

2021-01-09

Examining the Modelling Capabilities of Defeasible Argumentation and non-Monotonic Fuzzy Reasoning

Luca Longo

Technological University Dublin, luca.longo@tudublin.ie

Lucas Rizzo

Technological University Dublin, lucas.rizzo@tudublin.ie

Pierpaolo Dondio

Technological University Dublin, pierpaolo.dondio@tudublin.ie

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomart>



Part of the [Computer and Systems Architecture Commons](#)

Recommended Citation

Longo, L., Rizzo, L. & Dondio, P. (2021) Examining the Modelling Capabilities of Defeasible Argumentation and non-Monotonic Fuzzy Reasoning, *Knowledge-Based Systems*, 211 (2021) 106514. DOI: 10.1016/j.knosys.2020.106514

This Article is brought to you for free and open access by the School of Computer Sciences at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, gerard.connolly@tudublin.ie, vera.kilshaw@tudublin.ie.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/346115212>

Examining the modelling capabilities of defeasible argumentation and non-monotonic fuzzy reasoning

Article in Knowledge-Based Systems · October 2020

DOI: 10.1016/j.knosys.2020.106514

CITATION

1

READS

52

2 authors:



[Luca Longo](#)

Technological University Dublin - City Campus

82 PUBLICATIONS 894 CITATIONS

[SEE PROFILE](#)



[Lucas Rizzo](#)

Technological University Dublin - City Campus

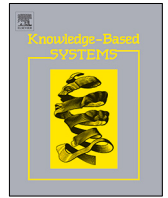
22 PUBLICATIONS 118 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Mental workload modeling [View project](#)



Examining the modelling capabilities of defeasible argumentation and non-monotonic fuzzy reasoning

Luca Longo^{*}, Lucas Rizzo, Pierpaolo Dondio

School of Computer Science, Technological University Dublin, Ireland

ARTICLE INFO

Article history:

Received 28 February 2020

Received in revised form 30 September 2020

Accepted 6 October 2020

Available online 31 October 2020

Keywords:

Defeasible reasoning

Non-monotonic reasoning

Fuzzy logic

Argumentation

Empirical research

Knowledge-representation

Mental workload

ABSTRACT

Knowledge-representation and reasoning methods have been extensively researched within Artificial Intelligence. Among these, argumentation has emerged as an ideal paradigm for inference under uncertainty with conflicting knowledge. Its value has been predominantly demonstrated via analyses of the topological structure of graphs of arguments and its formal properties. However, limited research exists on the examination and comparison of its inferential capacity in real-world modelling tasks and against other knowledge-representation and non-monotonic reasoning methods. This study is focused on a novel comparison between defeasible argumentation and non-monotonic fuzzy reasoning when applied to the representation of the ill-defined construct of human mental workload and its assessment. Different argument-based and non-monotonic fuzzy reasoning models have been designed considering knowledge-bases of incremental complexity containing uncertain and conflicting information provided by a human reasoner. Findings showed how their inferences have a moderate convergent and face validity when compared respectively to those of an existing baseline instrument for mental workload assessment, and to a perception of mental workload self-reported by human participants. This confirmed how these models also reasonably represent the construct under consideration. Furthermore, argument-based models had on average a lower mean squared error against the self-reported perception of mental workload when compared to fuzzy-reasoning models and the baseline instrument. The contribution of this research is to provide scholars, interested in formalisms on knowledge-representation and non-monotonic reasoning, with a novel approach for empirically comparing their inferential capacity.

© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Several approaches in the field of Artificial Intelligence (AI) have been proposed and investigated for modelling reasoning under uncertainty [1]. These include probability calculus [2] and its variations such as Possibility Theory [3] and Imprecise Probabilities [4], Dempster-Shafer Theory [5], Argumentation Theory [6] and Fuzzy reasoning [7]. Many scholars have used these approaches for modelling reasoning activities across many application domains. However, most of these are based upon a monotonic consequence relation: adding a formula to a theory never produces a reduction of its set of consequences. In other words, monotonicity supports the fact that learning a new piece of knowledge cannot reduce the set of what is known. Because of this limitation, they have been deemed not suitable for many knowledge representation problems that are non-monotonic in nature. Here, a conclusion or claim, derived from the application of some knowledge, can be retracted in the light of new

knowledge [8]. Defeasible Reasoning [9–11], a formalism for non-monotonic reasoning [12], is built upon reasons, conclusions and inferences that can be defeated. In practical terms, an implementation of defeasible reasoning is offered by argumentation [6,13,14]. It has gained interest because it provides the basis for computational models inspired by the way humans reason [15]. These models have extended classical reasoning approaches, based on deductive logic, that were proving increasingly inadequate for problems requiring non-monotonic reasoning and explanatory reasoning that are not available in standard non-monotonic logics [16]. In details, AT focuses on how pieces of evidence, beliefs and sometimes intuitions, seen as arguments, can be represented, supported or discarded in a defeasible reasoning process, and it investigates formal approaches to assess the validity of the conclusions inferred [17]. Despite the interest surrounding formal argumentation [18], limited work exists to compare the inferential capacity of models built with it against those of other non-monotonic reasoning approaches in knowledge-representation and reasoning under uncertainty [19–21].

This paper attempts at bridging this gap by proposing an empirical investigation whereby the inferential capacity of AT

^{*} Corresponding author.

E-mail address: luca.longo@tudublin.ie (L. Longo).

is compared against that of non-monotonic fuzzy reasoning. To achieve this goal, a knowledge representation problem has been chosen: the representation and assessment of the ill-defined construct of mental workload (MWL), via non-monotonic reasoning. Defining and modelling mental workload is an open problem within the disciplines of Human Factors, Psychology and Neuroscience. It is ill-defined because no clear and widely accepted definition exist. Mental workload can be intuitively described as the amount of necessary cognitive work invested in a certain task for a period of time. Nevertheless, this is an oversimplified definition and other factors such as stress, time pressure and mental effort can all influence mental workload whose levels can change over the execution of a task. Additionally, different scholars and reasoners from different disciplines might define mental workload according to their theoretical background, their knowledge and availability of different theories. This justifies the subjectivity and uncertainty surrounding the problem of mental workload definition and modelling. However, a reasonable assumption made in this research study is that mental workload is a complex construct built upon a network of pieces of evidence, beliefs, intuitions and understanding the interactions among these, is essential in defining and assessing it. These assumptions are also the key components of a defeasible concept: a concept built upon a set of reasons that can be defeated by additional reasons. Eventually, since mental workload is an ill-defined construct, no ground truth exists to validate different models of mental workload itself. However, over the past 50 years, designers and scholars have proposed a set of validation criteria that can be used to validate models of mental workload [22]. Among them, convergence and face validity have been chosen in this study. The former refers to the extent to which two measures of mental workload, that should be theoretically related, are in fact related, while the latter assess whether these measures appear effective in terms of their stated aims, that means measuring mental workload.

The specific research question under investigation is: *to what extent can defeasible argumentation allow the construction of models of mental workload with a higher convergent and face validity when compared to those constructed via non-monotonic fuzzy reasoning?*

The reminder of the paper is organised by presenting related work on argumentation theory and fuzzy reasoning in Section 2. A short description of the construct of mental workload follows, aimed at providing the readers with the relevant notions to understand the modelling problem under investigation. In Section 3, the design of a comparative experiment and the methodologies for the development of argument-based and fuzzy-reasoning models are detailed. Section 4 presents the results followed by a discussion. Section 5 concludes this research by highlighting its contribution to knowledge and proposing future avenues of development.

2. Related work

Reasoning and explanation under incomplete and uncertain knowledge have been investigated for several decades in AI. On one hand, classical propositional logic has demonstrated to be inadequate for dealing with real-world reasoning activities often involving inconsistent and conflicting information [23]. This is because of its monotonicity property that is based upon a notion of consequence relation by which adding a formula to a theory never produces a reduction of its set of consequences. In a nutshell, monotonicity is grounded on the fact that learning a new piece of knowledge cannot reduce the set of what is known.

In detail, if a conclusion p follows from a set of premises A (denoted as $A \vdash p$), in standard monotonic reasoning it also holds

that $A, B \vdash p$ if any additional set of premises B is added to A , the conclusion p is still valid.

However, this property does not allow the retraction of what is known, which is what often happens with human-like reasoning, which is generally non-monotonic. Reasoning is non-monotonic when a conclusion, supported by one or a number of premises built over some information, can be retracted in the light of new information [24,25].

In other words, in non-monotonic reasoning, even if the premises are still true, then the conclusion might not hold, whereas in monotonic reasoning, if the premises are true the conclusion necessarily follows. The property of non-monotonicity relies on the idea that a claim can be derived from premises partially specified, but in the case of an exception arising, this can be withdrawn [26,27]. Many non-monotonic reasoning formalisms have been investigated in AI over the last few decades [1,8,11,16,28–31] with many applications in many domains, with some examples in [10,32–40]. A type of reasoning that accounts for the property of non-monotonicity is defeasible reasoning, routinely used by humans. Here, the main feature is represented by *default knowledge* employable in a reasoning process even if the preconditions for its application are only partially known. Default knowledge is represented by using *defaults* that are specific inference rules. These are expressions of the form: $p(x) : j_1(x), \dots, j_n(x) \rightarrow c(x)$ where $p(x)$ is the prerequisite of the default, $j_1(x), \dots, j_n(x)$ are the justifications and $c(x)$ is the consequent. If $p(x)$ is known and if $j(x)$ is consistent with what is known, then $c(x)$ can be defeasibly deduced [23]. In other words, if it is believed that the prerequisite is true, and each of the n conditions (justifications) can be assumed since they are consistent with current beliefs, then this leads to believe the truth of the conclusion. The truth surrounding preconditions is not explicitly verified and they are assumed to hold defeasibly. In other words, this means they are true in the absence of explicit information to the contrary [8]. As soon as new information becomes available and the falsity of such preconditions can be deduced, then the conclusions, derived from the application of the default knowledge can be retracted [1,10,12]. Intuitively, this means that adding new premises may lead to removing (non-monotonicity), rather than adding new conclusions (monotonicity).

One of the practical implementations of defeasible reasoning is offered by argumentation theory (AT). Classical argumentation, from its roots within philosophy and psychology, deals with the study of how arguments or assertions are defined, discussed and solved in case of divergent opinions. In AI, argumentation refers to that body of literature that focuses on techniques for constructing computational models of arguments. Such models have become increasingly important for operationalising non-monotonic reasoning [1,15]. Example of application areas include dialogue and negotiation [15], knowledge representation [39] and decision-making in health-care [34], practical reasoning, decision support, dialogue and negotiation [15,17,41]. Argumentation deals with the interactions between possibly conflicting arguments, arising when different parties argue for and against some conclusions or when different pieces of evidence are available [42]. Arguments can be seen as 'tentative proofs for propositions' [43] in a logical language whose axioms represent premises in the domain under consideration. In general reasoning problem, usually the premises are not consistent because they may lead to incompatible conclusions. Argumentation systems are usually formed by several components and layers. A good summary of these components and their role can be found in [6] and summarised in the structure of Fig. 1. The first and second layers deals with the internal structure of arguments, its components such as premises, types of rules, their conclusions, their connection [44] as well as the definition of the conflicts among

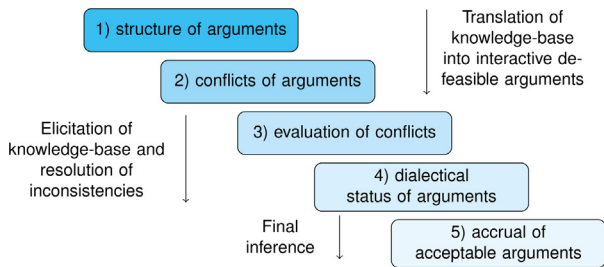


Fig. 1. Five layers upon which single-agent argument-based reasoning systems are generally built [6].

them. The subsequent layers deal with the dialogical structure of arguments and focus instead on conflict resolution. They typically regard arguments as monolithic entities, whose internal structure is abstracted away as far as the conflict resolution process is concerned. Roughly speaking, the first two layers concern the production and construction of arguments and their conflicts: they are aimed at formalising a knowledge-base. The subsequent dialogical layers concern with the resolution of potential inconsistencies and the extraction of a rational point of view that can be subsequently used, for instance, for supporting decision-making.

The structure described above is in some way aligned to the structure of [45] where 4 layers were proposed: logical, dialectical, procedural, heuristic. The logical layer deals with argument constructions, similar to layer one (Fig. 1). The second layer introduces the notion of attack, rebuttal, defeats, counter-arguments and deal with conflictual information, introducing the notion of reinstatement, respectively similar to layers 2 and 4 of Fig. 1. It also introduces the notion of argument comparison aimed at determining whether an attack is successful, as in layer 3 of the structure of [6]. The third layer regulates how a dispute among arguments can be conducted and allows the introduction of new information and support new arguments formation. This does not have correspondence to the structure proposed in this paper, as static knowledge can be only represented in the first 2 layers using the notions of argument and attack. The fourth layer provides rationale ways for resolving conflicts emerged in the previous layer and it is connected to the notion of rhetoric, influencing the preceding layers. These strategical ways or heuristics influence the previous layers by suggesting which premises to use and which arguments to construct as well as which arguments to attacks and claim to make or deny. This does not have correspondence with the structure proposed in this study (Fig. 1). The main differences (summarised in Fig. 2) is that the structure proposed in this paper (left), on the one hand, is linear and mainly for single-agent argument-based systems, sometimes referred to autonomous agent reasoning [46], and it is tailored to knowledge-representation problems, belief revision and decision-making under uncertainty. On the other hand, the structure proposed in [45] (right) is non-linear and mainly for multi-agent systems where different parties engage in a dispute and might form new arguments during the dialectical exchange of information. The former is more static and adds a final 5th layer devoted to the accrual of arguments, with the assumption that sometimes, in some particular domain, a final rational point has to be always extracted. The latter is more dynamic and is tailored to those situations in which rhetoric is required and reasoning is conducted by multi-parties such as in legal reasoning and dialogues facilitating multi-agent interaction [45,46].

The **first layer** of the structure used in this research (Fig. 1) focuses on constructing arguments. An example of an internal, monological structure has been proposed by Toulmin [47]. Toulmin's model is useful for highlighting the elements that might

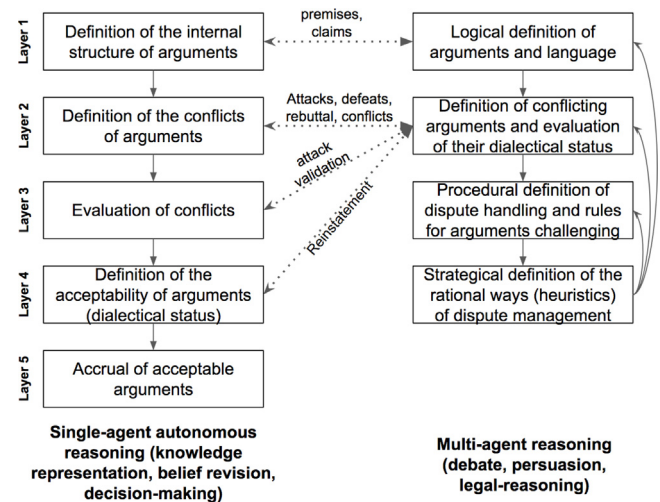


Fig. 2. Comparison of a single-agent argument-based reasoning scheme for knowledge representation [6] versus a multi-agent reasoning scheme for legal reasoning [45].

form a natural argument, and it provides a useful basis for knowledge representation. Another well-known approach has been proposed by Reed and Walton to model the notion of *arguments as product* [48,49]. It is based upon the notion of *argumentation schemes* and it is useful for identifying and evaluating a variety of argumentation structures in everyday discourse [44]. A recent attempt for reconstructing the internal structure of arguments in natural language has been proposed in [50]. This combines the linguistic representation framework of *Constructive Adpositional Grammars* (CxADGrams) [51] with the argument classification framework of the *Periodic Table of Arguments* (PTA) [52]. In detail, the **second layer** is built upon the notion of *conflict*, also referred to as attack or defeat, sometimes with slightly different meanings, key notions in argumentation. Several kinds of conflicts have been emerged in the literature but three core classes exist [53]:

- **undermining attack** - an argument can be attacked on one of its premises by another argument whose conclusion negates that premise. Example: 'soda consumption is low according to X so X has a low risk of cholesterol' can be undermined by 'the blood pressure emerging from a test is high so X has a high consumption of sodas'.
- **rebutting attack** - it occurs when an argument negates the conclusions of another argument. Example: 'X consumes minimal amount of soda so X has a low risk of heart attack' can be rebutted by 'X is an obese person, the strongest risk for heart attack is obesity, so the risk of heart attack is high'.
- **undercutting attack** - it occurs when an argument uses a defeasible inference rule and is attacked by arguing that there is a special case that does not allow the application of the rule itself [54]. Example: 'aspirin treatment minimises the risk of heart attack so X has a low risk of heart attack' can be undercut by 'paper Z demonstrated that aspirin failed several times in minimising the treat of heart attack so it is not always an effective method to reduce heart attacks'.

Conflict between arguments, although a key notion in argumentation, does not embody any approach for evaluating an attack, which is instead handled in the **third layer**. Generally speaking, an attack often has a form of a binary relation between two arguments. Some author distinguishes this relation in a weak form, attacking another argument and not weaker (defeat) or in a strong form, attacking another argument and stronger

(strict defeat) [53]. The construction of defeat relations are often influenced by the domain of application and are usually defeasible, that means attackable. For example, consider those domains where observations are important: defeat relations might be influenced by the reliability of tests and the expertise of the observers. To establish whether an attack can be considered a successful defeat, a trend in argumentation considers the notion of *strength* of arguments. Here, the assumption is the inequality of the strength of arguments that should be accounted for in the computation of sets of arguments and counterarguments [55]. Several researchers have adopted the notion of *preferentiality* among arguments [56]. In these researches, the assumption is that the information, necessary to evaluate the successfulness of an attack between two arguments is often pre-specified, and implemented as an ordering of values or a given partial preference. However, according to [56], also the information concerning preferentiality might be contradictory, and the preferences may change according to the context and to different observers who can assign different strengths, to different arguments, employing specific different criteria. Therefore, the notion of *meta-level argument* has emerged: a simple node in a graph where preferentiality is abstractly defined, and can be implemented by creating a new attack relation from a preference argument. Meta-level arguments have the goal of making a reasoning process more intuitive, as it allows no commitment regarding the definition of the preferences of arguments. As opposed to the preferentiality approach, another branch of argumentation is devoted to attach weights to attack relations rather than to arguments [55,57]. For example, the traditional binary relation of attack has been extended in [58] through the notion of *fuzzy relations* borrowed from Fuzzy Logic. This approach allows an expert to represent the degree of truth on an attack from an argument to another. Similarly in [59], probabilities can be assigned to arguments and attack relations. Here, probabilities refers to the likelihood of their existence and aim at capturing the inherent uncertainties in an argumentative reasoning activity.

Defeat relations, as defined in the third layer, focus on the relative strength of two conflicting arguments. However, they do not express yet which ones can be regarded as justifiable. The final state of each argument depends on the interaction with the others thus a notion of *dialectical status* is required in a **fourth layer**. Often, the outcome of an argumentation system is achieved by splitting the set of arguments in two, those that support a certain decision/action/claim and those that do not. Sometimes an additional class can contain those arguments that leave the dispute among arguments in an undecided status. In practical, real-world argumentation, multiple actions, decisions or claims can be considered thus the number of classes can increase. A seminal approach for assessing the dialectical status of arguments has been proposed by Dung [16], in line with other more practical and concrete works on argumentation [54,60]. Dung's abstract argumentation frameworks (AF) has been proven very useful for comparing different argumentative systems by translating them into his abstract format [60]. Here, given a set of abstract arguments whereby their internal structure is abstracted away, and a set of defeat relations, a decision has to be taken for determining which ones can ultimately be accepted. The core idea is that by uniquely looking at the defeaters of an argument to decide its dialectical status is not enough. It is also important to assess whether the defeaters are defeated themselves. Argumentation frameworks are strictly connected to the notions of *semantics*, specific criteria for the assessment of the dialectical status of arguments. The ideas is that, given an argumentation framework, a semantics specifies zero or more sets of acceptable arguments, called argument-based extensions, intuitively corresponding to different points of view. Examples include the popular grounded

and preferred semantics proposed by Dung [16] respectively a sceptical and credulous criteria for argument acceptability. Other semantics can be found in [14,61–68]. For practical purposes, as it often happens with human, a unique rational inference has to be made, a unique decision or has to be taken or a unique rationale claim is necessary. Thus another layer, a **fifth layer**, is sometimes added to the previous structure, as depicted in Fig. 1, aimed at accruing remaining arguments and producing a unique inference employable for practical purposes, such as informing decision making or explaining a rationale outcome. Here, accrual occurs at the level of the consequents of arguments and not at the level of an argument's claim. In other words, as stated in [69], in the former case, the accrual refers to a combination of arguments and their claims, leading to accruing reasons or arguments, while in the latter case, it refers to the decision or belief which is the subject of a reason (argument), namely referred to the claim (conclusion). Examples of approaches for accruing arguments at the consequent level include the average of the consequents of the arguments in an extension, in case of quantitatively assessable claims, or the consideration of claims of the top x% of arguments in case they were ranked in the previous layer with a ranking-based semantics. The reader is reminded to [69–72] for other approaches of accrual of arguments.

Despite the abundance of theoretical work, one of the main issues surrounding argumentation, its theories and approaches, is the lack of research devoted to the examination of its impact on the quality of the inferences produced by reasoning models built upon it. Similarly, situating and comparing argument-based systems among and against other non-monotonic reasoning approaches is negligible. However, as previously mentioned, many other approaches for implementing non-monotonic reasoning and enabling knowledge representation under partial, conflicting and uncertain information exist. One of these is represented by fuzzy reasoning, a type of reasoning grounded in the well-known fuzzy logic. Fuzzy reasoning, as originally proposed in the seminal article by Zadeh [73], is a generalisation of standard logic in which a concept can possess a degree of truth, and not only being completely true or completely false.

For this reason, fuzzy logic aims at modelling vague concepts with varying degrees of truth. In practical terms, these concepts are implemented via fuzzy sets and membership functions. The former are particular sets whose elements have degrees of membership and have similar notions to classical set theory such as inclusion, union and intersection. The latter are particular functions that models this membership, that means, they assign a grade of membership to an element, in the interval $[0, 1] \in \mathbb{R}$. These formal representations can support the construction of "IF-THEN" rules using natural language terms, as in "IF the *temperature* is *very high* THEN *significantly* increase the *speed* of the fan". *Temperature* and *speed* are the concepts, modelled as fuzzy sets, while *very high* and *significantly* are modelled by membership functions. Note that other membership functions might be designed as for instance, *very low* for *temperature*, or *moderately* for *speed*. In other words, *Temperature* is a fuzzy variable, while *very high* is a value that such variable can take, which is modelled by a membership function. A set of fuzzy rules can be employed for reasoning and along with the above notions, they have enabled the development of fuzzy control systems [74,75]. A traditional fuzzy control system is usually composed by a fuzzification module, an inference engine and a defuzzification module, sometimes referred to as defuzzifier [76, 77]. The fuzzification module takes the input variables, identified from the knowledge-base of an expert related to an underlying application domain. The universe of information is split into a number of fuzzy subsets, and each is assigned a linguistic label. Subsequently, a membership function for each fuzzy subset is

Table 1
Definition of Convergent, face validity and associated statistical tests for assessing a model of mental workload.

| Validity | Definition | Statistical test |
|------------|--|--|
| Convergent | It refers to the extent to which different mental workload measures that should be theoretically related, are in fact related | Correlation (model inferences vs MWL baseline instrument) |
| Face | It determines the extent to which a measure of mental workload logically appears effective in terms of its stated aims (measuring mental workload) | Correlation and Mean Squared Error (model inferences vs self-reported MWL score) |

created by an expert or a human reasoner, aimed at modelling the uncertainties associated to each input variable. These form the fuzzy linguistic variables that, jointly with the rule-base of the expert/reasoner, are translated into fuzzy rules that are aimed at assigning relationship between fuzzy input and output. Each rule can be built upon a number of fuzzy linguistic terms, joined with specific operators.

These might include the more traditional t-norm and t-conorms (intersection/disjunction), or others such as those used in intuitionistic fuzzy sets [78,79]. These are taken by the inference engine that actually performs the reasoning. Usually, these rules are all evaluated in parallel and are combined with another operator (usually the fuzzy OR operator) to obtain one fuzzy output distribution. This is then taken by the defuzzification module that applies a strategy (example the centre of mass, or mean of max approaches), to produce a single crisp output.

One of the limitations of this reasoning approach is that it does not incorporate any explicit notion of conflict among fuzzy rules. In other words, conflicting fuzzy rules might give rise to different outcomes when they are aggregated, therefore the behaviour of a fuzzy rule system admits that some degrees of truth may decrease in the presence of new information, which could be interpreted as a form of non-monotonicity. However, beside this interpretation, no explicit mechanism exists in standard fuzzy reasoning systems to handle conflicts among rules. Some scholars have proposed different extensions of standard fuzzy reasoning systems by incorporating a non-monotonic layer for dealing with conflicting information. Unfortunately, these are sporadic and not backed up by empirical research. For example, in [30], conflicting rules have their conclusions aggregated by an averaging function while in [31], a rule-base compression method is proposed for the reduction of non-monotonic rules. This method identifies all redundant rules after their fuzzification and it removes them while preserving the defuzzified output from the fuzzy system. Another approach has been proposed in [80, Chapter 8], and it is built upon Possibility Theory [3] which is employed within the inference engine to handle conflicting information. *Possibility* and *necessity* are measures of uncertainty and a special form of imprecise probability [81]. They can be linked to the notion of truth and its degree: *Possibility* indicates the extent to which data fail to refute its truth while *necessity* indicates the extent to which data supports its truth (both real values in $[0, 1] \in \mathbb{R}$). The possibility of a proposition can also be seen as the upper bound of the related necessity ($Pos \geq Nec$).

Not a lot of work exist in the literature towards comparing different non-monotonic reasoning approaches and their inferential capacity. The authors in [82] attempted at comparing Normal Default Theory against Defeasible Logic Programming. Similarly, [83] performed a comparison of first order predicate logic, fuzzy logic and non-monotonic logic as part of a knowledge-representation problem whereas [84] discussed and compared preferential and explanatory non-monotonic reasoning. However, the research studies above mainly focused on formal comparisons of different properties of different non-monotonic approaches for inferences and not on empirical evidence gathered from experimental studies involving humans. In contrast, a similar work to

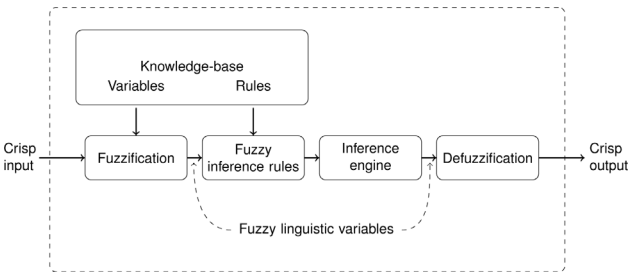


Fig. 3. General structure of a Mamdani fuzzy control system and its inferential process including fuzzification of natural language rules, their elicitation via an inference engine and their defuzzification towards a crisp output.

the research approach followed in this paper, has been performed by [85]. It focused on an empirical comparison between non-monotonic preferential logic and screened belief revision, a particular version of belief revision theories, involving human data. Most of the work on non-monotonic reasoning are strictly related to various knowledge-representation problems and this is the reason why the ill-defined concept of human mental workload has been selected for experimental purposes.

Mental workload (MWL) is a construct coming from psychology and mainly applied within ergonomics and education [86] with novel applications in medicine and human-computer interaction. It can be intuitively described as the amount of cognitive activity exerted to accomplish a specific task under a finite period of time [87]. However, this definition is very simplistic and many factors influence mental workload. Despite 50 years of research, unfortunately it is still an ill-defined construct, with many definitions, discipline-specific and with many application-dependent models [40]. Similarly, many formalisms exist to represent and assess mental workload [22,39,88,89]. Along with the various ad hoc models of mental workload, the many definitions and context-specific models, a number of criteria for evaluating these have been proposed [90], including reliability, validity, sensitivity and diagnosticity. Two of them have been selected in this study, as indicated in the research question of Section 1: convergent and face validity, both defined in Table 1. The following section is devoted to the description of the construction of various models of mental workload by employing different declarative knowledge-bases, provided by different human reasoners, formalised by using the two reasoning methods described so far, namely defeasible argumentation and non-monotonic fuzzy reasoning, and elicited with data acquired by human participants in an educational context.

3. Design and methodology

A primary research study has been designed and performed. This is aimed at comparing the inferential capacity of defeasible argumentation and non-monotonic fuzzy reasoning for the problem of mental workload modelling. In particular, a well-known self-reporting subjective mental workload assessment instrument has been chosen as baseline: the Nasa Task Load Index [88].

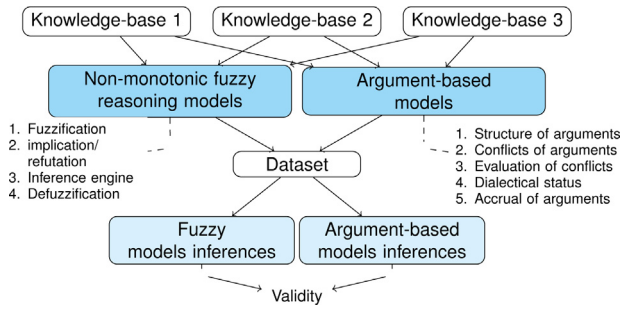


Fig. 4. Summary of the comparison of defeasible argumentation and non-monotonic fuzzy reasoning inferences.

This is a combination of six factors believed to influence mental workload: mental, temporal and physical demand, stress, effort and performance.

Each factor is quantified with a subjective judgement coupled with a weight w computed via a paired comparison procedure, that lead to 15 possible comparisons. The questionnaire designed for the quantification of each factor, and the pairwise comparison, can be found in [88] and in Tables 7 and 8 of Appendix. Eventually, the final mental workload score is the weighted average of the subjective rating associated to each attribute d_i :

$$\text{NASA-TLX} = \left(\sum_{i=1}^6 d_i \times w_i \right) \frac{1}{15}$$

As it is intuitive to grasp, this model is a weighted average, with no notion of non-monotonicity and no consideration of the relationships among the factors.

The information associated to these factors along with that related to the pairwise comparison (preferentiality) has been used to construct three knowledge-bases of different topology and complexity with a human reasoner. Note that no automatic procedure for inducing rules from data have been used, but only the expertise of human reasoners. In detail, defeasible argument-based model and a non-monotonic fuzzy reasoning-based models were constructed for each knowledge-base. Eventually, their inferential capacity is evaluated according to two criteria usually employed for the evaluation of models of mental workload: face and convergent validity (as detailed in Table 1). The overall research design is summarised in Fig. 4.

3.1. Non-monotonic fuzzy reasoning models

The non-monotonic fuzzy reasoning models are built according to the traditional Mamdani control system of Fig. 3 and extended by employing the notions of *possibility* and *necessity* to support its non-monotonicity.

3.1.1. Fuzzification module

Each encoded knowledge-base can be represented by rules of the form “IF - THEN”. The antecedent is a set of premises associated to a number of mental workload features, while the consequent is associated to a possible mental workload level (*underload*, *fitting lower load*, *fitting upper load*, *overload*). Examples of rules are:

- Rule 1: **IF** *low mental demand* **THEN** *underload*
- Rule 2: **IF** *low effort* **THEN** *fitting lower load*

Each consequent of a rule can be represented as a fuzzy term and described by a fuzzy membership function (FMF) (examples in Fig. 17). According to the domain expert’s knowledge, two options

Table 2

T-Norms and t-Conorms employed for two propositions a, b .

| Fuzzy operator | T-Norm | T-Conorm |
|----------------|----------------------|---------------------|
| Zadeh | $\min(a, b)$ | $\max(a, b)$ |
| Lukasiewicz | $\max(a + b - 1, 0)$ | $\min(a + b, 1)$ |
| Product | $a \cdot b$ | $a + b - a \cdot b$ |

were designed: the universe of mental workload is represented with an interval $[0, 100] \in \mathbb{R}$ with 4 membership functions. Fuzzy membership functions were also defined for all linguistic variables associated to the antecedents of rules such as *low* for mental demand and *low* for effort (examples in Fig. 16).

Membership functions for both the antecedents and consequent of a rule have been defined considering a human reasoner, with experience with the construct of human mental workload, and not automatically induced from data. Since each feature of the NASA-Task Load Index instrument are rated by human participants using a 20-point scale (Table 7 in Appendix), the input for these membership functions was scaled to the interval $[0, 100] \in \mathbb{R}$ in order to be the same as the membership functions defined for the consequents (workload levels). This was purely a practical decision in order to deal with similar scales. Additionally, the definition of

3.1.2. Inference engine

Once the knowledge-bases are fully translated into fuzzy inference rules, the next step is to evaluate their initial truth values. Each membership grade on the antecedent of these rules needs to be evaluated according to input data individually. If more than one antecedent is contained in a rule, fuzzy logics are necessary to aggregate them via the notions of union and intersection. Three known operators are selected for this: *Zadeh*, *Product* and *Lukasiewicz*. Table 2 lists the t-norms and t-conorms (AND, OR) respectively for each selected operator. Subsequently to the calculation of the initial truth values of the rules, it is necessary to solve any contradiction among them. For example:

- ‘IF *high effort*
THEN mental demand cannot be *low*’.

This rule indicates that if *effort* is *high* then any rule whose antecedent contains ‘*low mental demand*’ is refuted and its consequent should be evaluated again and the truth value associated to it updated. An example is:

- Exception 1: *high effort refutes Rule 1*

Exceptions represent the non-monotonic nature of the information in the knowledge-bases. Here, the underlying reasons who brought to life the exception is abstracted away that means, the undercutting, rebutting or undermining nature of the refutation is not formally modelled. A way of enabling non-monotonicity in fuzzy reasoning is the use of Possibility Theory, as implemented in [80]. Much of the literature on fuzzy mathematics is concerned with possibility, a measure of the extent to which the data fail to refute a conclusion. However, in the real world, we are primarily concerned with necessity, a measure of the extent to which the data support a conclusion. *Possibility* (Pos) and *necessity* (Nec) are values bounded in the interval $[0, 1] \in \mathbb{R}$. The reasoning process of establishing necessary conclusions is not the same as the process of establishing possible conclusions. For example, on one hand, when possible truth values are initialised in the lack of any data, a value of 1 is used. On the other hand, when necessary truth values are initialised in the absence of data, a value of 0 is used [80]. According to these considerations, in this study *necessity* is considered as the membership grade of a proposition while *possibility* is always set to 1 for all propositions, due to

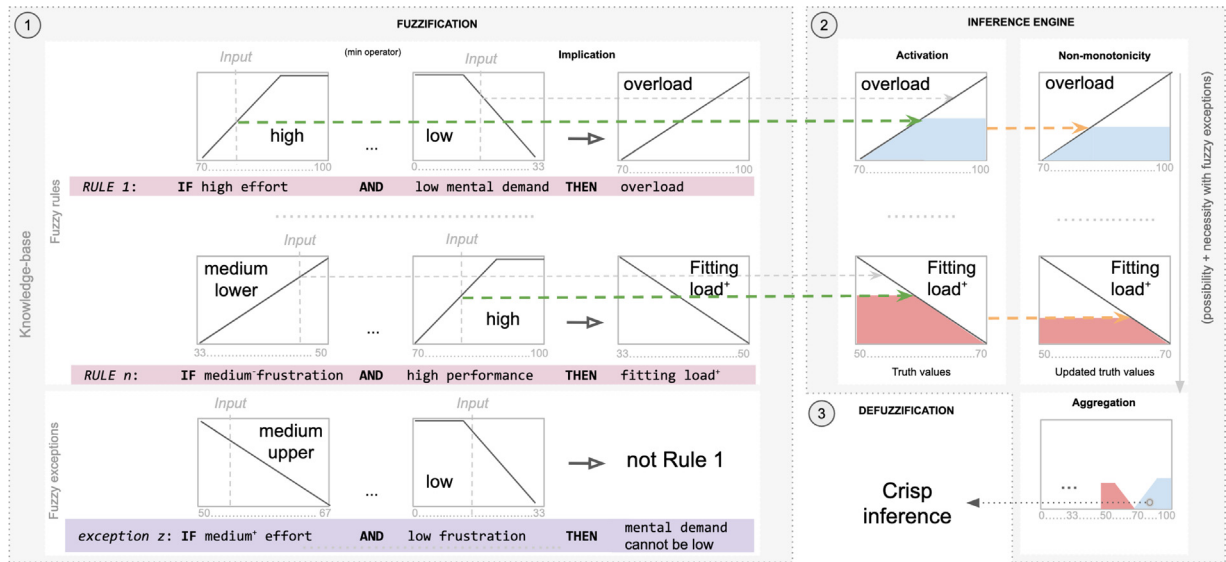


Fig. 5. A graphical summary of the inferential process followed by the non-monotonic fuzzy reasoning model. (1) A fuzzification layer in which a knowledge-base is operationalised via a human reasoner. Natural language rules are converted into computational fuzzy rules via membership functions for their antecedents and consequents. This layer includes also the formalisation of fuzzy exceptions, specific refutations of fuzzy rules. (2) An inference engine includes an activation layer aimed at evaluating fuzzy rules and fuzzy exceptions with real-world data, providing them with initial degrees of truth. Then, a non-monotonic layer deals with the application of fuzzy exceptions via the notions of possibility and necessity, and then it updates the degrees of truth of activated fuzzy rules. Subsequently, an aggregation layer merges the truncated membership functions of the consequents of the fuzzy rules. This aggregation is performed using the *min* operator, resulting in an outer envelope of the truncated membership functions (3) A defuzzification layer produces a final crisp inference from merged membership functions.

the lack of any data. Under this circumstance, the effect on the necessity of a proposition *A* by a set of propositions *Q* which refutes *A* is given by:

$$Nec(A) = \min(Nec(A), \neg Nec(Q_1), \dots, \neg Nec(Q_n)) \quad (1)$$

with $\neg Nec(Q) = 1 - Nec(Q)$. In Eq. (1), the addition of supporting evidence can only affect the necessity but not the possibility of a proposition. In this research study, there is no addition of supporting information but only attempts to refute information. Thus, the above equation is sufficient to model the contradictions in the knowledge-bases. An example is given below. Let us consider Rule 1 assuming it is refuted only by Exception 1, then the truth value of its consequent is given by:

- Truth value of Rule 1's consequent = $\min(Nec(low \text{ mental demand}), 1 - Nec(high \text{ effort}))$

$Nec(low \text{ mental demand})$ is the membership grade of the linguistic variable *low* of the attribute mental demand. For instance, if the mental demand = 1, then $Nec(low \text{ mental demand}) = 1$, according to the membership function *low* of Fig. 16. Similarly, if $Nec(high \text{ effort}) = 0$ then Exception 1 has no impact on Rule 1 while if $Nec(high \text{ effort}) = 1$ then the new truth value of Rule 1 is 0. Values between 1 and 0 indicates that Rule 1 is partially refuted. The truth value of Rule 1 is the truth value of the consequent *underload* in this particular case. It is important to highlight that the theory developed in [80] was for a multi-step forward-chaining reasoning system. This means that rules were activated in a chain, one by another, defining a precedence order of rules. However, in the current research study, the activation of rules is done in a single step, in the sense that data is imported and all rules are activated (or not) at once. Despite this constrain, it is still possible to define a precedence order of refutations also in this case. In detail, a tree structure in which the consequent of a refutation is the antecedent of the next refutation can be constructed. In this way, Eq. (1) can be applied from the root or roots to the leaves. This approach is sufficient for knowledge-bases that do not contain loops (cyclic exceptions). However, this is not

the case as the knowledge-bases employed here contain loops of rules or rebutting information. For instance, let us consider the following "IF-THEN" rules and their refutations:

- Rule 3: IF *low* temporal demand THEN *underload*
- Rule 4: IF *high* frustration THEN *overload*
- Exception 2: *low* temporal demand refutes Rule 4
- Exception 3: *high* frustration refutes Rule 3

In this case it is not clear whether exceptions 2 or 3 should be solved first. Given that there is no information within the knowledge bases to decide whether a mental workload feature (premise of a rule) or an exception is more important than another, then the proposal here is to activate them simultaneously. In detail, the original truth value of rules are initially computed before updating them due to exceptions (refutations). For instance, the truth values of rules 3, 4 are:

- $Tmp_1 = Nec(\text{Rule 3}) = Nec(low \text{ temporal demand})$
- $Tmp_2 = Nec(\text{Rule 4}) = Nec(high \text{ frustration})$
- Truth value of Rule 3 = $\min(Nec(low \text{ temporal demand}), 1 - Tmp_2)$
- Truth value of Rule 4 = $\min(Nec(high \text{ frustration}), 1 - Tmp_1)$

The above mechanism handles conflictual information by updating the degrees of truth of each rule. In turn, this degree of truth also represents the degree of truth of the rule's consequent. Once each consequent (mental workload level) has an associate degree of truth, it is necessary to aggregate them before proceeding to the defuzzification step. A disjunctive approach is employed for performing this aggregation. This approach groups the consequent levels inferred by each "IF-THEN" rule using the 'max' operator. In this case, at least one rule is satisfied, leading to a more flexible proposal. For instance, the truth value of *underload* in a context where only Rule 1 and Rule 3 infer *underload* is 'max(Truth value of Rule 1, Truth value of Rule 3)'. A conjunctive approach, employing the 'min' operator, is also possible. However, the set of rules would need to be jointly satisfied, representing a stricter proposal. Since exceptions are already defined in

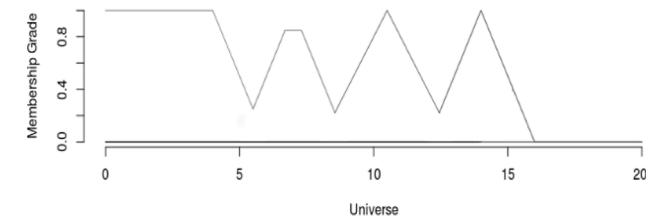


Fig. 6. An illustrative defuzzification process with an aggregation of 5 membership functions associated to 5 mental workload levels. The final truth values are: underload = 1, fitting lower = 0.83, fitting load = 1, fitting upper = 1 and overload = 0. The coordinates of the centroid are (6.89, 0.39) and the mean of max is 7.12.

Table 3

Summary of designed fuzzy reasoning models with different operators, defuzzification methods, membership functions for antecedents and consequents across knowledge-bases.

| Model | Operator | Defuzzification | Antecedents | Consequents | KB |
|-------|-------------|-----------------|-------------|-------------|----|
| F1 | Zadeh | Centroid | Fig. 16 | Fig. 17 | 1 |
| F2 | Zadeh | Mean of max | Fig. 16 | Fig. 17 | 1 |
| F3 | Product | Centroid | Fig. 16 | Fig. 17 | 1 |
| F4 | Product | Mean of max | Fig. 16 | Fig. 17 | 1 |
| F5 | Lukasiewicz | Centroid | Fig. 16 | Fig. 17 | 1 |
| F6 | Lukasiewicz | Mean of max | Fig. 16 | Fig. 17 | 1 |
| F7 | Zadeh | Centroid | Fig. 16 | Fig. 17 | 2 |
| F8 | Zadeh | Mean of max | Fig. 16 | Fig. 17 | 2 |
| F9 | Product | Centroid | Fig. 16 | Fig. 17 | 2 |
| F10 | Product | Mean of max | Fig. 16 | Fig. 17 | 2 |
| F11 | Lukasiewicz | Centroid | Fig. 16 | Fig. 17 | 2 |
| F12 | Lukasiewicz | Mean of max | Fig. 16 | Fig. 17 | 2 |
| F13 | Zadeh | Centroid | Fig. 20 | Fig. 21 | 3 |
| F14 | Zadeh | Mean of max | Fig. 20 | Fig. 21 | 3 |
| F15 | Product | Centroid | Fig. 20 | Fig. 21 | 3 |
| F16 | Product | Mean of max | Fig. 20 | Fig. 21 | 3 |
| F17 | Lukasiewicz | Centroid | Fig. 20 | Fig. 21 | 3 |
| F18 | Lukasiewicz | Mean of max | Fig. 20 | Fig. 21 | 3 |

the knowledge-base, this paper only investigates the disjunctive case. The inference engine stops when each consequent level has a truth value assigned.

3.1.3. Defuzzification module

The output of the inference engine can be graphically seen as the aggregation of the consequents (mental workload levels) of fuzzy inference rules, output of the inference engine, as exemplified in Fig. 6. Several methods can be used for calculating a single defuzzified scalar but two are selected here: *mean of max* and *centroid*. The first returns the average of all the consequents (here mental workload levels) of the fuzzy rules, with maximal membership grade. The second returns the coordinates (x, y) of the centre of gravity of the geometric shape formed by the aggregation of the membership functions associated to each consequent (MWL level). The defuzzified scalar is subsequently represented by the x coordinate of the centroid and we assume it represents a rational inference of mental workload. Following the non-monotonic fuzzy reasoning process described so far, a set of models is constructed with different fuzzy logic operators and defuzzification methods (as listed in Table 3). A graphical summary of the inferential process followed by these models is depicted in Fig. 5 (see Table 4).

3.2. Argument-based models

The definition of argument based-models follows the five-layer modelling approach depicted in Figs. 1 and 7. Arguments are rule-based and not of any other form such as natural language or logical propositions. They are of two types: *defeasible* and *defeater*

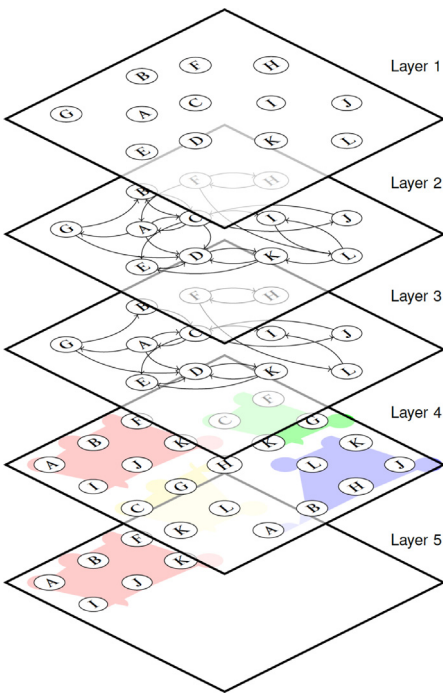


Fig. 7. The multi-layer argumentative structure for knowledge representation and defeasible reasoning. Layer 1: definition of the internal structure of arguments; Layer 2: definition of the conflicts among arguments using the notion of attack; Layer 3: activation of arguments and evaluation of their conflicts; Layer 4: definition of the acceptability status of arguments; Layer 5: accrual of acceptable arguments at their consequent level.

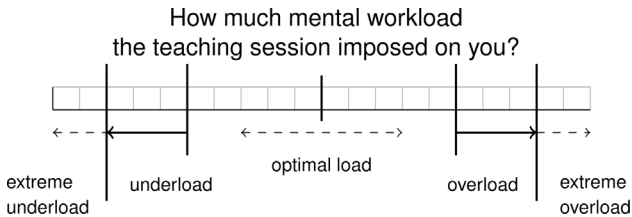


Fig. 8. A uni-dimensional self-reporting scale of perceived mental workload.

Table 4

Summary of the non-monotonic fuzzy reasoning models constructed via a human reasoner grouped by knowledge-base with references to their graphical representation and details of the membership functions adopted and their activation ranges both for the antecedents (MfA) and consequents (MfC) of fuzzy rules.

| KB | Graph | Fuzzy rules | MfA (ranges) | MfC (ranges) |
|----|---------|--------------|--------------------|--------------------|
| 1 | Fig. 18 | Tables 11–15 | Fig. 16 (Table 9) | Fig. 17 (Table 10) |
| 2 | Fig. 19 | Tables 16–17 | Fig. 16 (Table 9) | Fig. 17 (Table 10) |
| 3 | Fig. 22 | Tables 20–25 | Fig. 20 (Table 18) | Fig. 21 (Table 19) |

rules. The former rules contain premises that only create presumptions in favour of their conclusion contrarily to strict rules, which logically entail their conclusion [91]. The latter are rules that cannot be used to draw any conclusion and are essentially used to prevent some conclusion. These definitions of rules are in line with defeasible logics [92]. In detail, defeasible arguments are rules that can be defeated by contrary knowledge while defeater arguments are used to defeat some defeasible rule by producing knowledge to the contrary [92].

Layer 1 - Definition of the internal structure of arguments. The first step focuses on the construction of *forecast arguments* as it

Table 5

Designed argument-based models and their parameters across each layer.

| Model | Layer 1 | Layer 2 | Layer 3 | Layer 4 | Layer 5 |
|-------|-----------------|-----------------------|---------------------|-----------|-----------------------|
| | Args. | Conflicts | Conflict evaluation | Semantics | Accrual |
| M1 | Tables 11, 13 | Tables 12, 14, 15 | Binary | Grounded | Cardinality + Average |
| M2 | Tables 11, 13 | Tables 12, 14, 15 | Binary | Preferred | Cardinality + Average |
| M3 | Tables 11,13,16 | Tables 12,14, 15,17 | Binary | Grounded | Cardinality + Average |
| M4 | Tables 11,13,16 | Tables 12,14, 15,17 | Binary | Preferred | Cardinality + Average |
| M5 | Tables 20, 22 | Tables 21, 23, 24, 25 | Binary | Grounded | Cardinality + Average |
| M6 | Tables 20, 22 | Tables 21, 23, 24, 25 | Binary | Preferred | Cardinality + Average |

Table 6

Number of students across topics and delivery methods employed by lecturer (1: direct instruction with PDF; 2: direct instruction with video; 3: 2 + collaborative activity).

| Topic | Delivery method | | |
|-------------------|-----------------|----|----|
| | 1 | 2 | 3 |
| Science | 13 | 36 | 9 |
| Scientific method | 21 | 18 | 22 |
| Research planning | 22 | 22 | 16 |
| Literature review | 21 | 22 | 9 |

Table 7

The questions associated to the self-reporting NASA Task Load Index mental workload assessment instrument. Each answer is a 20-point scale.

| Dimension | Question |
|-----------------|---|
| Mental demand | How much mental AND perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy OR demanding, simple OR complex, exacting OR forgiving? |
| Physical demand | How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy OR demanding, slow OR brisk, slack OR strenuous, restful OR laborious? |
| Temporal demand | How much time pressure did you feel due to the rate OR pace at which the tasks OR task elements occurred? Was the pace slow AND leisurely OR rapid AND frantic? |
| Effort | How hard did you have to work (mentally AND physically) to accomplish your level of performance? |
| Performance | How successful do you think you were in accomplishing the goals, of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals? |
| Frustration | How insecure, discouraged, irritated, stressed AND annoyed versus secure, gratified, content, relaxed AND complacent did you feel during the task? |

follows.

Forecast argument: *premises* \rightarrow *conclusion*

A forecast argument is composed by a set of premises and a conclusion derivable by applying an inference rule \rightarrow . It is a defeasible rule that can be defeated by contrary knowledge [92]. To keep the terminology consistent with previous sections, we also refer to this as an “IF-THEN” rule. An example of a single-premise forecast argument is “ARG 1: IF *low* mental demand THEN *underload*”. An example of a multiple-premise forecast argument is: “HPF1: IF (*high* OR *medium upper*) effort AND (*medium lower* OR *medium OR low*) performance AND (*low* OR *medium lower*) frustration AND (*low* OR *medium lower*) mental demand THEN *underload*”. The linguistic terms of the attributes used in the premises of an argument and its conclusion are strictly bounded in well defined ranges and these are

Table 8

The pair-wise comparison procedure associated to the self-reporting NASA Task Load Index mental workload assessment instrument.

| Pair | Factor 1 | | Factor 2 |
|------|-----------------|--|-----------------|
| 1 | Temporal demand | <input type="checkbox"/> OR <input type="checkbox"/> | Frustration |
| 2 | Performance | <input type="checkbox"/> OR <input type="checkbox"/> | Mental demand |
| 3 | Mental demand | <input type="checkbox"/> OR <input type="checkbox"/> | Physical demand |
| 4 | Frustration | <input type="checkbox"/> OR <input type="checkbox"/> | Performance |
| 5 | Temporal demand | <input type="checkbox"/> OR <input type="checkbox"/> | Effort |
| 6 | Physical demand | <input type="checkbox"/> OR <input type="checkbox"/> | Frustration |
| 7 | Performance | <input type="checkbox"/> OR <input type="checkbox"/> | Temporal demand |
| 8 | Mental demand | <input type="checkbox"/> OR <input type="checkbox"/> | Effort |
| 9 | Physical demand | <input type="checkbox"/> OR <input type="checkbox"/> | Temporal demand |
| 10 | Frustration | <input type="checkbox"/> OR <input type="checkbox"/> | Effort |
| 11 | Physical demand | <input type="checkbox"/> OR <input type="checkbox"/> | Performance |
| 12 | Temporal demand | <input type="checkbox"/> OR <input type="checkbox"/> | Mental demand |
| 13 | Effort | <input type="checkbox"/> OR <input type="checkbox"/> | Physical demand |
| 14 | Frustration | <input type="checkbox"/> OR <input type="checkbox"/> | Mental demand |
| 15 | Performance | <input type="checkbox"/> OR <input type="checkbox"/> | Effort |

Table 9

Linguistic terms used to build the antecedents of rules and their levels for knowledge base 1 and 2.

| Linguistic term | Range |
|------------------------|-----------|
| Low | [0, <33] |
| Medium lower (Medium-) | [33, <50] |
| Medium upper (Medium+) | [50, <67] |
| High | [70, 100] |

established by the human reasoner. Formally:

$$\text{IF}(i_1 \in [l_1, u_1] \text{ AND } i_2 \in [l_2, u_2]) \text{ OR } (i_3 \in [l_3, u_3] \text{ AND } i_4 \in [l_4, u_4]) \text{ THEN } \textit{consequent} \in [l_c, u_c]$$

where $i_n \in \mathbb{R}$ is the activation value of premise n which is bounded in the numerical range $[l_n, u_n] \in \mathbb{R}$ (with $l_n < u_n$); $[l_c, u_c] \in \mathbb{R}$ (with $l_c < u_c$) is the numerical range of the consequent; AND, OR are logical operators.

The inference \rightarrow is formalised as a mapping from the ranges of the premises to that of the conclusion, as proposed in [93]. Formally, with the simple case which can be extended to an arbitrary number of premises, the value associated to the conclusion c is:

$$c = \frac{|u_c - l_c|}{R_{\max} - R_{\min}} \times (v - R_{\max}) + u_c \quad (2)$$

with

$$v = \min[\max(i_1, i_2), \max(i_3, i_4)]$$

$$R_{\max} = \min[\max(u_1, u_2), \max(u_3, u_4)]$$

$$R_{\min} = \min[\max(l_1, l_2), \max(l_3, l_4)]$$

Table 10

Linguistic terms used to build the consequents of rules and their levels for knowledge bases 1 and 2.

| Linguistic term | Range |
|---|-----------|
| Underload | [0, <33] |
| Fitting lower load (Fitting load ⁻) | [33, <50] |
| Fitting upper load (Fitting load ⁺) | [50, <67] |
| Overload | [70, 100] |

Table 11

Forecast rules designed for knowledge-base 1.

| Label | Rule |
|-------|---|
| MD1 | IF low mental demand THEN underload |
| MD2 | IF medium- mental demand THEN fitting load ⁻ |
| MD3 | IF medium+ mental demand THEN fitting load ⁺ |
| MD4 | IF high mental demand THEN overload |
| TD1 | IF low temporal demand THEN underload |
| TD2 | IF medium- temporal THEN fitting load ⁻ |
| TD3 | IF medium+ temporal THEN fitting load ⁺ |
| TD4 | IF high temporal THEN overload |
| EF1 | IF low effort THEN underload |
| EF2 | IF medium- effort THEN fitting load ⁻ |
| EF3 | IF medium+ effort THEN fitting load ⁺ |
| EF4 | IF high effort THEN overload |
| PF1 | IF low performance THEN overload |
| PF2 | IF medium- performance THEN fitting load ⁻ |
| PF3 | IF medium+ performance THEN fitting load ⁺ |
| PF4 | IF high performance THEN overload |
| FR1 | IF low frustration THEN underload |
| FR2 | IF high frustration THEN overload |

Table 12

Rebutting (symmetrical) attacks designed for knowledge-base 1.

| label | Attack |
|-------|---------------------------|
| R1 | MD1 mutually excludes FR2 |
| R2 | TD1 mutually excludes FR2 |
| R3 | FR1 mutually excludes MD4 |
| R4 | FR1 mutually excludes TD4 |
| R5 | FR1 mutually excludes EF4 |
| R6 | EF1 mutually excludes FR2 |
| R7 | EF1 mutually excludes MD4 |

Note that the value of a rule will always lie between the numerical range $[l_c, u_c]$ of its consequent but three options are possible depending on their values:

- if $l_c < u_c$ then Eq. (2) models a linear relationship (the higher the value of the premises the higher the value of the conclusion)
- if $l_c > u_c$ then Eq. (2) models an inverse linear relationship (the higher the value of the premises, the lower the value of the conclusion)
- if $l_c = u_c$ then Eq. (2) models a constant function whose inputs results in the same output (u_c). This is useful to model consequents with categorical levels.

Briefly, the above mapping provides a formula for rules that employ logical operators AND/OR, replacing them for *max* and *min* operators [94]. Example of forecast arguments can be found in Tables 11 and 20.

Layer 2 - Definition of the conflicts of arguments. In order to model inconsistencies, the notion of *mitigating argument* [42] is introduced. This is a defeater rule that cannot be used to draw any conclusion but only to prevent some other conclusion [92]. This type of argument is formed by a set of premises and an *undercutting inference* \Rightarrow to another argument, either forecast or mitigating:

Mitigating argument: *premises* \Rightarrow \neg *argument*

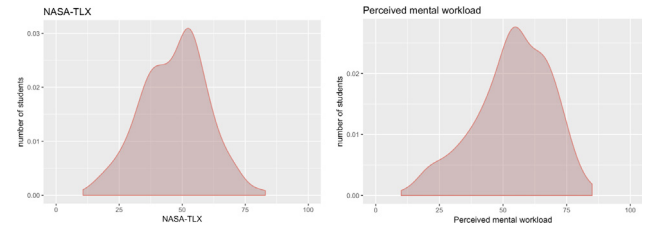


Fig. 9. Density plots of the distributions of the NASA-TLX scores (left) and perceived mental workload scores (right) for the entire population (231 participants).

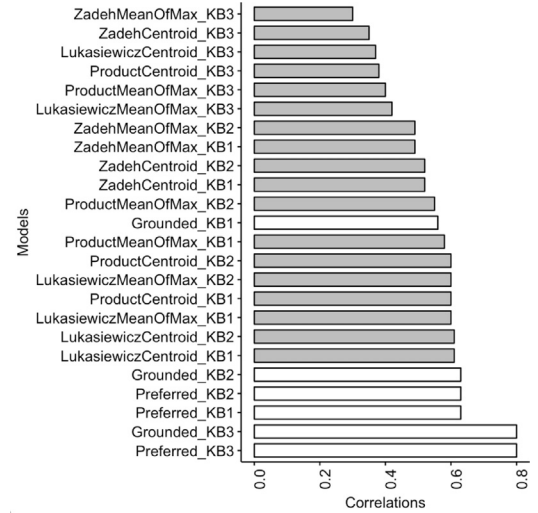


Fig. 10. Convergent validity of the non-monotonic fuzzy reasoning models (grey), the argument-based models (white) against the Nasa Task Load scores measured by the Pearson Correlation Coefficient, ordered lower to higher.

The inference \Rightarrow is an *undercutting attack* and a mitigating argument can be seen as an exception which has the effect of undermining the validity of another argument. Mitigating arguments are used to express the uncertainties of a reasoner concerning the validity of forecast arguments or other mitigating arguments. An example of a mitigating argument that can be constructed from Exception 1 (Section 3.1) is: “UA1: *high effort* \Rightarrow \neg ARG1”. Undercutting attacks do not allow partial refutation, so its target is always discarded. This means that no notion of strength of arguments nor strength of attack is considered.

Another type of attack relation is the *rebutting attack*, as mentioned in Section 2. This occurs when a *forecast argument* negates the conclusion of another forecast argument. A rebuttal occurs when the conclusions of two forecast arguments are believed to be mutually exclusive according to a reasoner. They attack each other generating rebuttals from contradictions, resulting in symmetric attacks:

Rebutting attack: *forecast argument* \Leftrightarrow *forecast argument*

An example of rebutting attack is: “R1: MD1 \Leftrightarrow FR2” (Table 12 and Fig. 18 in Appendix).

The third type of attack is the *undermining attack* which occurs when a forecast argument challenges the premises of another forecast argument. In other words, the conclusion of a forecast argument attacks the premises of another forecast argument. This allows a reasoner to model a situation in which the premises used to construct a particular forecast argument are no longer applicable given the inference of another forecast argument.

Undermining attack: *forecast argument* \Rightarrow \neg *forecast argument*

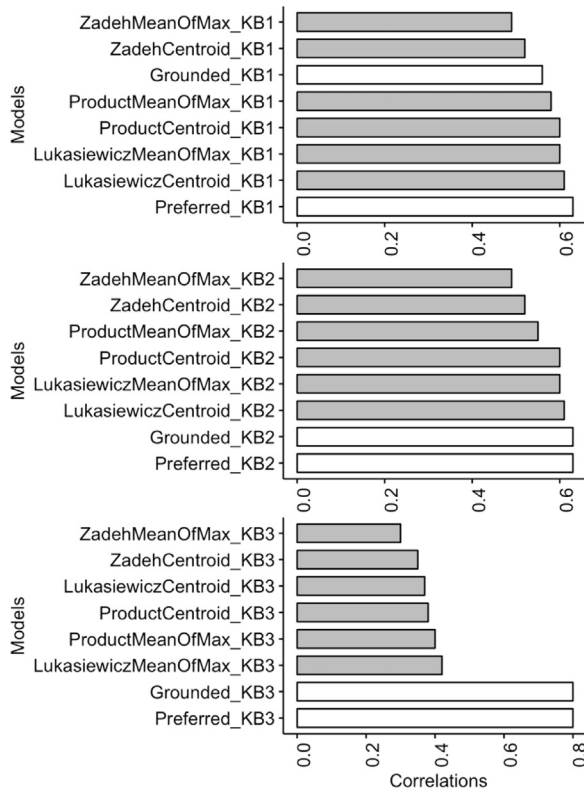


Fig. 11. Convergent validity of the non-monotonic fuzzy reasoning models (grey), the argument-based models (white) against the Nasa Task Load index, measured by the Pearson Correlation Coefficient, grouped by knowledge-base and ordered from lower to higher.

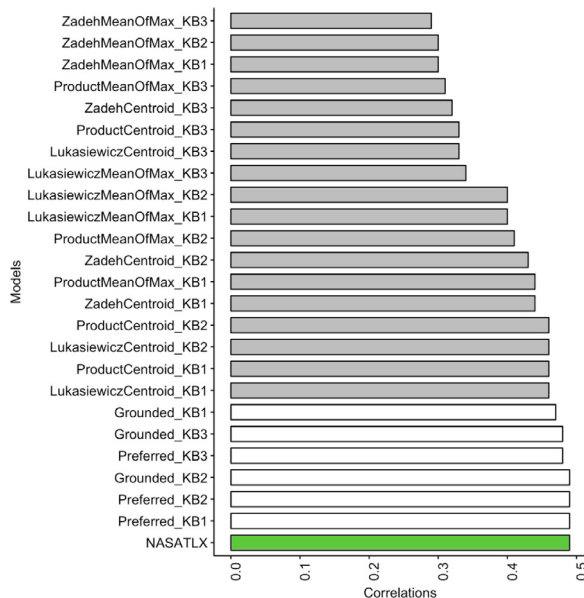


Fig. 12. Face validity of the non-monotonic fuzzy reasoning models (grey), the argument-based models (white) against the subjective perception of mental workload, ordered by lower to higher.

An example of an undermining attack is: “U1: R1 undercuts FR2”. More on the benefits of having undermining attacks in addition to rebutting and undercutting attacks is discussed in detail in [91]. The union of designed arguments and attacks can now be seen as an *argumentation framework* (AF). Formally, an argumentation

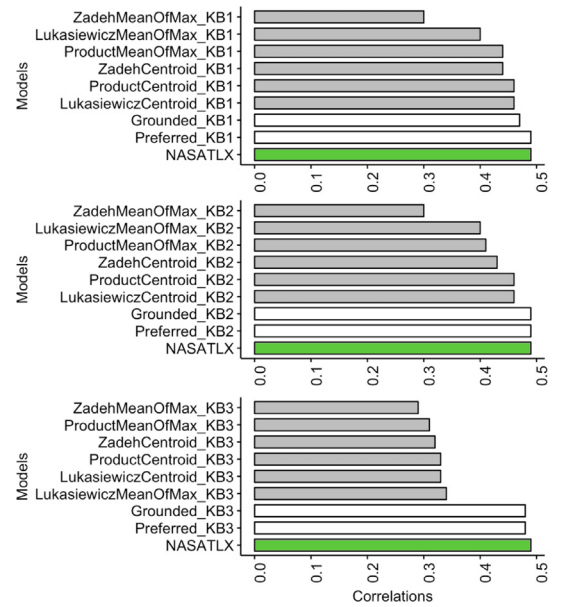


Fig. 13. Face validity between the non-monotonic fuzzy reasoning models (grey), the argument-based models (white) against the subjective perception of mental workload, group by knowledge-base and ordered by lower to higher.

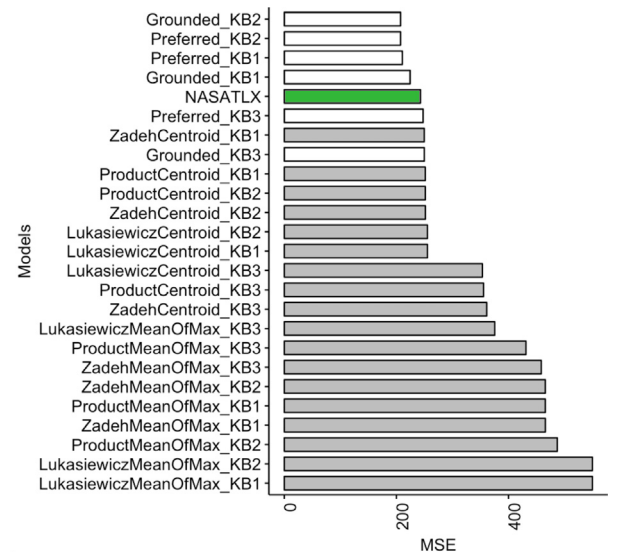


Fig. 14. Mean squared errors of the inferences of the non-monotonic fuzzy reasoning models (grey), the argument-based models (white) against the subjective perception of mental workload, ordered by lowered to higher.

framework is a pair:

$$AF = \langle \text{Args}, \text{Attacks} \rangle$$

where *Args* is a set of arguments (either forecast or mitigating) and $\text{Attacks} \subseteq \text{Args} \times \text{Args}$ is the list of attacks (either undercutting, rebutting or undermining). Examples of argumentation frameworks are those depicted in Figs. 18, 19 and 22 that also represent the three knowledge-bases used in this research.

Layer 3 - Evaluation of the conflicts of arguments At this stage an AF can be elicited with data. Forecast and mitigating arguments can be activated or discarded, based on whether their premises evaluate true or false. Attacks between activated arguments are considered valid, while all the others designed argument are discarded. Contrarily to fuzzy reasoning systems,

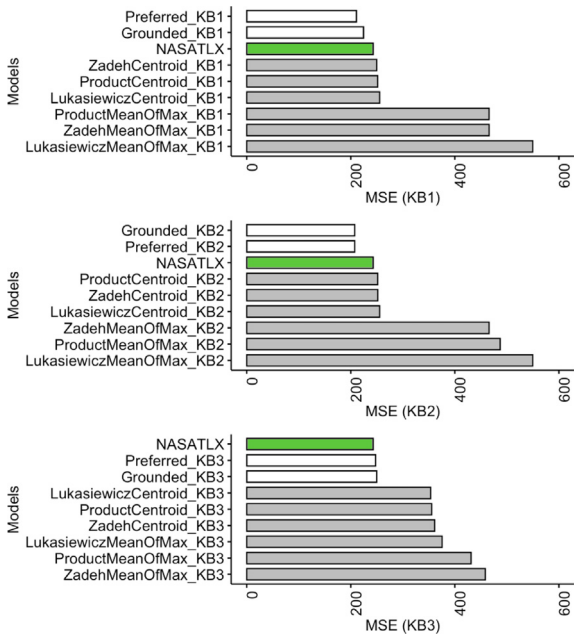


Fig. 15. Mean squared errors of the inferences of the non-monotonic fuzzy reasoning models (grey), the argument-based models (white) against the subjective perception of mental workload, grouped by knowledge-base and ordered by lowered to higher.

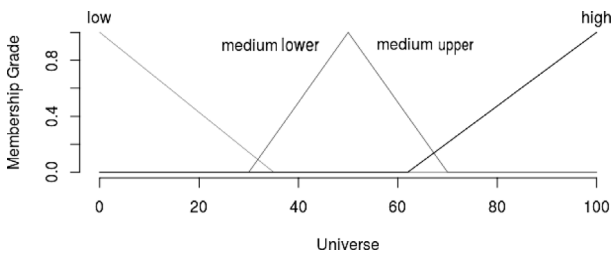


Fig. 16. Fuzzy membership functions for the designed linguistic terms for the antecedents of rules for knowledge bases 1 and 2.

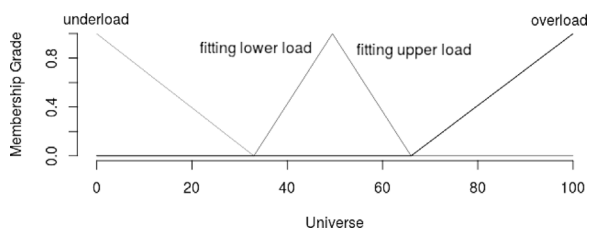


Fig. 17. Designed fuzzy membership functions for the designed linguistic terms for the consequents of rules for knowledge bases 1 and 2.

as designed in Section 3.1, there is no partial refutation, so a successful attack always refutes its target. In other words, no notion of strength of arguments or strength of attack is considered here. The set of activated forecast and mitigating arguments as well as the set of valid attacks form a *sub-argumentation framework* (*subAF*). A *subAF* is the same or a sub-set of the original *AF* and it contains that portion of a knowledge-base (rules and attacks) that has been activated with data and evaluates true:

$$subAF \subseteq AF$$

Layer 4 - Definition of the dialectical status of arguments
Given a *subAF*, acceptability semantics are applied in order to

accept or reject its arguments. Here, a *subAF* is equivalent to the notion of Abstract Argumentation Framework (AAF) proposed by Dung [16]) because the internal structure of arguments is not considered, that means it is abstracted away for the acceptance of arguments. In detail, arguments are considered in a dialogical structure, an interconnect graph of nodes (the arguments) and semantics are applied for evaluating which of them are defeated. An argument *A* is *defeated* by *B* if there is a valid attack from *A* to *B* [16]. Not only that, but it is also necessary to evaluate if the defeaters are defeated themselves. A set of non defeated arguments is called *extension* (conflict free set of arguments). Formally, An AAF is a pair $\langle Arg, attacks \rangle$ where: *Arg* is a finite set of arguments, abstractly evaluated, $attacks \subseteq Arg \times Arg$ is binary relation over the set *Arg*. Given sets of arguments $X, Y \subseteq Arg$, *X attacks Y* if and only if there exists $x \in X$ and $y \in Y$ such that $(x, y) \in attacks$. A set $X \subseteq Arg$ of arguments is:

- *admissible* iff *X* does not attack itself and *X* attacks every set of arguments *Y* such that *Y* attacks *X*;
- *complete* iff *X* is admissible and *X* contains all arguments it *defends*, where *X defends x* if and only if *X* attacks all attackers of *x*;
- *grounded* iff *X* is minimally complete (w.r.t. \subseteq);
- *preferred* iff *X* is maximally admissible (w.r.t. \subseteq).

Extensions are in turn used in the 5th layer of the diagram of Fig. 1, to produce a final inference.

Layer 5 - Accrual of acceptable arguments Eventually, in the last step of the reasoning process, a final inference sometimes has to be produced for practical purposes. For example, in case multiple extensions are computed, one extension might be preferred over the others. In this study, one assessment of mental workload has to be produced for practical purposes, thus a mechanism for accruing the arguments in multiple extensions, is needed. Intuitively, we argue that a larger conflict-free extension of arguments might be seen as carry more consistent information towards the description of the underlining phenomena being modelled (mental workload) than smaller extensions. Aware that other justifications are possible, we believe that the *cardinality* of an extension (number of conflict-free accepted arguments within it) can be used as a practical mechanism for the selection of the most suitable one. In the case more extensions exist with the same highest cardinality, then the proposal is to take into consideration of all them and their arguments, since a clear consistent inferential point of view (extension) has not emerged. After the selection of the most suitable extension/s, a single scalar is produced through the accrual of the consequents of its/their arguments. This is a simple average of the consequents of accepted forecast arguments within an extension (those that support a mental workload level and as computed in layer 1). Mitigating arguments already had their role by contributing to the resolution of conflictual information (layer 4) and they do not support any mental workload level and thus are not considered here. Formally, the overall inference of mental workload, brought forward by an extension (or multiple extensions), is computed by aggregating the scalars of its forecast arguments:

$$MWL = \frac{1}{m} \sum_{z=1}^m \frac{1}{n} \sum_{i=1}^n c_{arg^i} \in z$$

with *c* the value of the consequent of forecast arguments *arg_i* in extension *z*, as computed in Eq. (2), *n* the number of arguments in extension *z* and *m* is the number of extensions with highest cardinality, with $m, n \geq 1$).

Table 5 summarises the design of the argument-based models and their configuration following the 5-layer structures described above (Fig. 7).

3.3. Participant and procedures

A number of third-level classes have been delivered to students at the (anonymous institution). Each student has been provided with a study information sheet and a consent form, approved by the ethical committee of the university. After each class, students had to fill in the questionnaire associated to the NASA-TLX instrument (Tables 7–8, Appendix). Students were from 23 different countries (age 21–74, mean 30.9, std= 7.67). Four different topics of the module ‘research methods’ delivered in the school of computing, at the above institute, were delivered in different semesters during the academic terms 2015–2018. Three different delivery methods of instructional material were used by the lecturer:

- 1 traditional direct instruction, using slides projected to a white board. The lecturer introduced and explained the contents as in a traditional one-way class, without interaction with students;
- 2 multimedia video of the same content of 1 projected to a white board. In this case the delivery of the videos were significantly shorter than the deliveries of the instructional material with method 1;
- 3 constructivist collaborative activity performed after the delivery method 2. The lecturer randomly formed groups of students (max 4), provided them with the handouts of the class and a set of trigger questions on the instructional material previously delivered, to allow iteration of information and to enhance learning.

These delivery methods are assumed to impose on students different experience of mental workload. Overall, 231 students have participated in the study (summary statistics described in Table 6). After completing the questionnaire associated to the multi-dimensional NASA-TLX instrument for mental workload assessment, participants were required to answer another question providing an indication of their experienced mental workload using a uni-dimensional self-reporting scale (Fig. 8). The assumption behind self-reporting measures of mental workload is that only the person executing a task can provide a precise account of the own experienced workload. This question was also designed to support the computation of the face validity of each model, as defined in Table 1. Each argument-based and fuzzy reasoning model was elicited with the data associated to each student, and a mental workload inference was produced for each of them. The following section is aimed at presenting and interpreting these inferences.

4. Results

The answers of each item of the NASA-TLX questionnaire (Tables 7 and 8) provided by each student were used to elicit each of the designed non-monotonic fuzzy reasoning model (eighteen models in Table 3) and argument-based model (6 models in Table 5). Fig. 9 depicts the distribution of the NASA-TLX scores for the entire population (avg: 46.82; sd: 13.05) and the distribution of the uni-dimensional self-reported mental workload answers for the entire population of students (avg: 53.76; sd: 14.83). Both are normal and follow a Gaussian distribution. Fig. 23 depicts the scatterplots of the inferences produced by the non-monotonic fuzzy reasoning and argument-based models against the uni-dimensional self-reported mental workload scores as reported by students.

Convergent validity

Figs. 10 and 11 depict the Spearman correlation coefficients ($p < 0.05$) of the inferences produced by the designed models

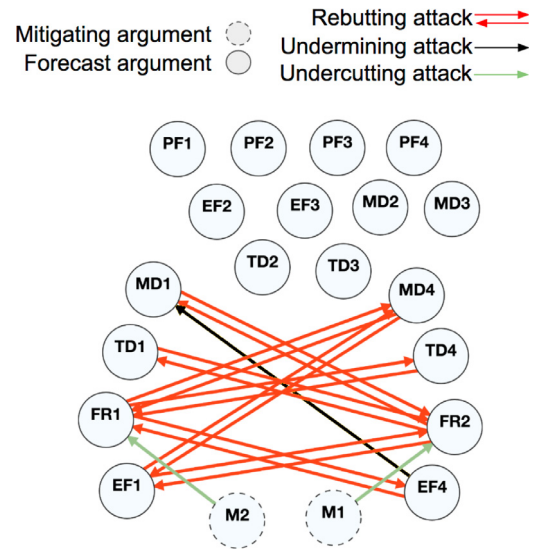


Fig. 18. Graphical representation of knowledge-base 1 as an argumentation framework.

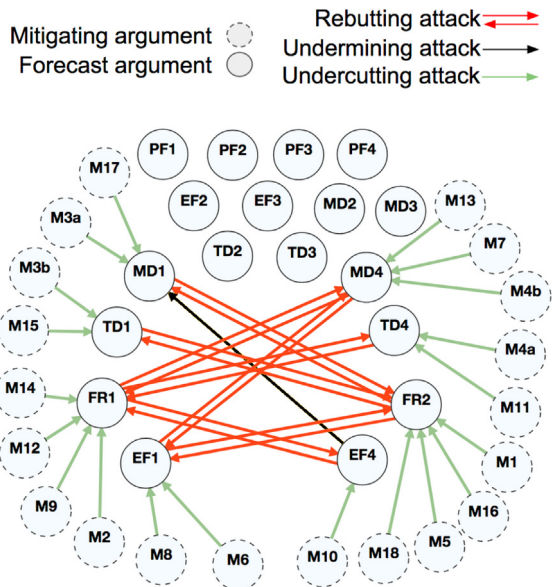


Fig. 19. Graphical representation of knowledge-base 2 as an argumentation framework.

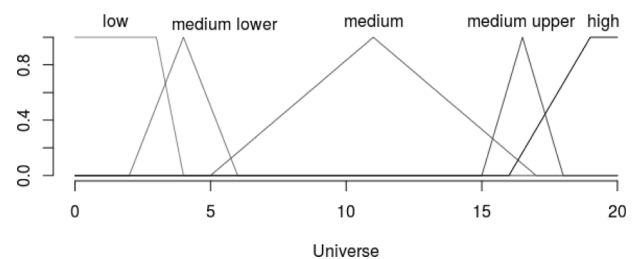


Fig. 20. Designed fuzzy membership functions for the designed linguistic terms for the antecedents of rules for knowledge base 3.

(mental workload scores) against the NASA-TLX scores, respectively ordered overall and grouped by knowledge base. The Spearman test was used because the assumptions behind the Pearson

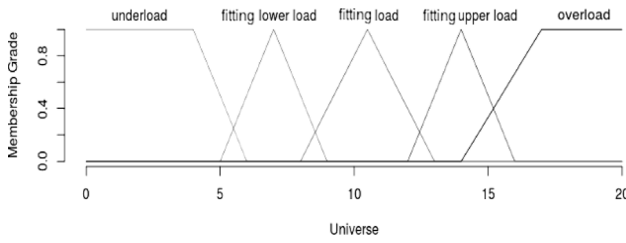


Fig. 21. Designed fuzzy membership functions for the designed linguistic terms for the consequents of rules for knowledge base 3.

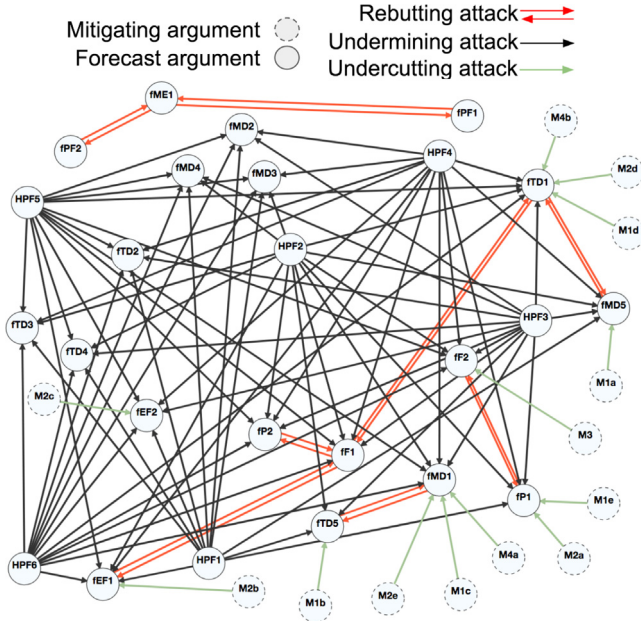


Fig. 22. Graphical representation of knowledge-base 3 as an argumentation framework.

Table 13

Mitigating arguments designed for knowledge-base 1.

| Label | Rule |
|-------|----------------------------------|
| M1 | IF high performance THEN not FR2 |
| M2 | IF low performance THEN not FR1 |

Table 14

Undercutting attacks originating from designed mitigating arguments for knowledge-base 1.

| Label | Attack |
|-------|------------------|
| U1 | R1 undercuts FR2 |
| U2 | R2 undercuts FR1 |

Table 15

Undermining attacks originating from designed forecast arguments for knowledge-base 1.

| Label | Attack |
|-------|--------------------|
| UM1 | EF4 undermines MD1 |

correlation test were not met. Moderate to high correlation coefficients were generally observed (ρ within 0.3–0.8). This indicates that the assumption of the theoretical relationship between the NASA-TLX measure, known in the literature to fairly model the ill-defined construct of mental workload, and the designed models, in fact exists. Additionally, a deeper analysis of these correlations reveals that argument-based models (white bars of Fig. 11)

Table 16

Mitigating arguments designed for knowledge-base 2. These are in addition to the mitigating arguments designed in knowledge-base 1. Note that in the case attribute A is preferred than attribute B, it means attribute A was chosen over B on the pair-wise comparison procedure of the NASA Task Load Index.

| label | Rule |
|-------|---|
| M3a | IF high effort AND low performance THEN not MD1 |
| M3b | IF high effort AND low performance THEN not TD1 |
| M4a | IF high performance AND low effort THEN not TD4 |
| M4b | IF high performance AND low effort THEN not MD4 |
| M5 | IF effort is preferred than frustration AND low effort THEN not FR2 |
| M6 | IF frustration is preferred than effort AND high frustration AND THEN not EF1 |
| M7 | IF effort is preferred than mental demand AND low effort THEN not MD4 |
| M8 | IF mental demand is preferred than effort AND high mental demand THEN not EF1 |
| M9 | IF effort is preferred than frustration AND high effort THEN not FR1 |
| M10 | IF frustration is preferred than effort AND low frustration THEN not EF4 |
| M11 | IF frustration is preferred than temporal demand AND low frustration THEN not TD4 |
| M12 | IF temporal demand is preferred than frustration AND high temporal demand THEN not FR1 |
| M13 | IF frustration is preferred than mental demand AND low frustration THEN not MD4 |
| M14 | IF mental demand is preferred than frustration AND high mental demand THEN not FR1 |
| M15 | IF frustration is preferred than temporal demand AND high frustration THEN not TD1 |
| M16 | IF temporal demand is preferred than frustration AND low temporal demand AND THEN not FR2 |
| M17 | IF frustration is preferred than mental demand AND high frustration THEN not MD1 |
| M18 | IF mental demand is preferred than frustration AND low mental demand THEN not FR2 |

generally lead to inferences closer to the baseline (NASA-TLX) when compared to those of the non-monotonic fuzzy reasoning models. In detail, this difference is more accentuated (Fig. 11 bottom) when considering the third knowledge-base (Fig. 22). A reasonable interpretation is that this knowledge-base contains a larger amount of conflictual information when compared to the other knowledge-bases. This suggests that argument-based models seems to handle non-monotonicity of knowledge and information on average significantly better than fuzzy reasoning models. In respect to model parameters, the inferences produced by fuzzy reasoning models do not seem to be significantly affected by the different fuzzy operators used (Zadeh, Product, Lukasiewicz) nor the defuzzification method (centroid or mean of max) when analysed by considering the different knowledge bases (Fig. 11). A similar situation exists for the argument-based models whose inferences seem not to be affected by the semantic

Table 17

Undercutting attacks originating from additional designed mitigating arguments for knowledge-base 2.

| Label | Attack |
|-------|-------------------------|
| U3a | M3a <i>THEN not</i> Md1 |
| U3b | M3b <i>THEN not</i> TD1 |
| U4a | M4a <i>THEN not</i> TD4 |
| U4b | M4b <i>THEN not</i> MD4 |
| U5 | M5 <i>THEN not</i> FR2 |
| U6 | M6 <i>THEN not</i> EF1 |
| U7 | M7 <i>THEN not</i> MD4 |
| U8 | M8 <i>THEN not</i> EF1 |
| U9 | M9 <i>THEN not</i> FR1 |
| U10 | M10 <i>THEN not</i> EF4 |
| U11 | M11 <i>THEN not</i> TD4 |
| U12 | M12 <i>THEN not</i> FR1 |
| U13 | M13 <i>THEN not</i> MD4 |
| U14 | M14 <i>THEN not</i> FR1 |
| U15 | M15 <i>THEN not</i> TD1 |
| U16 | M16 <i>THEN not</i> FR2 |
| U17 | M17 <i>THEN not</i> MD1 |
| U17 | M18 <i>THEN not</i> FR2 |

Table 18

Linguistic terms used to build the antecedents of rules and their levels for knowledge-base 3.

| Linguistic term | Range |
|------------------------|-----------|
| Low | [0, <3] |
| Medium lower (Medium−) | [3, <6] |
| Medium | [6, <16] |
| Medium upper (Medium+) | [16, <18] |
| High | [18, 20] |

Table 19

Linguistic terms used to build the consequents of rules and their levels for knowledge base 3.

| Linguistic term | Range |
|---|-----------|
| Underload | [0, <6] |
| Fitting lower load (Fitting load [−]) | [6, <9] |
| Fitting Load (Fitting load) | [9, <13] |
| Fitting upper load (Fitting load ⁺) | [13, <16] |
| Overload | [16, 120] |

used (grounded, preferred). However, this was expected because when preferred semantics produces only one extension of arguments, this coincides with the grounded extension produced by grounded semantics.

Face validity Figs. 12 and 13 depict the Spearman correlation coefficients ($p < 0.05$) of the inferences produced by the designed models against the perception of mental workload reported by students, respectively ranked from lower to higher and grouped by knowledge base. Similarly to the analysis of convergent validity, the Spearman test was used because the assumptions behind the Pearson correlation test were not met. Moderate correlation coefficients were generally observed (ρ within 0.3 – 0.5). This indicates that the fuzzy non-monotonic reasoning models and the argument-based models appear moderately effective in following the perception of mental workload subjectively reported by students. However, since the correlations produced by argument-based models (white bars of Fig. 12) are nearly the same as the correlations of the mental workload scores of the baseline model (the NASA-TLX), considered to be the gold standard measure for subjective mental workload, against the subjective perception of mental workload, then this suggests that their face validity is very good and in line with state-of-the-art models [22]. Additionally, results of Fig. 13 clearly demonstrate how the correlations of the inferences produced by the argument-based models against the perception of subjective mental workload, reported by students, when grouped by knowledge-bases is

Table 20

Forecast rules designed for knowledge-base 3.

| Label | Rule |
|-------|---|
| HPF1 | <i>IF</i> (high OR medium upper) effort AND (medium lower OR medium OR low) performance AND (low OR medium lower) frustration AND (low OR medium lower) mental demand <i>THEN</i> underload |
| HPF2 | <i>IF</i> (high OR medium upper) performance AND (medium OR medium lower) effort AND medium mental demand AND medium lower frustration <i>THEN</i> fitting load [−] |
| HPF3 | <i>IF</i> high performance AND (medium OR medium upper) mental demand AND medium frustration AND low effort <i>THEN</i> fitting load |
| HPF4 | <i>IF</i> (high OR medium upper) performance AND medium upper frustration AND medium upper mental demand AND (medium lower OR medium OR medium upper) effort <i>THEN</i> fitting load |
| HPF5 | <i>IF</i> high effort AND high frustration AND (medium lower OR medium) performance AND high mental demand <i>THEN</i> overload |
| HPF6 | <i>IF</i> high effort AND high frustration AND high mental demand AND low performance <i>THEN</i> overload |
| fMD1 | <i>IF</i> low mental demand <i>THEN</i> underload |
| fMD2 | <i>IF</i> medium lower mental demand <i>THEN</i> fitting load [−] |
| fMD3 | <i>IF</i> medium mental demand <i>THEN</i> fitting load |
| fMD4 | <i>IF</i> medium upper mental demand <i>THEN</i> fitting load ⁺ |
| fMD5 | <i>IF</i> high mental demand <i>THEN</i> overload |
| fTD1 | <i>IF</i> low temporal demand <i>THEN</i> underload |
| fTD2 | <i>IF</i> medium lower temporal demand <i>THEN</i> fitting load [−] |
| fTD3 | <i>IF</i> medium temporal demand <i>THEN</i> fitting load |
| fTD4 | <i>IF</i> medium upper temporal demand <i>THEN</i> fitting load ⁺ |
| fTD5 | <i>IF</i> high temporal demand <i>THEN</i> overload |
| fEF1 | <i>IF</i> low effort <i>THEN</i> fitting load |
| fP1 | <i>IF</i> high effort AND high mental demand <i>THEN</i> overload |
| fEF2 | <i>IF</i> (medium upper OR medium OR medium lower) effort <i>THEN</i> fitting load [−] |
| fP2 | <i>IF</i> high effort AND low mental demand <i>THEN</i> underload |
| fF1 | <i>IF</i> high frustration <i>THEN</i> underload |
| fF2 | <i>IF</i> low frustration <i>THEN</i> underload |
| fME1 | <i>IF</i> high mental demand AND high effort <i>THEN</i> overload |
| fPF1 | <i>IF</i> medium lower performance AND medium lower frustration <i>THEN</i> fitting load [−] |
| fPF2 | <i>IF</i> medium performance AND medium frustration <i>THEN</i> fitting load |

always higher than those correlations produced by the fuzzy non-monotonic reasoning models and this is more accentuated again with knowledge-base 3.

A deeper analysis has been performed by investigating the error (distance) between the inferences produced by designed models and the subjective perception of mental workload reported by students (as per scale in Fig. 8). Figs. 14 and 15 depict the mean squared errors (MSEs) of the inferences of each designed model, respectively overall and grouped by knowledge base. As it can be observed, argument-based models not only have

Table 21
Rebutting (symmetrical) attacks designed for knowledge-base 3.

| label | Attack |
|-------|------------------------------------|
| R1 | fTD5 <i>mutually excludes</i> fMD1 |
| R2 | fTD1 <i>mutually excludes</i> fMD5 |
| R3 | fP1 <i>mutually excludes</i> ff2 |
| R4 | fP2 <i>mutually excludes</i> ff1 |
| R5 | ffF1 <i>mutually excludes</i> ff1 |
| R6 | ff1 <i>mutually excludes</i> fTD1 |
| R7 | fPF2 <i>mutually excludes</i> fME1 |
| R8 | fPF1 <i>mutually excludes</i> fME1 |

Table 22
Mitigating rules designed for knowledge-base 3.

| Label | Rule |
|-------|---------------------------------------|
| M1a | IF high performance THEN not fMD5 |
| M1b | IF high performance THEN not fTD5 |
| M1c | IF high performance THEN not fMD1 |
| M1d | IF high performance THEN not fTD1 |
| M1e | IF high performance THEN not fP1 |
| M2a | IF low performance THEN not fP1 |
| M2b | IF low performance THEN not ffF1 |
| M2c | IF low performance THEN not ffF2 |
| M2d | IF low performance THEN not fTD1 |
| M2e | IF low performance THEN not fMD1 |
| M3 | IF high effort THEN not ff2 |
| M4a | IF high physical demand THEN not fMD1 |
| M4b | IF high physical demand THEN not fTD1 |

Table 23
Undercutting attacks originating from designed mitigating arguments for knowledge-base 3.

| Label | Attack |
|-------|---------------------------|
| U1 | M1a <i>undercuts</i> fMD5 |
| U2 | M1b <i>undercuts</i> fTD5 |
| U3 | M1c <i>undercuts</i> fMD1 |
| U4 | M1d <i>undercuts</i> fTD1 |
| U5 | M1e <i>undercuts</i> fP1 |
| U6 | M2a <i>undercuts</i> fP1 |
| U7 | M2b <i>undercuts</i> ffF1 |
| U8 | M2c <i>undercuts</i> ffF2 |
| U9 | Md2 <i>undercuts</i> fTD1 |
| U10 | M2e <i>undercuts</i> fMD1 |
| U11 | M3 <i>undercuts</i> ff2 |
| U12 | M4a <i>undercuts</i> fMD1 |
| U13 | M4b <i>undercuts</i> fTD1 |

Table 24
Mitigating arguments designed knowledge-base 3 - Part a.

| Label | Attack |
|-------|-----------------------------|
| UM1 | HPF5 <i>undermines</i> fMD3 |
| UM2 | HPF5 <i>undermines</i> fTD4 |
| UM3 | HPF5 <i>undermines</i> fTD2 |
| UM4 | HPF5 <i>undermines</i> fTD3 |
| UM5 | HPF5 <i>undermines</i> ffF1 |
| UM6 | HPF5 <i>undermines</i> ffF2 |
| UM7 | HPF5 <i>undermines</i> ff1 |
| UM8 | HPF5 <i>undermines</i> ff2 |
| UM9 | HPF5 <i>undermines</i> fTD1 |
| UM10 | HPF5 <i>undermines</i> fMD1 |
| UM11 | HPF5 <i>undermines</i> ff2 |
| UM12 | HPF3 <i>undermines</i> fMD4 |
| UM13 | HPF3 <i>undermines</i> fMD2 |
| UM14 | HPF3 <i>undermines</i> fTD4 |
| UM15 | HPF3 <i>undermines</i> fTD2 |
| UM16 | HPF3 <i>undermines</i> ffF2 |
| UM17 | HPF3 <i>undermines</i> ff1 |
| UM18 | HPF3 <i>undermines</i> ff2 |
| UM19 | HPF3 <i>undermines</i> fTD1 |
| UM20 | HPF3 <i>undermines</i> fMD5 |
| UM21 | HPF3 <i>undermines</i> fMD1 |
| UM22 | HPF3 <i>undermines</i> fTD5 |
| UM23 | HPF3 <i>undermines</i> fP1 |
| UM24 | HPF3 <i>undermines</i> ff2 |
| UM25 | HPF2 <i>undermines</i> fMD4 |
| UM26 | HPF2 <i>undermines</i> fMD3 |
| UM27 | HPF2 <i>undermines</i> fTD4 |
| UM28 | HPF2 <i>undermines</i> fTD3 |
| UM29 | HPF2 <i>undermines</i> ffF1 |
| UM30 | HPF2 <i>undermines</i> ff1 |
| UM31 | HPF2 <i>undermines</i> ff2 |
| UM32 | HPF2 <i>undermines</i> fTD1 |
| UM33 | HPF2 <i>undermines</i> fMD5 |
| UM34 | HPF2 <i>undermines</i> fMD1 |
| UM35 | HPF2 <i>undermines</i> fTD5 |
| UM36 | HPF2 <i>undermines</i> fP1 |
| UM37 | HPF2 <i>undermines</i> ff2 |
| UM38 | HPF6 <i>undermines</i> fMD4 |
| UM39 | HPF6 <i>undermines</i> fMD2 |
| UM40 | HPF6 <i>undermines</i> fMD3 |
| UM41 | HPF6 <i>undermines</i> fTD4 |
| UM42 | HPF6 <i>undermines</i> fTD2 |
| UM43 | HPF6 <i>undermines</i> fTD3 |
| UM44 | HPF6 <i>undermines</i> ffF1 |
| UM45 | HPF6 <i>undermines</i> ffF2 |
| UM46 | HPF6 <i>undermines</i> ff1 |
| UM47 | HPF6 <i>undermines</i> ff2 |
| UM48 | HPF6 <i>undermines</i> fTD1 |
| UM49 | HPF6 <i>undermines</i> fMD1 |
| UM50 | HPF6 <i>undermines</i> ff2 |
| UM51 | HPF1 <i>undermines</i> fMD4 |
| UM52 | HPF1 <i>undermines</i> fMD2 |
| UM53 | HPF1 <i>undermines</i> fMD3 |
| UM54 | HPF1 <i>undermines</i> fTD4 |

almost always a lower MSE error than the fuzzy non-monotonic reasoning models, but this time they were also often better than the baseline instrument (NASA-TLX). Additionally, among the non-monotonic fuzzy reasoning models, the difference in the inferences of those employing the *centroid* as defuzzification method appear to be better than the ones employing the *mean of max*. This is also confirmed by the scatterplots of Fig. 23 that show how the fuzzy models employing the *mean of max* defuzzification method produced clustered inferences around different points with low spread (last three lines of scatterplots), when compared to those inferences produced by models applying the *centroid method* (third and fourth lines of scatterplots). In relation to the inferences by argument-based models, Fig. 15 confirms again that when knowledge-bases are characterised by higher uncertainty due to an higher number of interacting and conflicting pieces of knowledge (knowledge-base 3, bottom of figure), then they behave significantly better than fuzzy reasoning models. This can be also observed in the scatterplots of Fig. 23. In fact, when it comes to knowledge-base 3 (third column of scatterplots), then the inferences of argument-based are more spread than those produced by non-monotonic fuzzy reasoning models.

4.1. Discussion

An analysis of the convergent validity of designed models showed that their inferences can be considered valid. They are positively and moderately correlated to the well known subjective model of mental workload, namely the NASA Task Load Index, thus likely modelling mental workload too. A negative or null correlation would have implied the invalidity of designed models since they would have probably modelled another construct and not mental workload. An analysis of the convergent validity indicated a better inferential capacity of the argument-based models over the non-monotonic fuzzy reasoning models. This can be confidently claimed because their inferences are always superior of the non-monotonic fuzzy reasoning models, regardless of their respective internal configurations and parameters, and regardless of the knowledge-bases employed. Argument-based models consistently showed a significantly lower mean

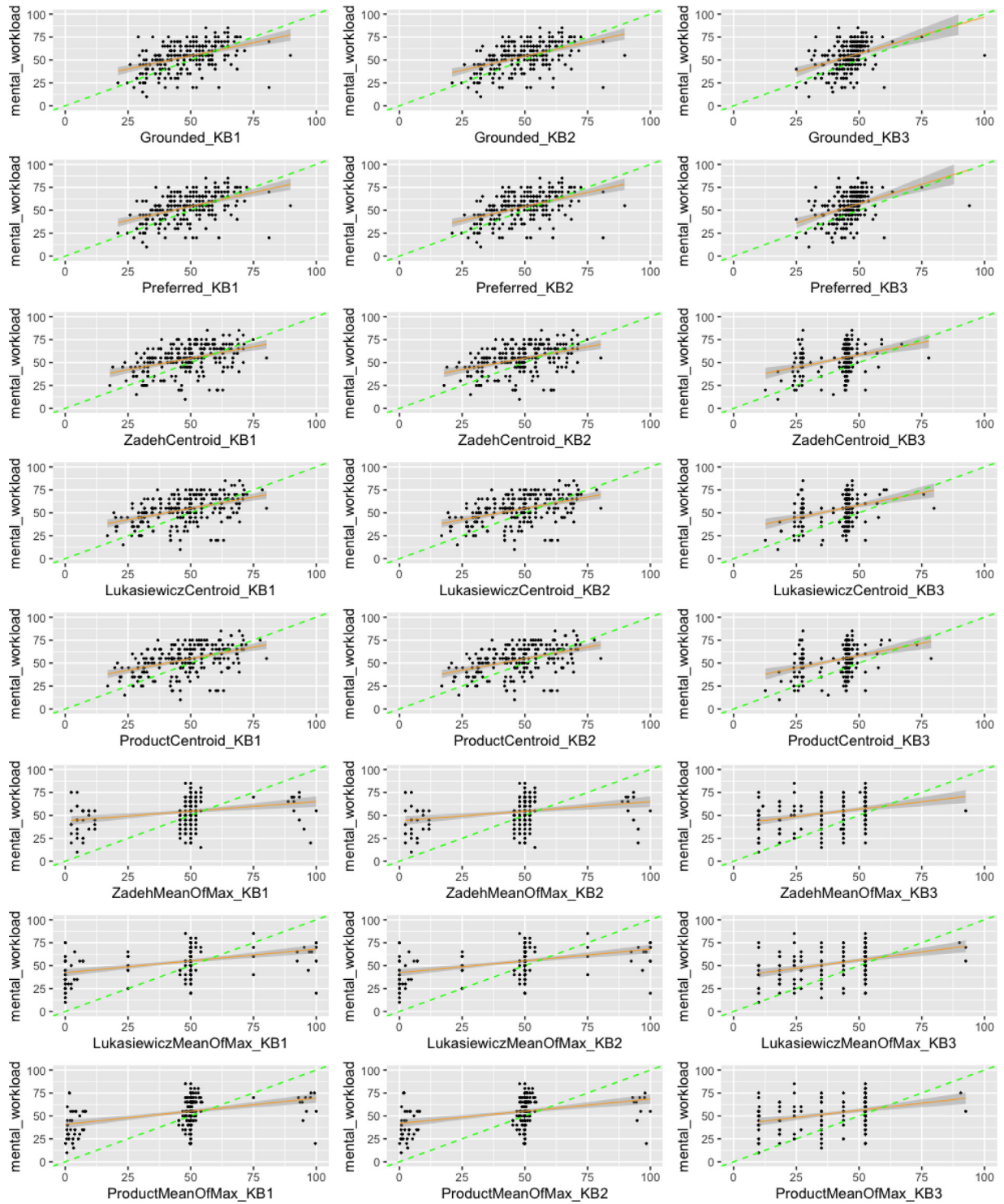


Fig. 23. Scatterplots of inferences produced by designed models and perception of mental workload reported by humans.

squared error when compared to the self reported mental workload over the non-monotonic fuzzy reasoning models, in addition to a slight improvement also against the baseline model (NASA-TLX). Eventually, when results were compared across knowledge-bases, it was evident that when a higher degree of conflictuality of information was present (as knowledge-base 3), then defeasible argumentation seems to be a better modelling tool for handling non-monotonicity when compared to fuzzy reasoning. In other words, defeasible argumentation allowed the construction of models of mental workload with a higher convergent

and face validity when compared to those constructed via non-monotonic fuzzy reasoning, especially when higher uncertainty and conflictuality of information characterise knowledge-bases.

5. Conclusion and future work

This study presented a comparison between non-monotonic fuzzy reasoning and non-monotonic defeasible argumentation using three different knowledge-bases of different complexity

Table 25
Mitigating arguments designed knowledge-base 3 - Part b.

| Label | Attack |
|-------|----------------------|
| UM55 | HPF1 undermines fTD2 |
| UM56 | HPF1 undermines fTD3 |
| UM57 | HPF1 undermines fEF1 |
| UM58 | HPF1 undermines fEF2 |
| UM59 | HPF1 undermines fMD5 |
| UM60 | HPF1 undermines fTD5 |
| UM61 | HPF1 undermines fP1 |
| UM62 | HPF4 undermines fMD2 |
| UM63 | HPF4 undermines fMD3 |
| UM64 | HPF4 undermines fTD2 |
| UM65 | HPF4 undermines fTD3 |
| UM66 | HPF4 undermines fEF1 |
| UM67 | HPF4 undermines fEF2 |
| UM68 | HPF4 undermines fP1 |
| UM69 | HPF4 undermines fP2 |
| UM70 | HPF4 undermines fMD5 |
| UM71 | HPF4 undermines fTD1 |
| UM72 | HPF4 undermines fMD1 |
| UM73 | HPF4 undermines fP1 |
| UM74 | HPF4 undermines fP2 |

coded from an expert human reasoner in the domain of mental workload. A primary research has been conducted including the construction of computational models using these two non-monotonic reasoning approaches to represent the construct of mental workload and to allow its assessment. Such models were constructed from the same information used by a well-known subjective model of mental workload, namely the Nasa Task Load Index, this also used as baseline. The elicitation of these models was made possible by using data gathered in third-level classrooms from students who attended different topics delivered in different ways. The inference produced by each of these models was a single scalar representing an assessment of mental workload that was used for comparison purposes. The selected metrics for the evaluation of the inferential capacity of these designed models were convergent and face validity. The former indicates whether the inferences of the models are coherent with the assessments produced by the selected baseline model (NASA-TLX). The latter is a form of logical validity and indicates whether these models are actually measuring what they are supposed to measure, namely mental workload. Findings indicated how both the models built with non-monotonic fuzzy reasoning and defeasible argumentation had a good convergent validity with the baseline, confirming these are also modelling mental workload as a construct. However, the argument-based models had a superior face validity over the non-monotonic fuzzy reasoning models across the three different knowledge bases. A deeper analysis of the inference of constructed models indicated that, when a knowledge-base is characterised by higher degree of uncertainty and conflictuality, then defeasible argumentation is more suitable for handling non-monotonicity when compared to fuzzy reasoning.

The first contribution of this research lies in the proposal and application of a comparative research aimed at evaluating the impact of different quantitative non-monotonic reasoning approaches for knowledge-representation problems under uncertainty, following and extending previous work in the field [21]. The second contribution is the execution of an experiment that is not purely based upon an analysis of the topological properties of graphs of arguments or the formal characteristics of different reasoning techniques, but rather on an empirical quantification of their inferential capacity by constructing models with human reasoners and by employing data collected in a real-world domain of application. Scholars in the field of logic are provided with a replicable approach for comparing different formalisms for non-monotonic reasoning empirically.

Future work will focus on the replication of this research study by considering further knowledge-bases of different complexities and degrees of conflictuality, by extending the comparison of defeasible argumentation with other reasoning approaches such as non-monotonic expert systems [95]. It will also concentrate on the creation of different models of arguments by employing different semantics for the computation of their dialectical status, such as ranking-based semantics [65] and their application in other real-world knowledge-representation and reasoning problems under uncertainty. Eventually, the comparison can be done also with other argument-based systems, hybrid systems such as fuzzy argumentation and an in depth evaluation of their explainability.

CRediT authorship contribution statement

Luca Longo: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration. **Lucas Rizzo:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - review & editing, Visualization. **Pierpaolo Dondio:** Conceptualization, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

See Figs. 16–23 and Tables 7–25.

References

- [1] Carlos Iván Chesñevar, Ana Gabriela Maguitman, Ronald Prescott Loui, Logical models of argument, *ACM Comput. Surv.* 32 (4) (2000) 337–383.
- [2] Benjamin N. Grosz, Non-monotonicity in probabilistic reasoning, in: *Machine Intelligence and Pattern Recognition*, vol. 5, Elsevier, 1988, pp. 237–249.
- [3] Didier Dubois, Henri Prade, Possibility theory: qualitative and quantitative aspects, in: *Quantified Representation of Uncertainty and Imprecision*, Springer, Dordrecht, 1998, pp. 169–226.
- [4] Lotfi A. Zadeh, Toward a perception-based theory of probabilistic reasoning with imprecise probabilities, in: *Soft Methods in Probability, Statistics and Data Analysis*, Springer, 2002, pp. 27–61.
- [5] Jean Gordon, Edward H. Shortliffe, The Dempster-Shafer theory of evidence, in: *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, vol. 3, Addison-Wesley, 1984, pp. 832–838.
- [6] Luca Longo, Argumentation for knowledge representation, conflict resolution, defeasible inference and its integration with machine learning, in: *Machine Learning for Health Informatics*, Springer, 2016, pp. 183–208.
- [7] Brian R. Gaines, Foundations of fuzzy reasoning, *Int. J. Man-Mach. Stud.* 8 (6) (1976) 623–668.
- [8] Pietro Baroni, Giovanni Guida, Silvano Mussi, Full non-monotonicity: a new perspective in defeasible reasoning, in: *European Symposium on Intelligent Techniques*, 1997, pp. 8–62.
- [9] John L. Pollock, Defeasible reasoning, *Cogn. Sci.* 11 (4) (1987) 481–518.
- [10] Luca Longo, Pierpaolo Dondio, Defeasible reasoning and argument-based systems in medical fields: An informal overview, in: *Computer-Based Medical Systems (CBMS)*, 2014 IEEE 27th International Symposium on, IEEE, 2014, pp. 376–381.
- [11] Daniel Bryant, Paul Krause, A review of current defeasible reasoning implementations, *Knowl. Eng. Rev.* 23 (3) (2008) 227–260.
- [12] Gerhard Brewka, Ilkka Niemelä, Mirosław Truszczyński, Nonmonotonic reasoning, *Found. Artif. Intell.* 3 (2008) 239–284.
- [13] Anthony Hunter, Matthew Williams, Argumentation for aggregating clinical evidence, in: *Tools with Artificial Intelligence (ICTAI)*, 2010 22nd IEEE International Conference on, vol. 1, IEEE, 2010, pp. 361–368.
- [14] Pietro Baroni, Martin Caminada, Massimiliano Giacomin, An introduction to argumentation semantics, *Knowl. Eng. Rev.* 26 (4) (2011) 365–410.

- [15] Trevor J.M. Bench-Capon, Paul E. Dunne, Argumentation in artificial intelligence, *Artificial Intelligence* 171 (10–15) (2007) 619–641.
- [16] Phan Minh Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artificial Intelligence* 77 (2) (1995) 321–358.
- [17] Francesca Toni, Argumentative agents, in: *Computer Science and Information Technology (IMCSIT)*, Proceedings of the 2010 International Multiconference on, IEEE, 2010, pp. 223–229.
- [18] Lucas Rizzo, Ljiljana Majnarić, Pierpaolo Dondio, Luca Longo, An investigation of argumentation theory for the prediction of survival in elderly using biomarkers, in: *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer, 2018, pp. 385–397.
- [19] Lucas Rizzo, Luca Longo, Inferential models of mental workload with defeasible argumentation and non-monotonic fuzzy reasoning: a comparative study, in: *Proceedings of the 2nd Workshop on Advances in Argumentation in Artificial Intelligence*, Co-Located with XVII International Conference of the Italian Association for Artificial Intelligence, AI³@AI*IA 2018, 20–23 November 2018, Trento, Italy, 2018, pp. 11–26.
- [20] Lucas Rizzo, Luca Longo, A qualitative investigation of the explainability of defeasible argumentation and non-monotonic fuzzy reasoning, in: *Proceedings for the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science* Trinity College Dublin, Dublin, Ireland, December 6–7th, 2018, 2018, pp. 138–149.
- [21] Lucas Rizzo, Luca Longo, An empirical evaluation of the inferential capacity of defeasible argumentation, non-monotonic fuzzy reasoning and expert systems, *Expert Syst. Appl.* 147 (2020) 113220.
- [22] Susana Rubio, Eva Díaz, Jesús Martín, José M. Puente, Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods, *Appl. Psychol.* 53 (1) (2004) 61–86.
- [23] Raymond Reiter, A logic for default reasoning, *Artificial Intelligence* 13 (1–2) (1980) 81–132.
- [24] Grigoris Antoniou, Mary-Anne Williams, *Nonmonotonic Reasoning*, MIT Press, 1997.
- [25] Gerhard Brewka, Jürgen Dix, Kurt Konolige, *Nonmonotonic Reasoning: An Overview*, vol. 73, CSLI publications Stanford, 1997.
- [26] Henry Prakken, Gerard Vreeswijk, Logics for defeasible argumentation, in: *Handbook of Philosophical Logic*, Springer, 2001, pp. 219–318.
- [27] Katie Atkinson, Pietro Baroni, Massimiliano Giacomin, Anthony Hunter, Henry Prakken, Chris Reed, Guillermo Simari, Matthias Thimm, Serena Villata, Towards artificial argumentation, *AI Mag.* 38 (3) (2017) 25–36.
- [28] John L. Pollock, *Knowledge and Justification*, Princeton University press, Princeton, 1974.
- [29] Georg Gottlob, Complexity results for nonmonotonic logics, *J. Logic Comput.* 2 (3) (1992) 397–425.
- [30] Juan Luis Castro, Enric Trillas, Jose Manuel Zurita, Non-monotonic fuzzy reasoning, *Fuzzy Sets and Systems* 94 (2) (1998) 217–225.
- [31] Alexander Gegov, Neelamugilan Gobalakrishnan, David Sanders, Rule base compression in fuzzy systems by filtration of non-monotonic rules, *J. Intell. Fuzzy Systems* 27 (4) (2014) 2029–2043.
- [32] Terrance Swift, Tabling for non-monotonic programming, *Ann. Math. Artif. Intell.* 25 (3–4) (1999) 201–240.
- [33] John Fox, David Glasspool, Jonathan Bury, Quantitative and qualitative approaches to reasoning under uncertainty in medical decision making, in: *Conference on Artificial Intelligence in Medicine in Europe*, Springer, 2001, pp. 272–282.
- [34] David Glasspool, John Fox, Ayelet Oettinger, James Smith-Spark, Argumentation in decision support for medical care planning for patients and clinicians, in: *AAAI Spring Symposium: Argumentation for Consumers of Healthcare*, 2006, pp. 58–63.
- [35] Alessandra Mileo, Davide Merico, Roberto Bisiani, Non-monotonic reasoning supporting wireless sensor networks for intelligent monitoring: The sindi system, in: *International Conference on Logic Programming and Nonmonotonic Reasoning*, Springer, 2009, pp. 585–590.
- [36] Vladik Kreinovich, Toward formalizing non-monotonic reasoning in physics: the use of Kolmogorov complexity, *Inteligencia Artificial. Rev. Iberoam. Intel. Artif.* 13 (41) (2009).
- [37] Peter Darcy, Bela Stantic, Abdul Sattar, A fusion of data analysis and non-monotonic reasoning to restore missed RFID readings, in: *2009 International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, IEEE, 2009, pp. 313–318.
- [38] Changyu Dong, Naranker Dulay, Shinren: Non-monotonic trust management for distributed systems, in: *IFIP International Conference on Trust Management*, Springer, 2010, pp. 125–140.
- [39] Lucas Rizzo, Luca Longo, Representing and inferring mental workload via defeasible reasoning: a comparison with the NASA Task Load Index and the Workload Profile, in: *1st Workshop on Advances in Argumentation in Artificial Intelligence*, 2017, pp. 126–140.
- [40] Luca Longo, A defeasible reasoning framework for human mental workload representation and assessment, *Behav. Inf. Technol.* 34 (8) (2015) 758–786.
- [41] Iyad Rahwan, Peter McBurney, Argumentation technology (guest editors), *IEEE Intell. Syst.* 22 (6) (2007) 21–23.
- [42] Paul-Amaury Matt, Morge Morgem, Francesca Toni, Combining statistics and arguments to compute trust, in: *9th International Conference on Autonomous Agents and Multiagent Systems*, Toronto, Canada, vol. 1, ACM, 2010, pp. 209–216.
- [43] Paul Krause, Simon Ambler, Morten Elvang-Gøransson, John Fox, A logic of argumentation for reasoning under uncertainty, *Comput. Intell.* 11 (1) (1995) 113–131.
- [44] Jamal Bentahar, Bernard Moulin, Micheline Bélanger, A taxonomy of argumentation models used for knowledge representation, *Artif. Intell. Rev.* 33 (3) (2010) 211–259.
- [45] Henry Prakken, Giovanni Sartor, The role of logic in computational models of legal argument: a critical survey, in: *Computational Logic: Logic Programming and beyond*, Springer, 2002, pp. 342–381.
- [46] Nicolas Maudet, Simon Parsons, Iyad Rahwan, Argumentation in multi-agent systems: Context and recent developments, in: *International Workshop on Argumentation in Multi-Agent Systems*, Springer, 2006, pp. 1–16.
- [47] Stephen Toulmin, *The Uses of Argument*, Cambridge University Press, 1958.
- [48] Douglas Walton, *Argumentation Schemes for Presumptive Reasoning (Studies in Argumentation Theory)*, Lawrence Erlbaum Associates, Inc., 1996.
- [49] Chris Reed, Douglas Walton, Argumentation schemes in argument-as-process and argument-as-product, in: *Proceedings of the Conference Celebrating Informal Logic*, vol. 25.
- [50] F. Gobbo, J.H.M. Wagemans, A method for reconstructing first-order arguments in natural language, in: *Proceedings of AIXIA 2018*, 2019, pp. 23–27.
- [51] F. Gobbo, M. Benini, *Constructive Adpositional Grammars. Foundations of Constructive Linguistics*, Cambridge Scholars Publishing, 2011.
- [52] J.H.M. Wagemans, Constructing a periodic table of arguments, in: *Argumentation, Objectivity, and Bias: Proceedings of the 11th International Conference of OSSA*, 2016.
- [53] Henry Prakken, An abstract framework for argumentation with structured arguments, *Arg. Comput.* 1 (2) (2010) 93–124.
- [54] John L. Pollock, Justification and defeat, *Artificial Intelligence* 67 (2) (1994) 377–407.
- [55] Paul E. Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, Michael Wooldridge, Weighted argument systems: Basic definitions, algorithms, and complexity results, *Artificial Intelligence* 175 (2) (2011) 457–486.
- [56] Sanjay Modgil, Reasoning about preferences in argumentation frameworks, *Artificial Intelligence* 173 (9–10) (2009) 901–934.
- [57] Diego C. Martínez, Alejandro Javier García, Guillermo Ricardo Simari, An abstract argumentation framework with varied-strength attacks, in: *Int. Conf. on Principles of Knowledge Representation and Reasoning*, 2008, pp. 135–143.
- [58] Jeroen Janssen, Martine De Cock, Dirk Vermeir, Fuzzy argumentation frameworks, in: *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2008, pp. 513–520.
- [59] Pierpaolo Dondio, Probabilistic argumentation frameworks: Basic properties and computation, in: *Highlights on Practical Applications of Agents and Multi-Agent Systems*, 2013, pp. 263–279.
- [60] Gerard Vreeswijk, Defeasible dialectics: A controversy-oriented approach towards defeasible argumentation, *J. Logic Comput.* 3 (1993) 3–27.
- [61] Martin Caminada, Comparing two unique extension semantics for formal argumentation: ideal and eager, in: *Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence (BNAIC 2007)*, Utrecht University Press, 2007, pp. 81–87.
- [62] Pietro Baroni, Massimiliano Giacomin, Semantics of abstract argument systems, in: *Argumentation in Artificial Intelligence*, Springer, 2009, pp. 25–44.
- [63] Serena Villata, Guido Boella, Leendert Van Der Torre, Attack semantics for abstract argumentation, in: *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011, pp. 406–413.
- [64] Martin W.A. Caminada, Walter Alexandre Carnielli, Paul E. Dunne, Semi-stable semantics, *J. Logic Comput.* 22 (5) (2012) 1207–1254.
- [65] Leila Amgoud, Jonathan Ben-Naim, Ranking-based semantics for argumentation frameworks, in: *International Conference on Scalable Uncertainty Management*, Springer, 2013, pp. 134–147.
- [66] Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, Nicolas Maudet, A comparative study of ranking-based semantics for abstract argumentation, in: *AAAI*, 2016, pp. 914–920.
- [67] Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, Srdjan Vesic, Ranking arguments with compensation-based semantics, in: *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 2016, pp. 12–21.
- [68] Pierpaolo Dondio, Ranking semantics based on subgraphs analysis, in: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '18*, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2018, pp. 1132–1140.

- [69] Henry Prakken, A study of accrual of arguments, with applications to evidential reasoning, in: *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, ACM, 2005, pp. 85–94.
- [70] Mauro Javier Gómez Lucero, Carlos Iván Chesñevar, Guillermo Ricardo Simari, On the accrual of arguments in defeasible logic programming, in: *Twenty-First International Joint Conference on Artificial Intelligence*, 2009.
- [71] Mauro J. Gómez Lucero, Carlos I. Chesñevar, Guillermo R. Simari, Modelling argument accrual in possibilistic defeasible logic programming, in: Claudio Sossai, Gaetano Chemello (Eds.), *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 131–143.
- [72] Henry Prakken, Modelling accrual of arguments in aspic+, in: *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, ICAIL '19, ACM, New York, NY, USA, 2019, pp. 103–112.
- [73] Lotfi A. Zadeh, et al., Fuzzy sets, *Inf. Control* 8 (3) (1965) 338–353.
- [74] Chieh-Li Chen, Pey-Chung Chen, et al., Analysis and design of fuzzy control system, *Fuzzy Sets and Systems* 57 (2) (1993) 125–140.
- [75] Riza C. Berkan, Sheldon Trubatch, *Fuzzy System Design Principles*, Wiley-IEEE Press, 1997.
- [76] Chuen-Chien Lee, Fuzzy logic in control systems: fuzzy logic controller. II, *IEEE Trans. Syst. Man Cybern.* 20 (2) (1990) 419–435.
- [77] Kevin M. Passino, Stephen Yurkovich, Michael Reinfrank, *Fuzzy Control*, vol. 20, Citeseer, 1998.
- [78] P.A. Ejegwa, S.O. Akowe, P.M. Otene, J.M. Ikyule, An overview on intuitionistic fuzzy sets, *Int. J. Sci. Technol. Res.* 3 (3) (2014) 142–145.
- [79] Krassimir T. Atanassov, Janusz Kacprzyk, On some modal type intuitionistic fuzzy operators, in: Vassil Sgurev, Ronald R. Yager, Janusz Kacprzyk, Vladimir Jotsov (Eds.), *Innovative Issues in Intelligent Systems*, Springer International Publishing, Cham, 2016, pp. 295–304.
- [80] William Siler, James J. Buckley, *Fuzzy Expert Systems and Fuzzy Reasoning*, John Wiley & Sons, 2005.
- [81] Didier Dubois, Henri Prade, Fuzzy sets and probability: misunderstandings, bridges and gaps, in: [Proceedings 1993] *Second IEEE International Conference on Fuzzy Systems*, IEEE, 1993, pp. 1059–1068.
- [82] Telma Delladio, Nicolás D. Rotstein, Guillermo Ricardo Simari, A comparison between non-monotonic formalisms, in: *VIII Workshop de Investigadores En Ciencias de la Computación*, 2006.
- [83] Kyung Hoon Yang, David Olson, Jaekyung Kim, Comparison of first order predicate logic, fuzzy logic and non-monotonic logic as knowledge representation methodology, *Expert Syst. Appl.* 27 (4) (2004) 501–519.
- [84] Alexander Bochman, Two paradigms of nonmonotonic reasoning, in: *International Symposium on Artificial Intelligence and Mathematics*, ISAIM 2006, Fort Lauderdale, Florida, USA, January 4–6, 2006.
- [85] Catarina Dutilh Novaes, Herman Veluwenkamp, Reasoning biases, non-monotonic logics and belief revision, *Theoria* 83 (1) (2017) 29–52.
- [86] Giuliano Orru, Luca Longo, The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: A review, in: Luca Longo, M. Chiara Leva (Eds.), *Human Mental Workload: Models and Applications*, Springer International Publishing, Cham, 2019, pp. 23–48.
- [87] Brad Cain, A Review of the Mental Workload Literature, Technical Report, Defence research and development Toronto (Canada), 2007.
- [88] Sandra G. Hart, Lowell E. Staveland, Development of NASA-TLX (task load index): Results of empirical and theoretical research, *Adv. Psychol.* 52 (C) (1988) 139–183.
- [89] Karim Moustafa, Luca Longo, Analysing the impact of machine learning to model subjective mental workload: A case study in third-level education, in: Luca Longo, M. Chiara Leva (Eds.), *Human Mental Workload: Models and Applications*, Springer International Publishing, Cham, 2019, pp. 92–111.
- [90] R.D. O'Donnell, F.T. Eggemeier, Workload assessment methodology, in: K.R. Boff, L. Kaufman, J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, in: *Cognitive Processes and Performance*, vol. 2., John Wiley and Sons, Inc, 1986.
- [91] Sanjay Modgil, Henry Prakken, The aspic+ framework for structured argumentation: a tutorial, *Arg. Comput.* 5 (1) (2014) 31–62.
- [92] Guido Governatori, Michael J. Maher, Grigoris Antoniou, David Billington, Argumentation semantics for defeasible logic, *J. Logic Comput.* 14 (5) (2004) 675–702.
- [93] Lucas Rizzo, Evaluating the Impact of Defeasible Argumentation as a Modelling Technique for Reasoning under Uncertainty (Ph.D. thesis), School of Computer Science, Technological University Dublin, 2020.
- [94] Jack Durkin, John Durkin, *Expert Systems: Design and Development*, Prentice Hall PTR, 1998.
- [95] Lucas Rizzo, Pierpaolo Dondio, Sarah Jane Delany, Luca Longo, Modeling mental workload via rule-based expert system: a comparison with NASA-TLX and workload profile, in: *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer, 2016, pp. 215–229.