, amongst cases labelled $d_1, \cdots, d_{10}$ in Table 6, are most similar to the goal case (*i.e.*, the new case No. 355) can be formalised as a decision framework $df = \langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$ in which*

- $\mathtt{D} = \{d_1, \cdots, d_{10}\}$*;*
- $\mathtt{A} = \{a_1, \cdots, a_9\}$*;*
- $\mathtt{G} = \{g_1, g_2, g_3\}$*, where g_1, g_2 and g_3 stand for "older than 18", "extremely huge amount" and "goods found", respectively, and amount to attributes a_1, a_7 and a_8, respectively, that the goal case has;*
- DA *is Table 6 without case No. 355 and dropping the column with the sentences; and*
- GA *is Table 7.*

*No strongly dominant decisions exist here as no decisions can achieve all three goals g_1, g_2, and g_3. Thus, we need to rely on weak dominance to make a decision. Several weakly dominant decisions exist, as indeed decisions d_2 (No. 97), d_3 (No. 420), d_4 (No. 96), d_5 (No. 48), d_6 (No. 751), d_7 (No. 801), d_8 (No. 1962), and d_{10} (No. 686) are all weakly dominant. However, notice that the goal case (*i.e.*, case No. 355) is achieved by having three attributes a_1, a_7, and a_8. Thus, to achieve the goals, decisions with these three attributes are sufficient. The remaining attributes are actually* redundant *in the sense that they do not contribute to achieving any goal. For instance, both d_2 (case No. 97) and d_8 (case No. 1962) achieve two goals g_1 and g_2. Yet, d_2 has no* redundant *attributes, whereas d_8 has attribute a_9 satisfying no goals. Therefore, we conclude that case No. 97 is more similar to the goal case.*

3.2. Redundant attributes and minimal redundancy

In settings such as the ones outlined in Examples 3–6, *having the fewest redundant attributes* is an important property for best (most rational) decisions. Indeed, we can use the presence/absence of redundant attributes as a *filter* to further distinguish amongst (strongly or weakly) dominant decisions. Let us first formalise the notion of redundant attribute as follows:

Definition 1. *Let $\alpha \in \mathtt{A}$ and i be the column index of α in $\mathtt{DA}$ and $\mathtt{GA}$. Then α is a* redundant attribute *if and only if $\forall g \in \mathtt{G}$, if g has row index j in $\mathtt{GA}$, then $\mathtt{GA}_{j,i} \neq 1$. The set of* redundant attributes decision d has *is denoted as $\lambda(d)$.*

Here, $\alpha \in \lambda(d)$ indicates that d has attribute α but α fulfils no goals in $\mathtt{G}$. For our water polo example (*i.e.*, Example 3), since $\mathtt{GA}_{1,3} = 0$ and $\mathtt{GA}_{2,3} = 0$, a_3 is redundant and $\lambda(d_1) = \emptyset$, $\lambda(d_2) = \{a_3\}$. For our legal example (*i.e.*, Example 6), since $\mathtt{GA}_{1,9} = 0$, $\mathtt{GA}_{2,9} = 0$ and $\mathtt{GA}_{3,9} = 0$, a_9 is a redundant attribute. Similarly, $a_2, \cdots, a_6$ are redundant. Then, $\lambda(d_1) = \{a_2, a_4, a_6\}$, $\lambda(d_2) = \emptyset$ and so on.

Minimally redundant decisions have fewest redundant attributes with respect to set inclusion. Formally:

Definition 2. *A decision $d \in \mathtt{D}$ is* minimally redundant *in $\langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle \in \mathcal{DF}$ if and only if $\nexists d' \in \mathtt{D}$ such that $\lambda(d') \subset \lambda(d)$.*

In other words, a decision d is minimally redundant if and only if there is no other decision d' such that the set of redundant attributes that d' has is a proper subset of the set of redundant attributes that d has. In Example 3, it is easy to see that d_1 (the candidate not playing rugby) is minimally redundant. In Example 6, it is easy to see that d_2 (case No. 97) is minimally redundant.

Minimal redundancy only concerns the relation between decisions and attributes, hence it is orthogonal to *dominance* (see Section 2), concerning the relation between decisions and goals. Therefore, we will identify decisions with a two-step process: we first identify all dominant decisions, and then, amongst all dominant decisions, we select the minimally redundant ones. To refine decisions on grounds of redundancy (and thus support the second step of the aforementioned process), we introduce *sub-frameworks*, as given below:

Definition 3. *Given $\mathtt{D}' \subseteq \mathtt{D}$, the* sub-framework *of df* with respect to $\mathtt{D}'$ *is a decision framework $\langle \mathtt{D}', \mathtt{A}, \mathtt{G}, \mathtt{DA}', \mathtt{GA} \rangle$ such that $\mathtt{DA}'$ is the restriction of $\mathtt{DA}$ that contains only rows for d_i such that $d_i \in \mathtt{D}'$.*

We subsequently combine the concepts of strong/weak dominance and minimal redundancy below:

Definition 4. *Given df $= \langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle \in \mathcal{DF}$, $d \in \mathtt{D}$ is*

- Minimally Redundant Strongly Dominant (MRSD) *if and only if d is minimally redundant in df_s, the sub-framework of df with respect to the set of all strongly dominant decisions;*

- Minimally Redundant Weakly Dominant (MRWD) *if and only if d is minimally redundant in df_w, the sub-framework of df with respect to the set of all weakly dominant decisions.*

Thus, we adopt a lexicographic approach, selecting first the (weakly or strongly) dominant decisions, and then, amongst those, the minimally redundant decisions.

As an illustration, in Example 3, both decisions are strongly dominant and d_1 (the candidate not playing rugby) is MRSD. In Example 6, there are no MRSD decisions, as there are no strongly dominant decisions, and d_2 (case No. 97) is MRWD as this case is weakly dominant and has no redundant attributes. There are no other MRWD decisions in Example 6.

It is easy to see that, if there exists any strongly dominant decision, then the set of MRSD decisions and the set of MRWD decisions coincide. Formally, we have:

Lemma 1. *Let D_s, D_{ms}, and D_{mw} be the sets of all strongly dominant, MRSD, and MRWD decisions, respectively. If $D_s \neq \emptyset$ then $D_{ms} = D_{mw}$.*

Proof. Let df_s be the sub-framework of df with respect to D_s, and df_w be the sub-framework of df with respect to D_w, the set of all weakly dominant decisions. Since $D_s \neq \emptyset$, we have $D_s = D_w$ (by Proposition 4 in Fan and Toni (2013)). Thus, we have $df_w = df_s$. So the lemma holds. □

4. Comparing Decisions

So far, we have introduced two kinds of minimally redundant decisions (*i.e.*, MRSD and MRWD decisions), adding to two existing kinds of rational decisions (*i.e.*, the strongly dominant and weakly dominant decisions of Fan and Toni (2013)). Thus, in this section, we can formally discuss how to compare decisions in different categories (*i.e.*, we will present a mechanism to rank them).

We first define the *better-than* relation between decisions, given as follows:

Definition 5. *For any $d, d' \in \mathtt{D}$, d is* better than *d', denoted $d \succ d'$, if and only if:*

(i) d is strongly dominant and d' is not strongly dominant, or

(ii) d is weakly dominant and d' is not weakly dominant, or

(iii) d is MRSD and d' is strongly dominant and d' is not MRSD, or

(iv) d is MRWD and d' is weakly dominant and d' is not MRWD.

We also say that d is as good as *d', denoted $d \sim d'$, if and only if neither $d \succ d'$ nor $d' \succ d$.*

The above definition is given with the following intuition: non-strongly/non-weakly dominant decisions are the least favourable decisions; minimally redundant strongly/weakly dominant decisions are the most favourable decisions; and two decisions are equally good unless one is more favourable than the other.

Note that if a non-empty set of strongly dominant decisions D exists, then D is the set of weakly dominant decisions as well (by Proposition 4 in (Fan and Toni, 2013)). In other words, in this case, there does not exist a decision that is weakly dominant but not strongly dominant. Therefore, if there exists an non-empty set of strongly dominant decisions, then by Definition 5 any strongly dominant decision is as good as any weakly dominant decision (as well as any other strongly dominant decision).

As an illustration, for Example 6, since d_2 (case No. 97) is the only MRWD and d_8 (case No. 1962) is weakly dominant, we have $d_2 \succ d_8$ by Definition 5(iv).

The following three propositions sanction an important result: for any pair of decisions in our decision framework, we can always compare them using the notions of *better than* and *as good as* we defined. We start with $\succ$ being transitive and $\sim$ being an equivalence relation on $\mathtt{D}$, as shown below:

Proposition 1. *For any d, d', $d'' \in \mathtt{D}$, if $d'' \succ d'$ and $d' \succ d$, then $d'' \succ d$.*

Proof. By Definition 5, there are only five possibilities satisfying $d'' \succ d'$ and $d' \succ d$, so we check them one by one. (i) In the case that d'' is strongly dominant, d' is weakly dominant but not strongly dominant, and d is not weakly dominant, we have d'' weakly dominant by Proposition 4 in Fan and Toni (2013). Hence, we have $d'' \succ d$ by Definition 5(ii). (ii) In the case that d'' is MRSD, d' is strongly dominant but not MRSD, and d is not strongly dominant, we have d'' strongly dominant by Definition 4. Hence, we have $d'' \succ d$ by Definition 5(i). (iii) In the case that d'' is MRSD, d' is strongly dominant but not MRSD, and d is not weakly dominant, we have d'' weakly dominant by Definition 4 and Proposition 4 in Fan and Toni (2013). Since d is not weakly dominant, we have $d'' \succ d$ by Definition 5(ii). (iv) In the case that d'' is MRWD, d' is strongly dominant but not MRWD, and d is not strongly dominant, we have d'' MRSD by Lemma 1. Hence, we have $d'' \succ d$ by Definitions 4 and 5(i). (v) In the case that d'' is MRWD, d' is weakly dominant but not MRWD, and d is not weakly dominant, we have d'' weakly dominant by Definition 4. Hence, we have $d'' \succ d$ by Definition 5(ii). □

Proposition 2. *$\sim$ is an equivalence relation on $\mathtt{D}$.*

Proof. It is trivial to prove that $\sim$ is reflexive and symmetric. For transitivity, we have $\forall d_1, d_2, d_3 \in \mathtt{D}$, if $d_1 \sim d_2$ and $d_2 \sim d_3$, by Definition 5, we have $d_1 \not\succ d_2$, $d_2 \not\succ d_1$, $d_2 \not\succ d_3$, and $d_3 \not\succ d_2$. Suppose $d_1 \not\sim d_3$, we can have either $d_1 \succ d_3$ or $d_3 \succ d_1$. (i) Since $d_1 \succ d_3$, by Definition 5, there are four sub-cases: (a) d_1 is strongly dominant but d_3 is not, (b) d_1 is weakly dominant but d_3 is not, (c) d_1 is MRSD and d_3 is strongly dominant but not MRSD, and (d) d_1 is MRWD and d_3 is weakly dominant but not MRWD. Here we just consider the first sub-case, the others can be proven similarly. Since $d_1 \not\succ d_2$, we have that d_2 is strongly dominant, which leads to $d_2 \succ d_3$ by Definition 5. Contradiction. (ii) Similarly, the case of $d_3 \succ d_1$ can be proven. □

Given the quotient set $\mathtt{D}/\!\sim$ of all equivalence classes, we can define a binary relation $\geq$ on $\mathtt{D}/\!\sim$ as follows:

Definition 6. *Given $d \in \mathtt{D}$, let $[d]$ denote the equivalence class to which d belongs. For any $[d'], [d''] \in \mathtt{D}/\!\sim$, $[d'] \geq [d'']$ if and only if $d' \succ d''$ or $d' \sim d''$.*

Note that $[d'] = [d'']$ if and only if $d' \sim d''$.

Proposition 3. *$\geq$ is a total order on* $\mathtt{D}/\sim$.

Proof. To establish this result, it must be shown that $\geq$ is anti-symmetric, transitive and total. (i) Anti-symmetry. For any $d_1, d_2 \in \mathtt{D}$, assume $[d_1] \geq [d_2]$ and $[d_2] \geq [d_1]$. We then need to prove $[d_1] = [d_2]$. Since $[d_1] \geq [d_2]$, then $d_1 \succ d_2$ or $d_1 \sim d_2$. For the former, we have $d_2 \not\succ d_1$. Since $[d_2] \geq [d_1]$, we have $d_2 \sim d_1$, which leads to $[d_1] = [d_2]$. For the latter, $[d_1] = [d_2]$ certainly. (ii) Transitivity follows from Propositions 1 and 2. (iii) Totality can be proven by contradiction. Suppose there exist $[d_1]$ and $[d_2]$ (where $d_1, d_2 \in \mathtt{D}$) such that $[d_1] \ngeq [d_2]$ and $[d_2] \ngeq [d_1]$. By Definition 6, we have $d_1 \not\succ d_2$, $d_1 \not\sim d_2$, $d_2 \not\succ d_1$, and $d_2 \not\sim d_1$. Since $d_1 \not\succ d_2$ and $d_2 \not\succ d_1$, we have $d_1 \sim d_2$ and $d_2 \sim d_1$ by Definition 5. Contradiction. □

The above proposition is the main result in this section. It sanctions that *better than* or *as good as* given in Definition 5 can be used to compare any two decisions in a decision framework. The comparison is core in our method for explaining decisions in Section 6. This method relies upon the mapping from a decision framework and a decision criterion to an ABA framework, given next.

5. Argumentative Counterpart of MRSD and MRWD Decisions

In this section, we present a mapping that takes decision frameworks and gives ABA frameworks so that the two kinds of minimally redundant decisions presented in Section 3.2 can be computed by finding admissible sets of arguments. This mapping also gives a mechanism for producing dispute trees that *explain* best decisions argumentatively and that will be used in Section 6 to *explain*in natural language why one decision is better than or as good as another, in the sense of Definition 5.

Concretely, given a decision criterion (amongst MRSD and MRWD) and a decision framework, an *equivalent* ABA framework is constructed.

Now we start with defining the ABA mapping for MRSD decisions as follows:[7]

Definition 7. *The* MRSD ABA framework *for* $\langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$ *is* $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$*, where:*

[7] Note that we could have given a single ABA framework for both MRSD and MRSW decisions, amounting to the union of the ABA framework for MRSD decisions (Definition 7) and the ABA framework for MRSW decisions (Definition 8). However, for clarity of presentation and readability, we have opted for two separate definitions instead.

- *$\mathcal{R}$ consists of the following rules (and nothing else):*[8]

$$\begin{aligned}
&\textit{for each } d \in \mathtt{D}, a \in \mathtt{A} \textit{ such that } d \textit{ has } a\textit{:} \quad \textit{hasAttr}(d,a) \leftarrow && (1)\\
&\textit{for each } g \in \mathtt{G}, a \in \mathtt{A} \textit{ such that } g \textit{ is satisfied by } a\textit{:} \quad \textit{satBy}(g,a) \leftarrow && (2)\\
&\textit{de}(D,A) \leftarrow \textit{hasAttr}(D,A) \wedge \textit{redundant}(A) && (3)\\
&\textit{notRedundant}(A) \leftarrow \textit{satBy}(G,A) && (4)\\
&\textit{notMSD}(D) \leftarrow \textit{sd}(D') \wedge \textit{worse}(D,D') \wedge \textit{nWorse}(D',D) && (5)\\
&\textit{nSD}(D) \leftarrow \textit{nMet}(D,G) && (6)\\
&\textit{met}(D,G) \leftarrow \textit{hasAttr}(D,A) \wedge \textit{satBy}(G,A) && (7)\\
&\textit{worse}(D',D) \leftarrow \textit{de}(D',A) \wedge \textit{nAttr}(D,A) && (8)
\end{aligned}$$

- *the set of assumptions $\mathcal{A}$ is given by*

$$\begin{aligned}
\mathcal{A} =& \{\textit{ms}(d), \textit{sd}(d), \textit{redundant}(a), \textit{nAttr}(d,a) \mid d \in \mathtt{D}, a \in \mathtt{A}\} \cup \\
& \{\textit{nWorse}(d,d') | d, d' \in \mathtt{D}, d \neq d'\} \cup \{\textit{nMet}(d,g) \mid d \in \mathtt{D}, g \in \mathtt{G}\}. \qquad (9)
\end{aligned}$$

- *$\mathcal{C} : \mathcal{A} \to 2^{\mathcal{L}} \setminus \{\emptyset\}$ is given by: for each $d, d' \in \mathtt{D}$ (such that $d \neq d'$), $a \in \mathtt{A}$, $g \in \mathtt{G}$:*

$$\begin{aligned}
&\mathcal{C}(\textit{ms}(d)) = \{\textit{notMSD}(d), \textit{nSD}(d)\}, \quad \mathcal{C}(\textit{sd}(d)) = \{\textit{nSD}(d)\}, && (10)\\
&\mathcal{C}(\textit{redundant}(a)) = \{\textit{notRedundant}(a)\}, \quad \mathcal{C}(\textit{nAttr}(d,a)) = \{\textit{hasAttr}(d,a)\}, && (11)\\
&\mathcal{C}(\textit{nWorse}(d,d')) = \{\textit{worse}(d,d')\}, \quad \mathcal{C}(\textit{nMet}(d,g)) = \{\textit{met}(d,g)\}. && (12)
\end{aligned}$$

The intuition behind the above definition is as follows. A decision d can always be assumed to be MRSD, since ms(d) is an assumption (see (9)). This assumption can then be debated (argumentatively) by assessing arguments for the contraries of this assumption, nSD(d) and notMSD(d) (see (10)). Indeed, d is not MRSD under either of two conditions: (i) d is not strongly dominant (nSD(d), see (10)), or (ii) d is not minimally redundant (notMSD(d), see (10)). According to (5), a decision d is not minimally redundant (notMSD(d)) if there exists some other decision d' such that d' is strongly dominant (sd(d')) and d contains some redundant attribute which d' does not (worse(d, d')) and d' does not contain any more redundant attributes than d (nWorse(d', d)). According to (6), any decision d is not strongly dominant (nSD(d)) if there exists a goal g that d does not fulfil (nMet(d, g)). Since sd(d) is an assumption (see (9)), d can always be assumed to be strongly dominant, subject to debate, amounting to assessing arguments for its contrary nSD(d) (see (10)), that d is actually not strongly dominant, namely (by (6)) that d does not achieve some goal g (nMet(d, g)). Again, given that nMet(d, g) is an assumption (see (9)) with contrary met(d, g) (see (12)), the

[8] We sometimes use *schemata* with variables (D, D', A, G) to represent compactly all rules that can be obtained by instantiating the variables as follows: D, D' are instantiated to decisions, A to attributes, G to goals.

Table 8: DA (left) and GA (right) for Example 7

	a
d_1	1
d_2	0

	a
g	1

acceptance of nMet(d, g) is subject to checking whether or not d has some attribute satisfying g (see (7)). The other assumptions, contraries and rules in the given ABA framework can be understood similarly. Overall, the MRSD ABA framework can be used to determine whether or not a decision d is MRSD by checking whether or not the assumption ms(d) can be accepted.

The following example illustrates the notion of MRSD ABA framework for a very simple decision framework.

Example 7. *Consider the decision framework $df = \langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$ with $\mathtt{D} = \{d_1, d_2\}$, $\mathtt{A} = \{a\}$, $\mathtt{G} = \{g\}$, and tables DA and GA given in Table 8. The MRSD ABA framework for df is $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$, where:*

- *$\mathcal{R}$ consists of*

$$\begin{aligned}
&\mathit{hasAttr}(d_1, a) \leftarrow,\\
&\mathit{satBy}(g, a) \leftarrow,\\
&\mathit{de}(d_1, a) \leftarrow \mathit{hasAttr}(d_1, a) \wedge \mathit{Credundant}(a),\\
&\mathit{de}(d_2, a) \leftarrow \mathit{hasAttr}(d_2, a) \wedge \mathit{Credundant}(a),\\
&\mathit{notRedundant}(a) \leftarrow \mathit{satBy}(g, a),\\
&\mathit{notMSD}(d_1) \leftarrow \mathit{sd}(d_2) \wedge \mathit{worse}(d_1, d_2) \wedge \mathit{nWorse}(d_2, d_1),\\
&\mathit{notMSD}(d_2) \leftarrow \mathit{sd}(d_1) \wedge \mathit{worse}(d_2, d_1) \wedge \mathit{nWorse}(d_1, d_2),\\
&\mathit{nSD}(d_1) \leftarrow \mathit{nMet}(d_1, g),\\
&\mathit{nSD}(d_2) \leftarrow \mathit{nMet}(d_2, g),\\
&\mathit{met}(d_1, g) \leftarrow \mathit{hasAttr}(d_1, a) \wedge \mathit{satBy}(g, a),\\
&\mathit{met}(d_2, g) \leftarrow \mathit{hasAttr}(d_2, a) \wedge \mathit{satBy}(g, a),\\
&\mathit{worse}(d_1, d_2) \leftarrow \mathit{de}(d_1, a) \wedge \mathit{nAttr}(d_2, a),\\
&\mathit{worse}(d_2, d_1) \leftarrow \mathit{de}(d_2, a) \wedge \mathit{nAttr}(d_1, a).
\end{aligned}$$

- $\mathcal{A} = \{\mathit{ms}(d_1), \mathit{ms}(d_2), \mathit{sd}(d_1), \mathit{sd}(d_2), \mathit{redundant}(a), \mathit{nAttr}(d_1, a), \mathit{nAttr}(d_2, a),$ $\mathit{nWorse}(d_1, d_2), \mathit{nWorse}(d_2, d_1), \mathit{nMet}(d_1, g), \mathit{nMet}(d_2, g)\}$.

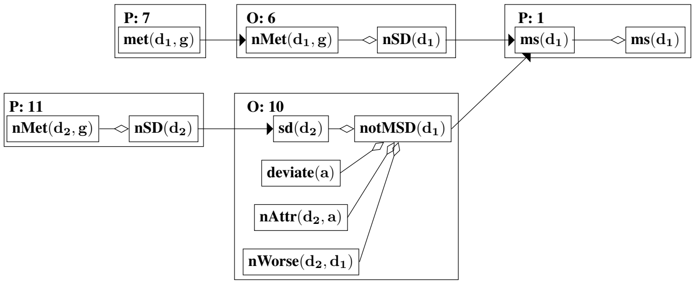

Figure 2: Admissible dispute tree for Example 7.

- *$\mathcal{C} : \mathcal{A} \rightarrow 2^{\mathcal{L}} \setminus \{\emptyset\}$ is given by:*

  $\mathcal{C}(ms(d_1)) = \{notMSD(d_1), nSD(d_1)\}, \mathcal{C}(ms(d_2)) = \{notMSD(d_2), nSD(d_2)\},$
  $\mathcal{C}(sd(d_1)) = \{nSD(d_1)\}, \mathcal{C}(sd(d_2)) = \{nSD(d_2)\},$
  $\mathcal{C}(redundant(a)) = \{notRedundant(a)\},$
  $\mathcal{C}(nAttr(d_1, a)) = \{hasAttr(d_1, a)\}, \mathcal{C}(nAttr(d_2, a)) = \{hasAttr(d_2, a)\},$
  $\mathcal{C}(nWorse(d_1, d_2)) = \{worse(d_1, d_2)\}, \mathcal{C}(nWorse(d_2, d_1)) = \{worse(d_2, d_1)\},$
  $\mathcal{C}(nMet(d_1, g)) = \{met(d_1, g)\}, \mathcal{C}(nMet(d_2, g)) = \{met(d_2, g)\}.$

Given this ABA framework, ms(d_1) is the claim of an argument in an admissible set, which is the defence set of the admissible dispute tree (adapted from the output of `proxdd`*) in Figure 2. The root argument of the tree (right-most rectangle labelled P:1) claims that d_1 is a MRSD decision. This claim is attacked by two different arguments (rectangles labelled O:6 and O:10 in Figure 2) as follows:*

- *O:6 represents an objection against d_1 on grounds that it does not achieve goal g; and*
- *O:10 represents an objection against d_1 on grounds that d_2 is strongly dominant and does not have presumably, the redundant attribute a.*

These two attacks are both counter-attacked by other arguments (rectangles labelled P:7 and P:11, respectively) as follows:

- *P:7 represents that d_1 achieves g; and*
- *P:11 represents that d_2 is not strongly dominant as it does not achieve g.*

Neither of these counter-attacks can be further attacked, and so the defence set of the dispute tree is admissible. Note that the "wrong" assumptions in O:6 and O:10 are essential to identify, via the counter-attacks P:7 and P:11, the reasons why d_1 is a MRSD decision and $ms(d_1)$ is a legitimate assumption (in an admissible set).

This example illustrates how the mapping onto ABA captures minimally redundant strong dominance by means of admissibility. In general, the following theorem gives the result that the mapping onto MRSD ABA frameworks gives an equivalent re-formulation for MRSD decisions.

Theorem 1. *Let AF be the MRSD ABA framework for $df = \langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$. Then, for each $d \in \mathtt{D}$, d is MRSD if and only if $\{ms(d)\} \vdash ms(d)$ belongs to an admissible set of arguments in AF.*

Proof. Let S be the set of all strongly dominant decisions and df' be the sub-framework of df with respect to S. Since both $\text{nSD}(d)$ and $\text{notMSD}(d)$ are contraries of $\text{ms}(d)$, the attacks against argument $\{\text{ms}(d)\} \vdash \text{ms}(d)$ have any of the following forms ($\forall d' \in \mathtt{D} \backslash \{d\}$, $g \in \mathtt{G}, a \in \mathtt{A}$):

$$\{\text{nMet}(d, g)\} \vdash \text{nSD}(d), \tag{13}$$

$$\{\text{nWorse}(d', d), \text{sd}(d'), \text{nAttr}(d', a), \text{redundant}(a)\} \vdash \text{notMSD}(d). \tag{14}$$

($\Longrightarrow$) With d being MRSD, we need to prove that $\forall d' \in \mathtt{D} \backslash \{d\}$, $g \in \mathtt{G}, a \in \mathtt{A}$, arguments (13) and (14) are counter-attacked. Since d is MRSD, there are arguments for $\text{met}(d, g)$ of the form $\emptyset \vdash \text{met}(d, g)$, so argument (13) is counter-attacked by arguments supported by empty sets of assumptions. For each $d' \in S \backslash \{d\}$, with d being MRSD, by Definitions 4 and 2, there are two cases to be considered: (i) In the case of $\lambda(d) \subseteq \lambda(d')$, there are arguments for $\text{hasAttr}(d', a)$. Hence, argument (14) is also counter-attacked by arguments supported by the empty set of assumptions. (ii) In the case that $\exists a' \in \mathtt{A}$ such that $a' \notin \lambda(d) \wedge a \in \lambda(d')$, there are arguments for $\text{worse}(d', d)$. Hence, argument (14) is counter-attacked by arguments of the form $\{deviate(a), nAttr(d, a)\} \vdash worse(d, d)$. These arguments are defended since $a' \notin \lambda(d) \wedge a \in \lambda(d')$. While $\forall d' \notin S \backslash \{d\}$, there are arguments for $\text{nSD}(d')$, so argument (14) is also counter-attacked by arguments of the form $\{nMet(d, g)\} \vdash nSD(d)$. These arguments are not attacked since d is not strongly dominant and there exists goal g which d does not achieve.

($\Longleftarrow$) The attacks against argument $\{\text{ms}(d)\} \vdash \text{ms}(d)$ have form (13) or (14) above. If $\{\text{ms}(d)\} \vdash \text{ms}(d)$ belongs to an admissible set in AF, then these attacks are all counter-attacked. Therefore, d is strongly dominant and for each $\forall d' \in \mathtt{D} \backslash \{d\}$, there are arguments for $\text{worse}(d', d)$ or $\text{nSD}(d')$ or $\text{hasAttr}(d', a)$ or $\text{notRedundant}(a)$ (where $a \in \mathtt{A}$). Then $\forall d' \in S \backslash \{d\}$, $\exists a \in \mathtt{A}$ such that $a \in \lambda(d') \wedge a \notin \lambda(d)$ or $\lambda(d) \subseteq \lambda(d')$. Hence, d is MRSD. □

Theorem 1 sanctions that the mapping, given in Definition 7, from decision frameworks to MRSD ABA frameworks is sound and complete in the sense that MRSD decisions in a decision framework correspond to arguments in admissible sets in the MRSD ABA framework. The following corollary states that strongly dominant decisions in a decision framework also correspond to arguments in admissible sets in the corresponding MRSD ABA frameworks.

Corollary 1. *Let AF be the MRSD ABA framework for $df = \langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$. Then for each $d \in \mathtt{D}$, d is strongly dominant if and only if $\{sd(d)\} \vdash sd(d)$ is in an admissible set of arguments in AF.*

Proof. This result can be shown easily from the proof of Theorem 1. □

As we will see in Section 6, this result is useful when we compare decisions according to the several criteria defined in Section 4.

The problem of identifying MRWD decisions can be also mapped onto the problem of computing admissible sets of arguments in an ABA framework, given next.

Definition 8. *Let $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$ be the MRSD ABA framework for df* $= \langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{DA}, \mathtt{GA} \rangle$. *The* MRWD ABA framework *for df is $\langle \mathcal{L}, \mathcal{R}', \mathcal{A}', \mathcal{C}' \rangle$, where:*

- *$\mathcal{R}'$ is obtained from $\mathcal{R}$ in Definition 7 by replacing rule schemata (5) and (6) with rule schemata:*

$$\begin{aligned} &\mathit{notMWD}(D) \leftarrow \mathit{wd}(D') \wedge \mathit{worse}(D, D') \wedge \mathit{nWorse}(D', D), \\ &\mathit{nWD}(D) \leftarrow \mathit{met}(D', G) \wedge \mathit{nMet}(D, G) \wedge \mathit{nMore}(D, D'), \\ &\mathit{more}(D, D') \leftarrow \mathit{met}(D, G) \wedge \mathit{nMet}(D', G). \end{aligned}$$

- *$\mathcal{A}'$=$\mathcal{A}_w \cup \{\mathit{nMore}(d, d') \mid d, d' \in \mathtt{D}, d \neq d'\}$, where $\mathcal{A}_w$ is $\mathcal{A}$ in Definition 7 with (for any $d \in \mathtt{D}$) $\mathit{ms}(d)$ replaced by $\mathit{mw}(d)$ and $\mathit{sd}(d)$ replaced by $\mathit{wd}(d)$.*

- *$\mathcal{C} : \mathcal{A}' \rightarrow 2^{\mathcal{L}} \setminus \{\emptyset\}$ is given by: for any $d, d' \in \mathtt{D}$ (such that $d \neq d'$):*

$$\begin{aligned} &\mathcal{C}'(\mathit{nMore}(d, d')) = \{\mathit{more}(d, d')\}, \\ &\mathcal{C}'(\mathit{mw}(d)) = \{\mathit{notMWD}(d), \mathit{nWD}(d)\}, \\ &\mathcal{C}'(\mathit{wd}(d)) = \{\mathit{nWD}(d)\}, \end{aligned} \tag{10'}$$

and $\mathcal{C}'(\alpha) = \mathcal{C}(\alpha)$, with $\mathcal{C}$ as in Definition 7, for all other assumptions in $\mathcal{A}'$.

Note that (10′) in the above definition is obtained by replacing ms(d), nSD(d), notMSD(d), and sd(d) in (10) in Definition 7 by mw(d), nWD(d), notMWD(d), and wd(d), respectively.

The theorem below sanctions a form of soundness and completeness for MRWD ABA frameworks:

Theorem 2. *Let AF be the MRWD ABA framework for df. Then for each $d \in \mathtt{D}$, d is MRWD if and only if $\{\mathit{mw}(d)\} \vdash \mathit{mw}(d)$ belongs to an admissible set of arguments in AF.*

This theorem follows from Definitions 4 and 8 as well as the definition of admissibility in ABA (see Section 2). Its proof is omitted since it is very similar to that of Theorem 1.

We illustrate the use of Theorem 2 in the context of our legal example, as follows:

Example 8 (Example 6 continued)**.** *Given the MRWD ABA framework for our legal df (obtained as per Definition 8), $\{\mathit{mw}(d_2)\} \vdash \mathit{mw}(d_2)$ belongs to an admissible set, as shown by the fragment, given in Figure 3, of an admissible dispute tree adapted from the output of* `proxdd`*. This tree illustrates the explanation aspect of the ABA mapping. The root argument of the tree (right-most rectangle labelled P:1) claims that d_2 (case No. 97) is MRWD. This claim is attacked by three different arguments (rectangles labelled O:23 - O:25), giving reasons why d_2 may not be MRWD as follows:*

- *(O:23) d_8 (case No. 1962) is better than d_2 (case No. 97);*
- *(O:24) d_2 is not weakly dominant since d_8 achieves goal g_1 (*i.e.*, older than 18) that is not achieved by d_2; and*

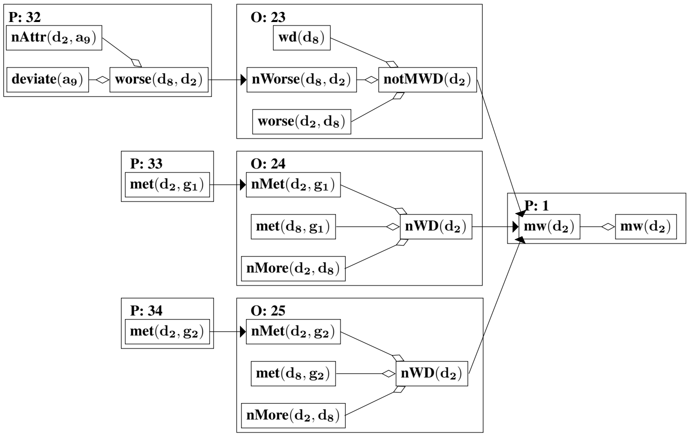

Figure 3: A fragment of an admissible dispute tree for our legal example.

- *(O:25) d_2 is not weakly dominant since d_8 achieves goal g_2 (*i.e.*, extremely huge amount) that is not achieved by d_2.*

These three arguments are counter-attacked by other arguments (rectangles labelled P:32, P:33, and P:34, respectively), as follows:

- *(P:32) d_8 has redundant attribute a_9 (*i.e.*, accessory) but d_2 does not, so d_8 is worse than d_2;*
- *(P:33 / P:34) d_2 achieves goal g_1 / g_2 (respectively).*

Indeed, as we have seen earlier, case No. 97 (d_2) is the case that is most similar to the goal case, because it is the only MRWD decision. Note that only a fragment of the dispute tree is shown here. The full tree contains several other arguments of the form:

$$\{\mathit{nMet}(d_2, g), \mathit{met}(d, g), \mathit{nMore}(d_2, d)\} \vdash \mathit{nWD}(d_2),$$
$$\{\mathit{wd}(d), \mathit{nWorse}(d, d_2), \mathit{worse}(d_2, d)\} \vdash \mathit{notMWD}(d_2),$$

for various $d \in \mathtt{D}$ but $d \neq d_2$, and $g \in \mathtt{G}$. These arguments are all counter-attacked by arguments similar to the ones shown as P:32, P:33, and P:34.

As shown in the examples, the ABA reformulations of the various decision criteria we propose allows us to use dispute trees as explanations for why these criteria sanction some decisions as best. In addition, these dispute trees pave the way towards linguistic explanations, as defined next.

6. Natural Language Explanation

In addition to giving a comprehensive explanation to best decisions, dispute trees can be used to provide explanations when comparing decisions, *e.g.* for giving answers

Table 9: Abbreviations for strings.

Abbreviation	**Full string**
$eq(d, d')$	d is as good as d'
$better(d, d')$	d is **better** than d'
$bmall(d, d')$	because **b**oth d and d' **m**eet **all** goals
$fmall(d, d')$	because d (the **f**irst decision) **m**eets **all** goals but d' (the second decision) does not
$bmnot\_met(d, d')$	because **b**oth d and d' **m**eet goals **not met** by the other
$fmmore(d, d')$	because d (the **f**irst decision) **m**eets **more** goals than d' (the second decision)
$nmmost(d, d')$	because **n**either d nor d' **m**eet **most** goals, moreover, d does not achieve more goals than d', and vice versa
$bfewest\_ra(d, d')$	because **b**oth d and d' have **fewest** redundant **a**ttributes
$ffewer\_ra(d, d')$	because d (the **f**irst decision) has **fewer** redundant **a**ttributes than d' (the second decision)

to questions such as *Why is a decision* d *more preferred to another* d′*?* Answering such questions transparently is important for human decision makers as it brings trust to the computer system recommended decisions. In this section, we present an algorithm to generate natural language explanations by extracting information from dispute trees. The explanations that we generate answer pairwise comparison questions. The presented algorithm is derived from the decision ranking scheme presented in Section 4, in the sense that it justifies a decision as *better than* or *as good as* another if the former is higher or the same in that ranking.

Based on our mapping from decision framework to ABA in Section 5, *good* decisions in different categories, *e.g.* MRSD or MRWD, are accompanied by different corresponding ABA frameworks. Dispute trees with, as roots, arguments with claims ms(d) or mw(d) contain all the required information for pairwise comparisons.

Algorithm 1 shows the main process for generating text for explaining the rationale for decision-making, with Algorithms 2 and 3 defining subprocedures used in Algorithm 1. In defining these algorithms, we adopt the following conventions, with respect to the relevant ABA framework: (i) a sentence $s \in \mathcal{L}$ is admissible if and only if s is the claim of an argument in some admissible set; (ii) the defence set of a sentence $s \in \mathcal{L}$ is the defence set of an admissible dispute tree, the root of which is an argument with claim s; (iii) $\forall s \in \mathcal{L}$ such that s is admissible, there is a total order over the admissible dispute trees, the root of which is an argument with claim s, and DT(s) denotes the first admissible dispute tree according to this order (we will refer to DT(s) as the "chosen" dispute tree for s); and (iv) $+$ represents string concatenation.

Moreover, we will make use of the abbreviations for strings given in Table 9. Further, the algorithms make use of natural language translations of the support of arguments, defined as follows:

- Given an argument for nSD(d) labelled $P : x$ in a chosen dispute tree, when its support contains assumption nMet(d, g), **nl-Support-for-nSD**$(d, P : x)$ denotes "decision d does not achieve goal g".

Algorithm 1 Comparison between decisions d and d'.

```
if $ms(d)$ is admissible in the MRSD ABA framework then
    COMPARE1($d$, $d'$, DT($ms(d)$))
else if $mw(d)$ is admissible in the MRWD ABA framework then
    COMPARE2($d$, $d'$, DT($mw(d)$))
else if $sd(d)$ is admissible in the MRSD ABA framework then
    if $ms(d')$ is admissible in the MRSD ABA framework then
        COMPARE1($d'$, $d$, DT($ms(d')$))
    else if $sd(d')$ is admissible in the MRSD ABA framework then
        output $eq(d, d') + bmall(d, d') +$ "."
    else
        output $better(d, d') + fmall(d, d') +$ ". For example: "+
                nl-Support-for-nSD($d'$, $P : 1$) + "."
    end if
else if $wd(d)$ is admissible in the MRWD ABA framework then
    if $mw(d')$ is admissible in the MRWD ABA framework then
        COMPARE2($d'$, $d$, DT($mw(d')$))
    else if $wd(d')$ is admissible in the MRWD ABA framework then
        output $eq(d, d') + bmnot\_met(d, d') +$ "."
    else
        output $better(d, d') + fmmore(d, d') +$ ". For example: "+
                nl-Support-for-more($d$, $d'$, $P : 1$) + "."
    end if
else
    if $wd(d')$ is admissible in the MRWD ABA framework then
        output $better(d', d) + fmmore(d', d) +$ ". For example: "+
                nl-Support-for-more($d'$, $d$, $P : 1$) + "."
    else
        output $eq(d, d') + nmmost(d, d') +$ "."
    end if
end if
```

- Given an argument for worse(d', d) labelled $P : x$ in a chosen dispute tree, when its support contains assumption nAttr(d, a), **nl-Support-for-worse**$(d', d, P : x)$ denotes "decision d does not have redundant attribute a but decision d' does".
- Given an argument for more(d, d') labelled $P : x$ in a chosen dispute tree, when its support contains assumption nMet(d', g), **nl-Support-for-more**$(d, d', P : x)$ denotes "decision d' does not achieve goal g but decision d does".
- Given an argument for nWD(d) labelled $P : x$ in a chosen dispute tree, when its support contains assumptions nMet(d, g) and nMore(d, d'), then we make use of **nl-Support-for-nWD**$(d, d', P : x)$ to denote "on the one hand, decision d does not achieve goal g but decision d' does; on the other hand, decision d does not achieve more goals than decision d'".

Here and subsequently, x in the label $P : x$ of an argument in a dispute tree denotes an element of the set of natural numbers $\mathbb{N}$, with $x = 1$ if the argument in the root. With an abuse of notation, when there is no ambiguity, we also use x to denote the argument labelled with $P : x$.

With the intuition of our ranking given in Section 4, for any decisions d and d', Algorithm 1 works as follows. (i) if one of ms(d) and ms(d') is admissible in the

Algorithm 2 Comparison between d and d' based on dispute tree with root argument $\{ms(d)\} \vdash ms(d)$ with respect to the MRSD ABA framework.

```
 1: procedure COMPARE1($d$, $d'$, $t$)
 2:     comment: $t$ is $DT(ms(d))$
 3:     if $ms(d')$ is admissible in the MRSD ABA framework then
 4:         output $eq(d,d') + bmall(d,d')$ + " and " + $bfewest\_ra(d,d')$ + "."
 5:     else if $nSD(d')$ is the claim of a proponent's argument $P : x$ in $t$ then
 6:         output $better(d,d') + fmall(d,d')$ + ". For example: "+
                nl-Support-for-nSD($d'$, $P : x$) + "."
 7:     else
 8:         if $worse(d', d)$ is the claim of a proponent's argument $P : y$ in $t$ then
 9:             output $better(d,d') + ffewer\_ra(d,d')$ + ". For example: "+
                    nl-Support-for-worse($d'$, $d$, $P : y$) + "."
10:         else
11:             output $eq(d,d') + bmall(d,d')$ + "."
12:         end if
13:     end if
14: end procedure
```

corresponding MRSD ABA framework, then the algorithm generates explanations by making use of the chosen admissible dispute tree according to Algorithm 2 (see lines 2 and 7 in Algorithm 1); (ii) else if one of mw(d) and mw(d') is admissible in the MRWD ABA framework, then the algorithm generates explanations by making use of the chosen admissible dispute tree according to Algorithm 3 (see lines 4 and 15 in Algorithm 1); (iii) else the explanation is constructed from the chosen admissible dispute tree of nSD(d) in the MRSD ABA framework (*e.g.* see line 11 in Algorithm 1), or from the chosen dispute tree of more(d, d') in MRWD ABA framework (*e.g.* see line 19 in Algorithm 1).

Moreover, to efficiently generate text for explaining why a particular decision is better than another, instead of translating the whole chosen tree, only some specific parts of the tree need to be looked at. Thus, Algorithms 2 and 3 just extract useful information from some chosen dispute tree to generate explanations for pairwise comparison. For instance, Algorithm 2 checks arguments for nSD(d') in the chosen admissible dispute tree DT(ms(d)) with respect to the MRSD ABA framework. If there exists a proponent argument for nSD(d'), because of the admissibility of the chosen dispute tree DT(ms(d)), an explanation "d' is not strongly dominant but d is" can support that "d is better than d'" (see lines 5-6 in Algorithm 2). Similarly, arguments for nWD(d') in Algorithm 3 are checked in the chosen dispute tree (see lines 5-12 in Algorithm 3). Moreover, proponent arguments for worse(d', d) in the chosen dispute tree are useful for explaining why a decision is better than another (see lines 8-12 in Algorithm 2 and lines 14-18 in Algorithm 3).

Let us illustrate our algorithm with the running legal example as follows.

Example 9 (Example 8 continued)**.** *From earlier discussions, we know that d_2 (case No. 97) is the best decision. Comparing d_2 with d_8 (case No. 1962), Algorithm 1 can provide an explanation. Firstly, since the argument $\{mw(d_2)\} \vdash mw(d_2)$ is admissible as d_2 is MRWD, the procedure for comparing d_2 and d_8 based on the dispute tree of mw(d_2) is invoked (line 4 in Algorithm 1). According to Algorithm 3 and the dispute*

Algorithm 3 Comparison between d and d' based on dispute tree with root argument $\{mw(d)\} \vdash mw(d)$ with respect to the MRWD ABA framework.

```
 1: procedure COMPARE2($d$, $d'$, $t$)
 2:     comment: $t$ is $DT(mw(d))$
 3:     if $mw(d')$ is admissible in the MRWD ABA framework then
 4:         output $eq(d, d') + bmnot\_met(d, d') +$ " and " $+ bfewest\_ra(d, d') +$ "."
 5:     else if $nWD(d')$ is the claim of a proponent's argument $P : x$ in $t$ then
 6:         if $d''$ does not appear in $t$ then
 7:             output $better(d, d') +$ " because, " $+$ nl-Support-for-nWD($d', d, P : x$) $+$ "."
 8:         else if $mw(d'')$ is admissible in the MRWD ABA framework then
 9:             output $better(d'', d') +$ " because, " $+$ nl-Support-for-nWD($d', d'', P : x$)$+$
                    ". Moreover, " $+ eq(d, d'') + bmnot\_met(d, d'') +$ "and " $+ bfewest\_ra(d, d'')+$
                    ". Thus, " $+ better(d, d') +$ "."
10:         else
11:             output $better(d, d'') + ffewer\_ra(d, d'') +$ ". Moreover, " $+ better(d'', d')+$
                    " because, "$+$ nl-Support-for-nWD($d', d'', P : x$) $+$ ". Thus, " $+ better(d, d') +$ "."
12:         end if
13:     else
14:         if $worse(d', d)$ is the claim of a proponent's argument $P : y$ in $t$ then
15:             output $better(d, d') + ffewer\_ra(d, d') +$ ". For example: "$+$
                    nl-Support-for-worse($d', d, P : y$) $+$ "."
16:         else
17:             output $eq(d, d') + bmnot\_met(d, d') +$ "."
18:         end if
19:     end if
20: end procedure
```

tree of mw(d_2) shown in Figure 3, since the argument $\{mw(d_8)\} \vdash mw(d_8)$ is not admissible as d_8 is not MRWD, and nWD(d_8) *is not the conclusion of any proponent argument (indicating that d_8 is not weakly dominant), whether or not* worse(d_8, d_2) *is the claim of any argument in a proponent node will be checked (see line 14 in Algorithm 3). Since* worse(d_8, d_2) *is the claim of the argument labelled with P:32, which contains* deviate(a_9)*, indicating that d_8 has more redundant attributes (a_9) than d_2, the system outputs:*

> d_2 is better than d_8 because d_2 has fewer redundant attributes than d_8. For example, d_2 does not have redundant attribute a_9 but d_8 does.

Furthermore, when being posed with the question: "why is d_2 (case No. 97) better than d_1 (case No. 245)?*", our algorithm gives:*

> d_2 is better than d_1 because, d_1 does not achieve goal "the defendant is older than 18" (g_1) but d_2 does, and decision d_1 does not achieve more goals than d_2.

The following theorem shows that our algorithms for generating natural language explanations match the decision ranking mechanism described in Section 4:

Theorem 3. *Given the ordering as defined in Definition 5, for any two decisions $d, d' \in$* D*, $d \succ d'$ if and only if Algorithm 1 outputs text containing* better(d, d')*; and $d \sim d'$ if and only if Algorithm 1 outputs text containing* eq(d, d').

Proof. (i) For any decisions $d, d' \in \mathtt{D}$, if $d \succ d'$, here we just consider the case of d is strongly dominant and d' is not, other cases can be proved similarly. Since d is strongly dominant and d' is not, we need to consider two cases: (a) If $\mathrm{ms}(d)$ is admissible, since d' is not strongly dominant, there must exist an proponent's argument for $\mathrm{nSD}(d')$ in the chosen dispute tree of $\{\mathrm{ms}(d)\} \vdash \mathrm{ms}(d)$ in the MRSD ABA framework, and thus according to Algorithm 2 (line 6), the output is: *better*(d,d')+$\mathit{fmall}(d,d')$ + "For example: "+**nl-Support-for-nSD**$(d', P : x)$+"."; and (b) if $\mathrm{ms}(d)$ is not admissible, by Lemma 1, we have d not MRWD, and thus by Algorithm 1 (line 11), the output is $better(d, d')$ +$\mathit{fmall}(d, d')$+"For example: "+**nl-Support-for-nSD**$(d', P : 1)$+".". And if Algorithm 1 outputs text containing $better(d, d')$, here we just consider the case as shown on line 6 in Algorithm 2 (*i.e.*, $better(d, d')$+$\mathit{fmall}(d, d')$+"For example: " + **nl-Support-for-nSD**$(d', P : x)$+".") (other cases can be proved similarly). We know that $\mathrm{nSD}(d')$ is the claim of a proponent's argument $(P : x)$ in the chosen admissible dispute tree with root $\{\mathrm{ms}(d)\} \vdash \mathrm{ms}(d)$. It is easy to see that $sd(d')$ is inadmissible. By Corollary 1, d' is not strongly dominant. Thus, since d is strongly dominant, we have $d \succ d'$.

(ii) For any decisions $d, d' \in \mathtt{D}$, if $d \sim d'$, we just focus on the case when $d \sim d'$, because both d and d' are MRSD (other cases can be proved similarly). By Theorem 1, we have $\mathrm{ms}(d)$ and $\mathrm{ms}(d')$ are admissible. Thus, according to Algorithm 2 (line 4), the output is $eq(d, d')$+$bmall(d, d')$ +" and " +$\mathit{bfewest\_ra}(d, d')$+".". And if Algorithm 1 outputs text containing $eq(d, d')$, here we just consider the case of output as in line 9 in Algorithm 1 (*i.e.*, $eq(d, d')$+$bmall(d, d')$+"."), other cases can be proved similarly. We know that both $sd(d)$ and $sd(d')$ are admissible. By Corollary 1, d is strongly dominant and d' is strongly dominant. Thus, by Definition 5, we have $d \sim d'$. □

7. Evaluation

This section presents an empirical study of our argumentation-based model of decision-making and explaining in the context of identifying most similar legal cases to a new case.

7.1. Hypotheses

In Section 3, we presented an illustration from the legal domain (see Examples 5 and 6), where past cases most similar to a given case are considered as best decisions. Thus, firstly we expect that, based on their professional experience, legal practitioners will agree that the case recommended by our minimal redundancy criterion is indeed the most similar alternative. That is, we have:

Hypothesis 1. *Our recommendation based on the idea of minimal redundancy is consistent with subjects' judgements based on their professional experience.*

Secondly, our explanations are based on the assumption that in influencing users' favorability and trust towards our recommendations, it is crucial to transparently explain why one case is more similar to an open case than another. In other words, we expect that our natural language explanations are useful for users to analyse the similarities between legal cases. So, we need to figure out how much users find our explanations helpful. Thus, we propose:

Hypothesis 2. *Among those users who have non-neutral attitudes about the usefulness of natural language explanations, a higher ratio of users rate positively than negatively.*

Thirdly, in order to evaluate the usefulness of our natural language explanations, also it is required to compare our natural language explanations to other alternative explanation techniques (Hoffmann, 2005). One of those alternatives is debate tree explanations, which is a standard output of argumentation engines. However, before we choose debate tree explanation as our target of comparison, we need to make sure it is meaningful (*i.e.*, debate tree explanations are useful for users to analyse the similarities between legal cases). Thus, we propose:

Hypothesis 3. *Among those users who have non-neutral attitudes about the usefulness of debate tree explanations, a higher ratio of users rate positively than negatively.*

Fourthly, the major difference between our natural language explanations and debate tree explanations is that with natural language explanations we can easily highlight similar and distinct factors between cases. Such highlighting cannot as easily be achieved using debate tree explanations, because debate trees prioritise the presentation of argumentation structures. As a result, after testing the above hypothesis, we should test the following one:

Hypothesis 4. *Our natural language explanations are more useful than graphical explanations via debate trees.*

7.2. Subjects

First, we need to determine the sample size of subjects. We follow the principle for evaluating knowledge based systems (such as ours in this paper), proposed in Menzies (1998):

> *Sample size (N) should be carefully controlled. Small sample sizes are hard to analyse. However, as random sample sizes get larger, they approach a bell shape (the normal distribution) that is a well-understood distribution. In practise, N greater than 20 is acceptable and N greater than 30 is encouraged. On the other hand, there may be no benefit to make N very large (Cohen also argues that sample sizes of N greater than 50 can be pointless (see Cohen (1995), p. 116)).*

Thus, we chose a sample size between 30 and 50 in our experiments. Specifically, we got 33 lawyers and judges, from seven different affiliations in the cities of Guangzhou and Foshan in China. We explained to them that the experiments aimed to evaluate the effectiveness of our methods. We did not train them in navigating debate trees, but conveyed to them the basic intuition for reading and understanding these trees.

7.3. Experimental procedure

Since the sample size is limited in our study, the repeated measure design could reduce the variance of estimates of treatment-effects, allowing statistical inference to be made with fewer subjects (Winkens et al., 2006; Vickers, 2003). According to Winkens et al. (2006), it is more efficient to have more than two repeated measures. Thus, we

Table 10: First Task.

No.	Case	Sentence
245	Li, the defendant, aged between 16 and 18, stole some private property in collusion with others repeatedly, and the value of the stolen goods was huge.	1 year imprisonment with ¥1,000 fine
97	Xie, the defendant, was older than 18, stole two trailers, property of Foshan AA Company and valuing 146,570 Chinese Yuan (*e.g.* an extremely huge amount of private property). The stolen goods could not be found.	10 years imprisonment, 3 years deprivation of political right and ¥10,000 fine
633	Liao, the defendant, was older than 18, and stole some private property in collusion with others. The value of the stolen goods was large (3480 Chinese Yuan). The stolen goods could not be found.	5 months imprisonment with ¥1,000 fine

New Case: Lu, the defendant, was older than 18 when he stole goods worth an extremely huge amount. The stolen goods have been found.

decided to have each experiment consisting of three tasks (see Tables 10-12). Each task includes a new case, without any sentence, and a table of past, real cases about theft, from the Nanhai District People's Court in the city of Foshan, Guangdong Province, China, with the sentence given by the Court. To make the assessment as objective as possible, all the cases in each task were selected randomly.

After presenting the outcome of our method, in terms of the most similar past case identified using the notions in Section 3.2, the subjects were asked:

Question 1. *How much do you agree that the past case is indeed the most similar, using a 5-point scale (*i.e.*, 5=strongly agree, 4=agree, 3=neither agree nor disagree, 2=disagree, and 1=strongly disagree)?*

After presenting the natural language explanation for comparison, as computed by the algorithms in Section 6, the subjects were asked:

Question 2. *How much do you agree that these comparisons are useful to analyse the similarities between legal cases (again according to the 5-point scale above)?*

After presenting the comparison using debate trees (*e.g.* Figure 4), offering a graphical view of how a similar case is chosen by our method of minimal redundancy, the subjects were asked:

Question 3. *How much do you agree that these debate trees are useful for you to analyse the similarities between legal cases (again according to the 5-point scale above)?*

Question 4. *Do you agree that the natural language explanation is more useful than debate trees, seen as explanations, on the same 5-point scale?*

Thus, the measures for Hypothesis 1 were collected through Question 1 and subjects' choices of the most similar case. Since Question 1 is related to the intuition of best decision, which has nothing to do with the fact that such an intuition reading is automatically generated or not, the subjects were not asked to compare the automatically

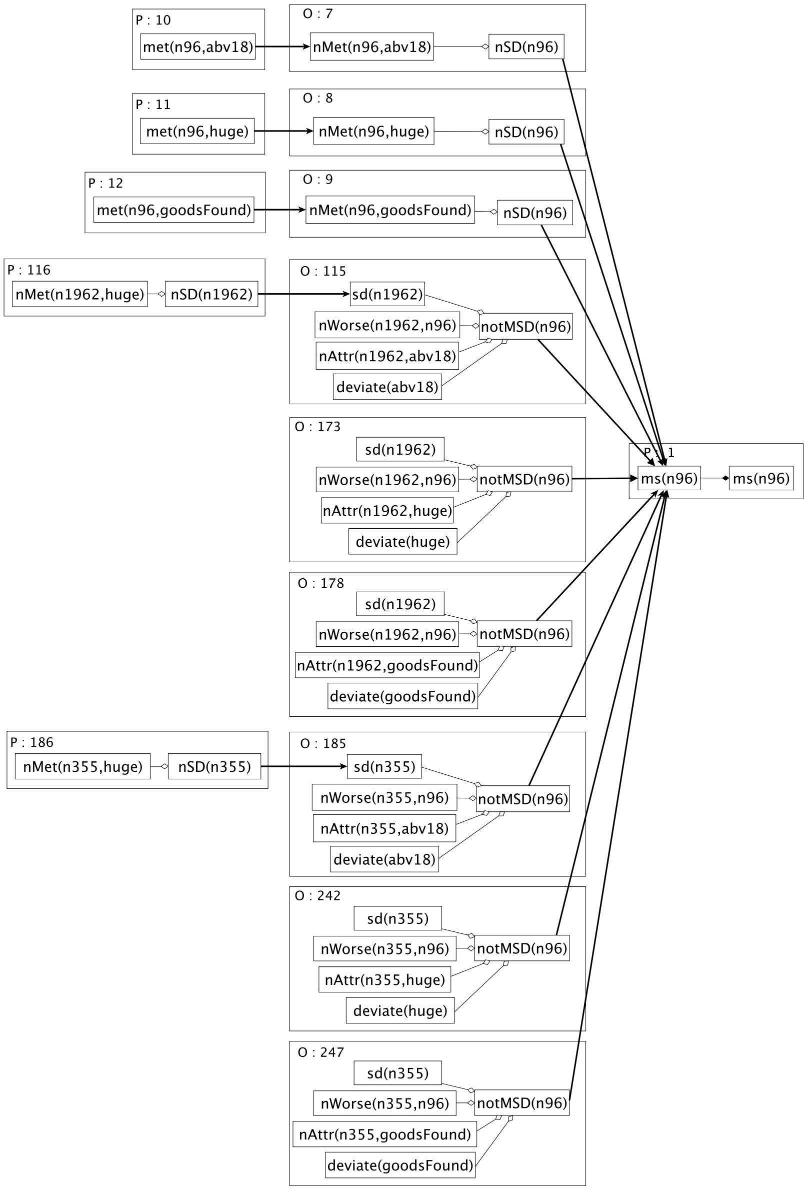

Figure 4: Example of debate tree used in experiments, for Questions 3 and 4.

Table 11: Second Task.

No.	Case	Sentence
1962	Zhou, the defendant, was older than 18, stole some private property in collusion with others, and the value of the stolen goods was extremely huge. Others played a leading role while Zhou was referred to as accessory. The stolen goods have been found.	7 years imprisonment with ¥7,000 fine
96	Chen, the defendant, was older than 18 and stole a Dongfeng van (worth 26736 Chinese Yuan) in collusion with others. The stolen goods are worth a huge amount and have been found.	3 years and 3 months imprisonment with ¥3,000 fine
355	Lu, the defendant, was older than 18 and stole goods worth an extremely huge amount. The stolen goods have been found.	10 years imprisonment, 3 years deprivation of political right and ¥10,000 fine

New Case: Mao, the defendant, was older than 18 when steeling some private property of the AA Bidding Company. The value of the stolen goods (found) was huge.

generated reading with the hand-crafted one. Questions 2 and 3 were used to collect measures for the usefulness of natural language explanations (*i.e.*, Hypothesis 2) and debate tree explanation (*i.e.*, Hypothesis 3), respectively. Hypothesis 4 was tested in two ways: (i) analysing the measures collected from answers to Question 4, and (ii) statistically comparing the measures collected from answers to Questions 2 and 3.

7.4. Hypotheses Test's Results

In this subsection, we test the four hypotheses on the basis of questionnaire data.

Figure 5 shows the overall average ratings elicited from the 33 subjects for each of the four questions. The average rating for each question is about 3.65 out of 5 (precisely, 3.90, 3.85, 3.40 and 3.83, respectively). This indicates that, in average, the subjects found our method of selecting the most similar case by minimal redundancy decision criterion acceptable, and for our natural language explanation, they agreed that it is useful and preferred it to debate trees.

To test Hypothesis 1, we need to investigate how many times subjects marked the same case as the most similar as recommended by our method. From the data shown in Table 13, we can conclude that, for each task, among those subjects who have selected what they believed to be the most similar cases (*i.e.*, 30, 31, and 31 subjects, respectively), more than 84% of them marked the same case as recommended by our method. Apart from the behavioural measurement, through Question 1, we asked the subjects to self-report their satisfaction with our recommendation. The average rating from them in all tasks is 3.9 out of 5 (Standard Deviation (SD) is 0.87). From Figure 6, we can see that for each task, at least 70% of the subjects either agreed or strongly agreed that the recommended case is the most similar to the new case, while less than 12% disagreed or strongly disagreed with this. After examining Cronbach alpha measures of internal consistency (*i.e.*, reliability) for responses collected from three tasks, it is reasonable for us to combine three responses to form a single index s for further analysis

Table 12: Third Task.

No.	Case	Sentence
751	Mao, the defendant, was older than 18 and stole some private property of the XX Bidding Company. The value of the stolen goods was huge. The stolen goods have been found.	3 years and 3 months imprisonment with ¥3,000 fine
801	Wen, the defendant, was older than 18, and stole some private goods in collusion with others. The value of the stolen goods was large (2,720 Chinese Yuan). Since the defendant committed the crime of theft within 5 years after serving a sentence of 1 year and 2 months of imprisonment due to committing a crime of theft on 7 April 2009, Wen is deemed a recidivist.	6 months imprisonment with ¥1,000 fine
802	Qiu, the defendant, was older than 18 and stole some private goods in collusion with others. The value of the stolen goods was large (6,100 Chinese Yuan). The stolen goods have been found.	6 months imprisonment with ¥1,000 fine

New Case: Yang, the defendant, was older than 18 when he stole some private property of the AA Company. The value of the stolen goods (found) was large. Since the defendant committed the crime of theft within 5 years after serving a sentence of 7 years of imprisonment due to illegal possession of drugs on 26 October 2006, Yang is a recidivist.

Table 13: Subject marked the same case as recommended by our method.

Task	Valid	Frequency	Percent (%)
1	30	29	97
2	31	26	84
3	31	30	97

(α = .765) (Zhou et al., 2008; Domino and Domino, 2006). A chi-square goodness-of-fit test was conducted with s to determine whether an equal number of non-neutral ratings from types of positive and negative were received, and the minimum expected frequency was 15. The chi-square goodness-of-fit test indicates that there is statistically significant difference in the subjects' ratings of positive and negative among those non-neutral ratings (87% versus 13% of 30 non-neutral ratings, $\chi^2(1) = 16.13, p <$ .0005). Therefore, the fact that a majority of votes are in favour of a positive answer, together with the result of the chi-square goodness-of-fit test, concludes a support in favour of Hypothesis 1.

Through Question 2, we can figure out whether or not the subjects found our natural language explanation useful (Hypothesis 2). The average rating from the 33 subjects on the usefulness of natural language explanation is more than 3.7 out of 5 for each task (for the first task: $M = 3.73$, $SD = 1.008$; for the second one: $M = 4.00$, $SD = .750$; and for the third: $M = 3.82$, $SD = .917$). According to the feedback shown in Figure 7, for each task, at least 67% of the subjects either agreed or strongly agreed that our natural language explanation is useful for analysing the similarities between cases,

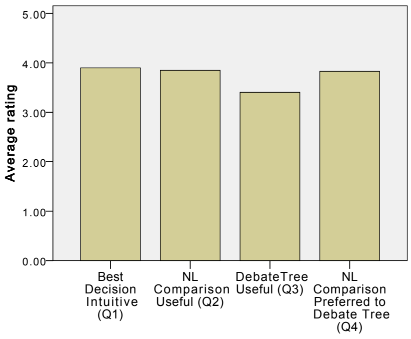

Figure 5: Average ratings across the 33 subjects.

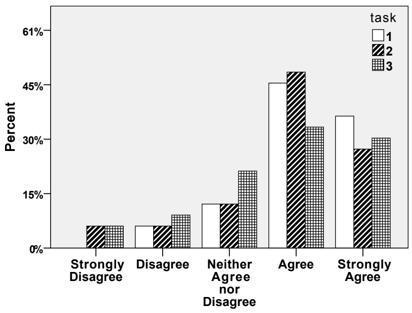

Figure 6: Responses to Question 1.

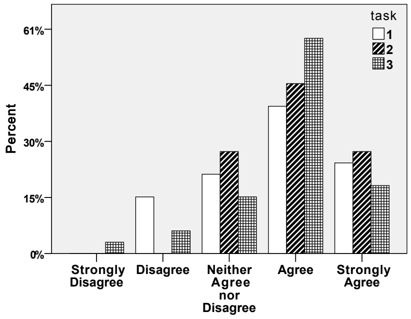

Figure 7: Responses to Question 2.

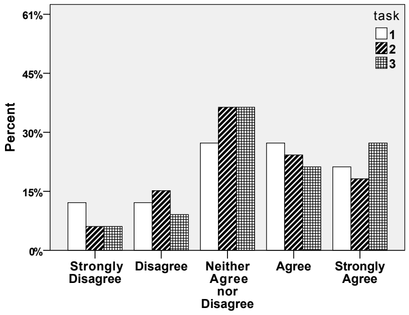

Figure 8: Responses to Question 3.

while less than 15% of them disagreed or strongly disagreed with this. Moreover, after examining Cronbach alpha measures of internal consistency for responses collected from three tasks, we combine three responses to form a single index nl for further analysis on the usefulness of natural language explanation ($\alpha = .869$) (Zhou et al., 2008; Domino and Domino, 2006). A chi-square goodness-of-fit test was done with nl to determine whether an equal number of non-neutral ratings about the usefulness of natural language explanations from types of positive and negative are received. Since we collected 31 non-neutral ratings about the usefulness of natural language explanations, the minimum expected frequency is $\frac{31}{2} = 15.5$. From the result of chi-square goodness-of-fit test, we can conclude that among those non-neutral ratings there is a statistically significant difference in the subjects' ratings of positive and negative (83% versus 17% of 31 non-neutral ratings, $\chi^2(1) = 14.23$, $p < .0005$). So, a majority of votes in favour of a positive answer, together with the result of the chi-square goodness-of-fit test, concludes a support in favour of Hypothesis 2.

Similarly, through Question 3 we can analyse whether or not the subjects found that debate trees are useful for their analysis (Hypothesis 3). Contrary to the natural language explanations, the collected responses shared higher standard deviations and the average rating from the 33 subjects is less than 3.6 out of 5 for each task (for the first task: M = 3.33, SD = 1.291; for the second task: M = 3.33, SD = 1.137; and for the third task: M = 3.55, SD = 1.175). As shown in Figure 8, for each task, except

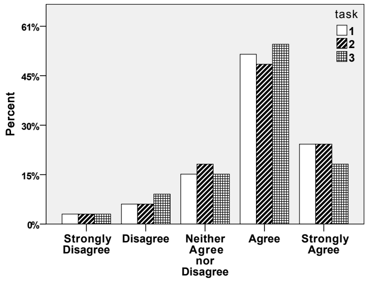

Figure 9: Responses to Question 4.

the neutral ratings, less than 20% of the 33 subjects chose the negative ratings (*i.e.*, strongly disagree or disagree) while a larger percentage of subjects (more than 47%) chose "agree" or "strongly agree". Moreover, we combined three responses based on different tasks to form a single index dt for further analysis on the usefulness of debate tree explanations (α = .965) (Zhou et al., 2008; Domino and Domino, 2006). From the result of chi-square goodness-of-fit test, we can conclude that there is a statistically significant difference in subjects' ratings of positive and negative among those non-neutral ratings (71% versus 29% of 24 non-neutral ratings, $\chi^2(1) = 4.17$, $p = .041$). Hence, Hypothesis 3 is supported.

As stated before, Hypothesis 4 is tested in two ways: (i) analysing the measures collected from subjects' self-report on Question 4, and (ii) statistically comparing the measures collected from Questions 2 and 3. Now we discuss the first way first. The average rating from the 33 subjects is more than 3.76 out of 5 for each task (for the first task: $M = 3.88$, $SD = .96$; for the second task: $M = 3.85$, $SD = .97$; and for the third task: $M = 3.76$, $SD = .97$). From Figure 9, we can see that in each task, more than 20% of the subjects strongly agreed that natural language explanations are more useful than debate tree explanations, at least 52% of them also agreed with such a preference, while a quite low percentage (less than 9%) disagreed or strongly disagreed. We combined three responses based on different tasks to form a single index pr for further analysis on the preference of explanation types (α = .871) (Zhou et al., 2008; Domino and Domino, 2006). The chi-square goodness-of-fit test indicates that agreements and disagreements on the preference were not equally reported by the subjects (93% versus 7% of 28 non-neutral ratings, $\chi^2(1) = 20.57$, $p < .0005$). That is, most of the subjects believed that natural language explanations are more useful when analysing the similarities between legal cases.

Then we turn to discuss the second way for testing Hypothesis 4 statistically: the paired sample t-test was conducted on the combined ratings of usefulness for the natural language explanations (*i.e.*, nl) and for the debate tree explanations (*i.e.*, dt).[9] The

[9] Note that the ratings on the usefulness of natural language explanations is slightly skewed (skewness $= -0.493$, also see Figure 7). However, according to the rule of thumb "skewness values within ± 1.0 are considered relatively normal" Hahs-Vaughn and Lomax (2012), it can still retain a relatively normal distribution. The case of kurtosis statistic is similar. Thus, the normality assumption of the paired t-test has been

paired sample t-test, whose applications include repeated-measures designs that this study contain, is a statistical procedure used to determine whether the mean difference between two sets of observations is zero (Hahs-Vaughn and Lomax, 2012). The procedure for a paired sample t-test can be summed up in four steps:

(i) Calculate the sample mean:
$$\overline{d} = \frac{(nl_1 - dt_1) + (nl_2 - dt_2) + \cdots + (nl_n - dt_n)}{n} = 0.444.$$

(ii) Calculate the sample standard deviation:
$$\hat{\sigma} = \sqrt{\frac{((nl_1 - dt_1) - \overline{d})^2 + ((nl_2 - dt_2) - \overline{d})^2 + \cdots + ((nl_n - dt_n) - \overline{d})^2}{n-1}} = 1.227.$$

(iii) Calculate the test statistic:
$$t = \frac{\overline{d} - 0}{\hat{\sigma}/\sqrt{n}} = \frac{0.444 - 0}{1.227/\sqrt{33}} = 2.08.$$

(iv) Look up the value T in the t-distribution table with $n - 1$ degrees of freedom and compare t to T. In our case, we have $T = 2.037$ with $p < 0.05$ (two-tailed).

Since $t > T$, we can reject the null hypothesis (*i.e.*, the mean difference between both ratings for natural language explanations and debate tree ones is zero) and conclude that there is a significant difference in the ratings for natural language and debate tree explanations. This finding together with the higher mean of rating for natural language explanations (M=3.85, SD=0.8) than that for debate tree explanations (M=3.40, SD=1.2) concludes a support in favour of Hypothesis 4.

Putting all the above feedbacks together, we can conclude that, based on our experiments, our proposed minimal redundancy criterion is approved by most legal practitioners, and our argumentation-based natural language explanations can support well legal practitioners to analyse the similarities between legal cases.

8. Related work

In this section, we discuss in which way our work advances the state-of-art.

8.1. Argumentation-based decision-making

There are a number of studies on argumentation-based decision-making and analysis. Amgoud and Prade (2009) propose the first (to the best of our knowledge) general argument-based framework for decision-making and explaining. Our work differs from theirs in several aspects: (i) they base their decision model on abstract argumentation, whereas we base ours on ABA; and (ii) they rank decisions by counting pro and con arguments without any further explanations for pairwise comparison, whereas

met.

we rank decisions on the basis of decision-making criteria with natural language explanations for pairwise comparison provided. Matt et al. (2009) introduce a special family of ABA frameworks for reasoning about the benefits of decisions, and thus view, as we do, best decisions as dominant decisions corresponding, within the family of argumentation frameworks considered, to admissible arguments. However, our best decisions are not only dominant, but also minimally redundant. Moreover, our model can explain why decisions are best, or better than or as good as other decisions. Müller and Hunter (2012) present a decision analysis system based on argumentation. Their model is developed from the ASPIC+ framework (Prakken, 2010; Modgil and Prakken, 2014), whereas our work uses ABA. Furthermore, they do not present any algorithm for the generation of natural language explanations for decisions, as we did in this paper. Visser *et al*. Visser et al. (2012b,a, 2013) use logic-based instances of abstract argumentation (Dung, 1995), taking into account preferences over attributes as well as decision criteria. Differently from these approaches, our methods focus on providing (argumentative and natural language) explanations for best decisions, using as a starting point an argumentative counterpart, in terms of ABA, of the problem of determining best (minimally redundant strongly/weakly dominant) decisions. Fan and Toni (2013) present basic decision frameworks and criteria, and illustrate how goal preferences can be included in decision making in Fan et al. (2013). Our work in this paper builds on top of theirs with new decision criteria introduced and an algorithm for natural language explanation for explaining the selected decisions constructed, but without considering preferences over goals. Heras et al. (2013) propose a customer support framework using agent societies with case-based argumentation. Their agents reach agreements for best solutions by engaging in argumentation dilogues. However, they use value-based argumentation (Nawwab et al., 2008), whereas our work is based on ABA; also they present dialogue graphs to justify their reasoning process, while our work generates natural language explanations, which in some cases may be easier for ordinary human users to understand, as our experiments seem to suggest. Ferretti et al. (2017) propose a method to compare alternative decisions using argumentation frameworks, which is similar to our method. However, they use abstract argumentation frameworks whereas we use ABA. Moreover and more significantly, their work does not include any method for automatically generating explanations of pairwise comparisons.

8.2. Multi-attribute decision-making

There are two categories of multi-attribute decision-making (MADM) models. The first one consists of those in which each attribute/criterion is either absolutely satisfied or absolutely not satisfied by alterntive decisions. In the second one, attributes/criteria may be partially satisfied, to some extent, by some decisions, and in this case the problem is a fuzzy MADM problem. In contrast, approaches in the first category are called crisp MADM ones. Our method in this paper can be deemed to be crisp.

Most crisp MADM models do not explain which factors cause a decision to be better than another. Our model instead is explainable, in the sense discussed in the paper, and could also be seen as providing argumentation-based explanation for any crisp MADM model coinciding with our decision model. Whereas in some MADM settings, notably Labreuche (2011), explanation for a selected decision is achieved by analysing properties' weights together with decision scores, our focus is on natural language generation of pairwise comparison explanations rather than weight-based ones.

Recently several researchers proposed fuzzy MADM models to handle criteria's evaluations with various complex fuzzinesses. As an example, to deal with degrees to which an alternative decision satisfies criteria, Park et al. (2017) develop two approaches based on the concepts of entropy, cross-entropy, and similarity measure; Zhang (2016) propose one based on prioritised aggregation operator (more operators of this sort can be found in Luo et al. (2003b, 2015)), and Yu et al. (2016) also do so by developing several operators to aggregate all the criteria's evaluations. However, to the best of our knowledge, none of these approaches have any functionality for explaining the decisions they recommend.

It is interesting that Ceballos et al. (2017) study how to choose an appropriate model among a number of fuzzy MADM models to solve a given fuzzy MADM problem. Specifically, they examine several fuzzy MADM models against over 1,200 randomly generated decision problems, and reveal their similarities and differences, the impact of their parameters settings, and how they can be clustered. According to their analyses, one may be able to explain why to choose one instead of another method to solve a given fuzzy MADM problem, but the kind of explanation has nothing to do with the explanation for why one decision is made by a fuzzy MADM model. Instead, our work provides human users with a natural language explanations for recommended decisions.

There are other kinds of uncertain MADM models. For example, Ma et al. (2017) propose a MADM model to deal with ambiguity, *i.e.*, the imprecise and uncertain evaluation of criterion weights as well as the ambiguous evaluations of the groups of decisions regarding a given criterion. Moreover, they study how cognitive factors may cause a deviation from rational decisions. Instead, in our model best (minimally redundant and dominant) decisions are guaranteed.

8.3. Case-based reasoning

As illustrated in all the legal examples in the paper, our model supports case-based reasoning through a different mechanism from existing ones. For instance, though both HYPO (Ashley, 1991; Ashley et al., 2008) and our model can identify cases that are most similar to given cases, our approach also produces argumentative explanations in natural language. CATO (Aleven, 2003) (an extension of HYPO) can instead provide argumentative explanations for outcomes for new cases based upon similarity with past cases. Their arguments are in favour of an outcome based on factors shared with past cases and against an outcome based on factors different from past cases. Instead, our argumentative explanations, in the specific legal setting we have used for illustration, are used for identifying most similar cases, rather than outcomes. Moreover, our explanations are generated from an argumentative reformulation of a decision-making method. Additionally, unlike HYPO and CATO, which are specific to law, our model is generic and can be used in other domains (as our medical Example 4 illustrates). Prakken et al. (2015) give a reformulation of CATO in terms of argument schemes represented in ASPIC+ (Prakken, 2010; Modgil and Prakken, 2014), thus allowing in principle for the automatic generation of CATO explanations within a framework for computational argumentation. We have used ABA and its computational machinery as a method for the automatic generation of explanations, coupled with template-based natural language generation algorithms, for any decision-making setting where minimal redundancy may be useful.

As another example of a CBR proposal, Athakravi et al. (2014) study legal reasoning from past cases, which allow prediction of judgements for new cases, in the same spirit of our motivating example. Nonetheless, our approach differs from theirs in the following aspects: (i) we are concerned with determining and explaining most similar cases to a new one, while they focus on the extraction of relevant attacks from past cases with an assumed default judgement for the new case; and (ii) we sort past cases according to multi-attributes decision criteria, while they sort cases according to the ability to overturn judgements of other past cases. Cyras et al. (2016) also propose an argumentation-based CBR method, but they are not concerned with the rationality of recommendations, and their explanations are in terms of dispute trees only, rather than in natural language.

Furthermore, Roth and Verheij (2004) propose a case comparison model for legal reasoning in terms of dialectical support for conclusions. Our work is different from theirs in that we determine the similarity of past cases to a new case by decision-making criteria. Moreover, although past cases can also be ranked in terms of their dialectical support in Roth and Verheij's work, their method does not offer a natural language explanation for the ordering.

In generic CBR settings, Armengol and Plaza (2012) explain to users why some cases are retrieved by providing an explanation scheme for similarities in CBR. However, using their method, users do not know why other cases are not retrieved. Instead, our work supports both functionalities by means of explanations as dispute trees via the ABA mapping and explanations in natural language for pairwise decision comparison.

McSherry (2005) studies recommender system with a CBR based approach. Their work improves the efficiency and transparency of the recommendation process by focusing on explaining relevant questions to users. Our approach focuses on providing explanations for decisions recommended by our model in multi-attribute decision-making, on its own and in comparison with alternatives.

8.4. Template-based Natural language generation

From complex data streams, the natural language processing task of producing readable text in ordinary language (Winograd, 1972; Reiter and Dale, 2000) is Natural Language Generation (NLG). Many NLG practical applications have been developed, including writing weather forecasts (Reiter et al., 2005), summarising medical data (Portet et al., 2009) or generating hypotheses (Quinlan et al., 2012). Our proposal for generating explanations in natural language for comparing decisions can be seen as a further application of template-based NLG. To the best of our knowledge, ours is the first NLG work in the context of argumentation-based decision-making.

9. Conclusion

In this paper, we have identified a new decision criterion, *minimal redundancy*, and together with two existing notions of *dominance*, we can select decisions that achieve most goals but with fewest *redundant* attributes (*i.e.*, attributes not contributing to achieving goals). Also we have developed mappings onto a form of structured argumentation, ABA, for the two resulting decision criteria, serving as a basis for explaining decision selection. Furthermore, by defining a total ordering amongst all decisions, we have formally defined a ranking mechanism for decisions. To construct natural language explanations for why one decision is more preferred than, or as preferred as, another, we rely on an underlying argumentative counterpart to identify best decisions. To support transparent explanation in decision-making, our natural language explanation algorithm takes advantage of the argumentation-based approach. To illustrate and evaluate our model, we have developed an application study in the practice of law. Namely, we identify the most similar law cases, from a past case repository, to a new given case. Moreover, we present a natural language explanation for why a case is more similar than others. Such an explanation is important for human users, *e.g.* lawyers, to better understand and trust the cases recommended by our approach. A preliminary user evaluation indicates that our natural language explanations are useful for supporting human decision makers.

Our work leaves many open issues such as the following. (i) We just focused on decision problems with Boolean attributes and goals, but many interesting real problems cannot be modelled under these restrictions (Luo et al., 2003b). So it would be interesting to extend our model to accommodate fuzzy attributes and goals, and apply it into other fields (*e.g.* automated multi-attribute negotiation (Luo et al., 2003a; Zhan et al., 2018)). (ii) The legal CBR landscape has considered much more sophisticated forms of CBR than our illustrative legal example in this paper. For example, Horty (2011); Horty and Bench-Capon (2012) distinguish factors (attributes) for or against an outcome, and rules extracted from factors and reasons extracted from cases as well as preferences amongst them. In the future, it would be worth exploring whether or not our method for extracting argumentative explanations in natural language could be fruitfully deployed to support legal CBR in all its richness. (iii) Although our user evaluation for the legal case-based reasoning example is encouraging, it would be worth doing more systematic and sizeable evaluations, for instance, to test whether or not our model works well if other accusations are concerned (as in the CATO legal case-based reasoning system developed by Aleven (2003)), to compare with other explanation techniques (*e.g.* the argument representation techniques presented by Hoffmann (2005)), and to check whether or not it outperforms methods and systems supporting argumentation-based case-based reasoning (*e.g.* the system recently developed by Al-Abdulkarim et al. (2016)). (iv) Each of our evaluation experiments also included requests for free suggestions from the participants, which could be studied further. For example, some participants suggested to consider that some goals/attributes may be more important than others, and define decision criteria involving both attribute redundancy, as defined in this paper, and preference ranking over goals (*e.g.* along the lines of Fan et al. (2013), or considering preferences of various kinds as in Horty and Bench-Capon (2012)).

References

Al-Abdulkarim, L., Atkinson, K., and Bench-Capon, T. (2016). A methodology for designing systems to reason with legal cases using abstract dialectical frameworks. *Artificial Intelligence and Law*, 24:1–49.

Aleven, V. (2003). Using background knowledge in case-based legal reasoning: A computational model and an intelligent learning environment. *Artificial Intelligence*, 150(1-2):183–237.

Amgoud, L. and Prade, H. (2009). Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3-4):413–436.

Armengol, E. and Plaza, E. (2012). Symbolic explanation of similarities in case-based reasoning. *Computing and Informatics*, 25(2-3):153–171.

Ashley, K. D. (1991). *Modeling Legal Arguments: Reasoning with Cases and Hypotheticals*. MIT press.

Ashley, K. D., Lynch, C., Pinkwart, N., and Aleven, V. (2008). A process model of legal argument with hypotheticals. In *Proceedings of the 21st International Conference on Legal Knowledge and Information Systems*, pages 1–10.

Athakravi, D., Satoh, K., Broda, K., and Russo, A. (2014). Generating legal reasoning structure by answer set programming. In *Proceedings of the 8th International Workshop on Juris-informatics*, pages 24–37.

Atkinson, K., Bench-Capon, T., and McBurney, P. (2004). Justifying practical reasoning. In *Proceedings of the 4th Workshop on Computational Models of Natural Argument*, pages 87–90.

Banks, V. A., Plant, K. L., and Stanton, N. A. (2017). Driver error or designer error: Using the perceptual cycle model to explore the circumstances surrounding the fatal tesla crash on 7th May 2016. *Safety Science*. https://doi.org/10.1016/j.ssci.2017.12.023.

Bondarenko, A., Dung, P. M., Kowalski, R., and Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1–2):63–101.

Cao, M., Luo, X., Luo, X. R., and Dai, X. (2015). Automated negotiation for e-commerce decision making: A goal deliberated agent architecture for multi-strategy selection. *Decision Support Systems*, 73:1–14.

Ceballos, B., Lamata, M. T., and Pelta, D. A. (2017). Fuzzy multicriteria decision-making methods: A comparative analysis. *International Journal of Intelligent Systems*, 32(7):722–738.

Cohen, P. (1995). *Empirical Methods for Artificial Intelligence*. MIT Press.

Cyras, K., Satoh, K., and Toni, F. (2016). Abstract argumentation for case-based reasoning. In Baral, C., Delgrande, J. P., and Wolter, F., editors, *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning*, pages 549–552. AAAI Press.

Ding, Y., Liu, Y., Luan, H., and Sun, M. (2017). Visualizing and understanding neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1150–1159.

Domino, G. and Domino, M. L. (2006). *Psychological testing: An introduction*. Cambridge University Press.

Dong, Y., Su, H., Zhu, J., and Bao, F. (2017a). Towards interpretable deep neural networks by leveraging adversarial examples. *arXiv preprint arXiv:1708.05493*.

Dong, Y., Su, H., Zhu, J., and Zhang, B. (2017b). Improving interpretability of deep neural networks with semantic information. *arXiv preprint arXiv:1703.04096*.

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

Dung, P. M., Kowalski, R. A., and Toni, F. (2006). Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence*, 170(2):114–159.

Dung, P. M., Kowalski, R. A., and Toni, F. (2009). Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pages 199–218. Springer.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669.

Fan, X., Craven, R., Singer, R., Toni, F., and Williams, M. (2013). Assumption-based argumentation for decision-making with preferences: A medical case study. In *Proceedings of the 14th International Workshop on Computational Logic in Multi-Agent Systems*, pages 374–390.

Fan, X. and Toni, F. (2013). Decision making with assumption-based argumentation. In *Proceedings of the 2nd International Workshop on the Theory and Applications of Formal Argumentation*, pages 127–142.

Ferretti, E., Tamargo, L. H., Garca, A. J., Errecalde, M. L., and Simari, G. R. (2017). An approach to decision making based on dynamic argumentation systems. *Artificial Intelligence*, 242:107 – 131.

Fox, J., Glasspool, D., Patkar, V., Austin, M., Black, L., South, M., Robertson, D., and Vincent, C. (2010). Delivering clinical decision support services: There is nothing as practical as a good theory. *Journal of Biomedical Informatics*, 43(5):831–843.

Hahs-Vaughn, D. L. and Lomax, R. G. (2012). *An Introduction to Statistical Concepts: Third Edition*. Routledge.

Heras, S., Jordán, J., Botti, V., and Julian, V. (2013). Argue to agree: A case-based argumentation approach. *International Journal of Approximate Reasoning*, 54(1):82–108.

Hoffmann, M. H. G. (2005). Logical argument mapping: A method for overcoming cognitive problems of conflict management. *International Journal of Conflict Management*, 16(4):304–334.

Horty, J. F. (2011). Reasons and precedent. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, pages 41–50.

Horty, J. F. and Bench-Capon, T. J. M. (2012). A factor-based definition of precedential constraint. *Artificial Intelligence and Law*, 20(2):181–214.

Huval, B., Wang, T., Tandon, S., Kiske, J., Song, W., Pazhayampallil, J., Andriluka, M., Rajpurkar, P., Migimatsu, T., Cheng-Yue, R., Mujica, F., Coates, A., and Ng, A. Y. (2015). An empirical evaluation of deep learning on highway driving. *arXiv preprint arXiv:1504.01716*.

Kakas, A. and Moraitis, P. (2003). Argumentation based decision making for autonomous agents. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 883–890.

Kronman, A. T. (1990). Precedent and tradition. *The Yale Law Journal*, 99(5):1029–1068.

Labreuche, C. (2011). A general framework for explaining the results of a multi-attribute preference model. *Artificial Intelligence*, 175(7):1410–1448.

Lacave, C. and Díez, F. J. (2002). A review of explanation methods for Bayesian networks. *Knowledge Engineering Review*, 17:107–127.

Luo, X., Jennings, N. R., Shadbolt, N., f Leung, H., and f Lee, J. H. (2003a). A fuzzy constraint based model for bilateral, multi-issue negotiations in semi-competitive environments. *Artificial Intelligence*, 148(1-2):53–102.

Luo, X., Lee, J. H.-m., Leung, H.-f., and Jennings, N. R. (2003b). Prioritised fuzzy constraint satisfaction problems: Axioms, instantiation and validation. *Fuzzy Sets and Systems*, 136(2):151–188.

Luo, X., Zhong, Q., and Leung, H.-f. (2015). A spectrum of weighted compromise aggregation operators: A generalization of weighted uninorm operator. *International Journal of Intelligent Systems*, 30(12):1185–1226.

Ma, W., Luo, X., and Jiang, Y. (2017). Multicriteria decision making with cognitive limitations: A ds/ahp-based approach. *International Journal of Intelligent Systems*, 32(7):686–721.

Mathur, V. (2015). Google autonomous car experiences another crash. *Government Technology*, 17.

Matt, P. A., Toni, F., and Vaccari, J. (2009). Dominant decisions by argumentation agents. In *Proceedings of the 6th International Workshop on Argumentation in Multi-Agent Systems*, pages 42–59. Springer.

McSherry, D. (2005). Explanation in recommender systems. *Artificial Intelligence Review*, 24(2):179–197.

Menzies, T. (1998). Evaluation issues for problem solving methods. In *Proceedings of the 11th Workshop on Knowledge Acquisition, Modeling and Management*.

Modgil, S. and Prakken, H. (2014). The *ASPIC*$^+$ framework for structured argumentation: A tutorial. *Argument & Computation*, 5(1):31–62.

Müller, J. and Hunter, A. (2012). An argumentation-based approach for decision making. In *Proceedings of the 24th IEEE International Conference on Tools with Artificial Intelligence*, pages 564–571.

Nawwab, F. S., Bench-Capon, T. J. M., and Dunne, P. E. (2008). A methodology for action-selection using value-based argumentation. In *Proceedings of COMMA 2008*, volume 172, pages 264–275. IOS Press.

Niu, Z., Cheng, D., Yan, J., Zhang, J., Zhang, L., and Zha, H. (2017). A hybrid approach for risk assessment of loan guarantee network. *arXiv preprint arXiv:1702.04642*.

Park, J. H., Kwark, H. E., and Kwun, Y. C. (2017). Entropy and cross-entropy for generalized hesitant fuzzy information and their use in multiple attribute decision making. *International Journal of Intelligent Systems*, 32(3):266–290.

Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., and Sykes, C. (2009). Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7):789–816.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124.

Prakken, H., Wyner, A. Z., Bench-Capon, T. J. M., and Atkinson, K. (2015). A formalization of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 25(5):1141–1166.

Quinlan, P. R., Thompson, A., and Reed, C. (2012). An analysis and hypothesis generation platform for heterogeneous cancer databases. In *Proceedings of the 4th International Conference on Computational Models of Argument*, pages 59–70.

Reiter, E. and Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge University Press.

Reiter, E., Sripada, S., J.Hunter, and Davy, I. (2005). Choosing words in computer-generated weather forecasts. *Artificial Intelligence*, 167(1-2):137–169.

Roth, B. and Verheij, B. (2004). Cases and dialectical argument: An approach to case-based reasoning. In *On the Move to Meaningful Internet Systems 2004, Lecture Notes in Computer Science, Vol. 3292*, pages 634–651.

Song, Q., Chan, S. H., and Wright, A. M. (2017). The efficacy of a decision support system in enhancing risk assessment performance. *Decision Sciences*, 48(2):307–335.

Sul, H. K., Dennis, A. R., and Yuan, L. I. (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, 48(3):454–488.

Teach, R. L. and Shortliffe, E. H. (1984). An analysis of physician's attitudes. In *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, pages 635–652.

Tintarev, N. and Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems: Methodological issues and empirical studies on the impact of personalization. *User Modeling and User-Adapted Interaction*, 22(4-5):399–439.

Tintarev, N. and Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In Ricci, F., Rokach, L., and Shapira, B., editors, *Recommender Systems Handbook*, pages 353–382. Springer.

Toni, F. (2013). A generalised framework for dispute derivations in assumption-based argumentation. *Artificial Intelligence*, 195:1–43.

Toni, F. (2014). A tutorial on assumption-based argumentation. *Argument & Computation*, 5(1):89–117.

Treich, N. (2010). The value of a statistical life under ambiguity aversion. *Journal of Environmental Economics and Management*, 59(1):15–26.

Vickers, A. J. (2003). How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC medical research methodology*, 3(1):1.

Visser, W., Hindriks, K. V., and Jonker, C. M. (2012a). Argumentation-based qualitative preference modelling with incomplete and uncertain information. *Group Decision and Negotiation*, 21(1):99–127.

Visser, W., Hindriks, K. V., and Jonker, C. M. (2012b). An argumentation framework for qualitative multi-criteria preferences. In Modgil, S., Oren, N., and Toni, F., editors, *Theories and Applications of Formal Argumentation*, volume 7132 of *Lecture Notes in Computer Science*, pages 85–98. Springer.

Visser, W., Hindriks, K. V., and Jonker, C. M. (2013). Reasoning about interest-based preferences. In *Proceedings 5th International Conference on Agents and Artificial Intelligence*, pages 115–130. Springer.

Wachter, S., Mittelstadt, B. D., and Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *CoRR*, abs/1711.00399.

Williams, M., Liu, Z., A.Hunter, and MacBeth, F. (2015). An updated systematic review of lung chemo-radiotherapy using a new evidence aggregation method. *Lung Cancer*, 87:290–5.

Winkens, B., Schouten, H. J., van Breukelen, G. J., and Berger, M. P. (2006). Optimal number of repeated measures and group sizes in clinical trials with linearly divergent treatment effects. *Contemporary Clinical Trials*, 27(1):57–69.

Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, 3(1):1–191.

Yu, Q., Hou, F., Zhai, Y., and Du, Y. (2016). Some hesitant fuzzy einstein aggregation operators and their application to multiple attribute group decision making. *International Journal of Intelligent Systems*, 31(7):722–746.

Zhan, J., Luo, X., and Jiang, Y. (2018). An atanassov intuitionistic fuzzy constraint based method for offer evaluation and trade-off making in automated negotiation. *Knowledge-Based Systems*, 139:170–188.

Zhang, X. (2016). A novel approach based on similarity measure for pythagorean fuzzy multiple criteria group decision making. *International Journal of Intelligent Systems*, 31(6):593–611.

Zhong, Q., Fan, X., Toni, F., and Luo, X. (2014). Explaining best decisions via argumentation. In *Proceedings of the European Conference on Social Intelligence*, pages 224–237.

Zhou, X., Sedikides, C., Wildschut, T., and Gao, D. G. (2008). Counteracting loneliness on the restorative function of nostalgia. *Psychological Science*, 19(10):1023–1029.