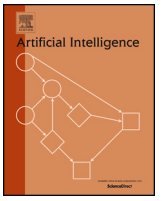
Artificial Intelligence 336 (2024) 104202

Contents lists available at ScienceDirect

Artificial Intelligence

journal homepage: www.elsevier.com/locate/artint



On generalized notions of consistency and reinstatement and their preservation in formal argumentation ☆

Pietro Baroni [a,*], Federico Cerutti [a,b], Massimiliano Giacomin [a]

[a] DII - University of Brescia, Via Branze 38, Brescia, 25123, Italy
[b] University of Cardiff, Cardiff, Wales, UK

ARTICLE INFO

MSC:
68T30
68T37

Keywords:
Formal argumentation
Consistency
Reinstatement

ABSTRACT

We present a conceptualization providing an original domain-independent perspective on two crucial properties in reasoning: consistency and reinstatement. They emerge as a pair of dual characteristics, representing complementary requirements on the outcomes of reasoning processes. Central to our formalization are two underlying parametric relations: incompatibility and reinstatement violation. Different instances of these relations give rise to a spectrum of consistency and reinstatement scenarios. As a demonstration of versatility and expressive power of our approach we provide a characterization of various abstract argumentation semantics which are expressed as combinations of distinct consistency and reinstatement constraints. Moreover, we conduct an investigation into preserving these essential properties across different reasoning stages. Specifically, we delve into scenarios where a labelling is derived from other labellings through a synthesis function, using the synthesis of argument justification as an illustrative instance. We achieve a general characterization of consistency preservation synthesis functions, while we unveil an impossibility result concerning reinstatement preservation, leading us to explore an alternative notion to ensure feasibility. Our exploration reveals a weakness in the traditional definition of argument justification, for which we propose a refined version overcoming this limitation.

1. Introduction

In many contexts, intelligent agents need to assess the elements pertaining to a scenario of their interest with respect to some criterion, *e.g.*, evaluating a set of propositions according to their credibility, or a set of actions according to their advisability.

Generally speaking, the production of this kind of assessment has to take into account two complementary needs. On the one hand, agents need to satisfy some constraints, inherent to the scenario of interest, which prevent the simultaneous positive evaluation of some elements. For instance, one cannot fully believe a proposition and its negation at the same time or cannot select together two

☆ This paper integrates and substantially extends the initial results presented at a conference [1] and a workshop [2] by the same authors. In particular, [1] deals with generalizing the notion of consistency only and thus provides some partial and preliminary results with respect to the present work. The subsequent workshop paper [2] extends [1] by providing some results concerning both notions, but does not address the problem of preservation. Moreover, the discussion of related literature is quite limited in [1] and [2].

* Corresponding author.
E-mail addresses: pietro.baroni@unibs.it (P. Baroni), federico.cerutti@unibs.it (F. Cerutti), massimiliano.giacomin@unibs.it (M. Giacomin).

https://doi.org/10.1016/j.artint.2024.104202
Received 21 September 2023; Received in revised form 4 June 2024; Accepted 9 August 2024

Available online 18 August 2024

incompatible actions, like moving straight and turning right. Thus, in a sense, these constraints provide a sort of upper bound to the positive evaluations one can produce.

On the other hand, agents also need to avoid unjustified, overly negative evaluations since they would lead to a situation of inertia and inability to act, as in the extreme case where nothing is believed and none of the available actions is adopted. Thus, the need to avoid an undesirable situation of nihilism and paralysis induces a sort of lower bound on the evaluations that is reasonable to produce.

The fact that the information used to produce the evaluation is typically incomplete and uncertain, together with the variety of attitudes an agent can adopt, ranging from cautiousness to braveness, introduces a spectrum of alternatives in the above sketched picture.

First, the choice of upper and lower bounds deemed sensible can be subjective. Further, there can be room for the production of a set of different assessments, each individually reasonable, being included in the range between the chosen upper and lower bounds. Last but not least, producing a set of different assessments may call for a further step of synthesis, whose outcome should also obey the chosen bounds.

The high-level description provided above finds a technical counterpart in the field of formal argumentation [3–6]. In a nutshell, it provides models to represent the reasoning activity of an agent in terms of the production and evaluation of arguments, where, in general terms, an argument can be conceived as a possibly (and typically) uncertain and fallible derivation of some conclusion from a set of premises.

In particular, in formal argumentation, the presence of conflicts between arguments is a key unavoidable aspect. It represents a challenge that calls for mechanisms able to produce sensible reasoning outcomes in terms of assessments of the acceptability of arguments, given their conflicts. These outcomes are typically required to satisfy two somewhat dual properties, intuitively corresponding to the upper and lower bounds previously discussed.

On the one hand, the outcomes are required to respect some notion of *consistency* related to the existence of attacks. For instance, in abstract argumentation semantics [7,8], the produced argument assessments are typically required to satisfy the property of *conflict-freeness*, namely if there is a conflict between two arguments they cannot be accepted at the same time.

On the other hand, the outcomes are also required to comply with some notion of *reinstatement* [9], entailing that arguments cannot be rejected without reason. For instance, a requirement common to many argumentation semantics is that an argument is accepted when all of its attackers are rejected.

While the properties mentioned above are included among the basic principles underlying the definition of abstract argumentation semantics [10,11], the following observations suggest that a deeper foundational investigation of their nature and properties is worth pursuing.

First, while these properties are usually defined with reference to a specific formal context (like *e.g.*, abstract argumentation), they correspond to basic intuitive requirements, which can be found in various, conceptually analogous, forms across different domains, like, for instance, multi-criteria decision making, voting systems, legal reasoning, and belief revision. This suggests the opportunity of devising a general domain-independent characterization of the underlying common notions, in order to shed light on the shared essential elements underlying different contexts, and of studying their properties at a level of abstraction that allows a broad reuse of the achieved results.

Second, even when focusing on a specific formal context, it emerges that consistency and reinstatement may allow a spectrum of actual realizations. Framing this spectrum of possibilities within a domain-independent model is very useful for systematically analyzing and comparing the set of available alternatives. This may support a better understanding and reappraisal from an original perspective of solutions already considered in the past and may also favour the investigation of novel ones.

Third, a reasoning process may involve a sequence of stages, each stage producing some assessment and passing it as input to the next one. As different as these assessments at different stages may be, they may share the need to comply with some consistency and reinstatement requirements. The question of ensuring the preservation of these requirements across different stages appears to be of high interest both from a theoretical and a practical point of view. Again, addressing this question imposes the use of a general model, able to capture and relate the properties of different stages within a unifying framework.

Building on the above observations, this paper contributes to the study of general notions of consistency and reinstatement and demonstrates their use in formal argumentation as follows.

To address the first point, Section 2 introduces a general representation of assessments based on labellings and proposes generalized notions of consistency and reinstatement, applicable in any context where a labelling approach is adopted. The essential ingredients of the approach are an intolerance relation, which indicates pairs of labelled elements that cannot stay together, and two relations between labels, called incompatibility and reinstatement violation, which express the constraints corresponding to consistency and reinstatement, respectively. Well-foundedness properties for these relations are investigated, and their dual nature is evidenced.

As a demonstration of the ability of the proposed notions to capture a spectrum of alternatives, we prove in Section 3 that a variety of traditional abstract argumentation semantics proposed by Dung [7] can be expressed as combinations of different consistency and reinstatement requirements, *i.e.*, varying the choice of the incompatibility and reinstatement violation relations.

Then, concerning the third point, in Section 4 we examine the issue of consistency and reinstatement preservation in the case where a set of labellings is aggregated through a simple synthesis function. This abstract notion captures, in particular, the process of deriving a synthetic argument justification status from a set of argument acceptability labellings. As to consistency preservation, we achieve a characterization of consistency preserving simple synthesis functions and use it to analyze the preservation capabilities of the traditional notion of argument justification. As to reinstatement preservation, we obtain an impossibility result in the general case and, to overcome this difficulty, propose a weaker requirement and then provide a characterization of weakly reinstatement

preserving simple synthesis functions. The developed analysis leads us to point out a limitation, from a reinstatement perspective, of the traditional notion of argument justification and to propose an improved version.

The paper is then completed by a discussion of related works and future research perspectives in Section 5 and some final remarks in Section 6.

2. Generalizing consistency and reinstatement for labelling-based systems

In a variety of contexts, the assessments of entities of various kinds are expressed by assigning them a label. To provide a common ground to characterize such different contexts, we present a three-layer model based on the proposal in [1], which includes the following levels:

- At the top level, the notion of assessment classes is introduced to provide a reference point to characterize different assessment labels and to relate and compare them. These classes have an underlying order, intuitively reflecting a notion of *positivity*, whatever the meaning assigned to *positive* in a given context.
- At an intermediate level, assessment labels are taken from a predefined set and classified on the basis of assessment classes, thus inheriting the relevant positivity degree.
- At the bottom level, a generic set of entities can be assessed by assigning each entity a label.

Example 2.1. An illustration of the three-layer model, which will be used as running example, is provided in Fig. 1. We provide a high level description here, while more details will be introduced later along with formal definitions. At the top layer we consider a set C^3 of three assessment classes, $\mathsf{pos}, \mathsf{mid}, \mathsf{neg}$, intuitively corresponding to a positive (or high), intermediate (or medium), and negative (or low) evaluation. These classes can be used to provide a common reference scale for sets of labels adopted in different contexts and constituting the second layer of the model. For instance, in Fig. 1 we consider two sets of labels that can be used to acquire customer opinions about the features of some products, e.g. through a website. Some websites may use a scale based on stars (from 1 to 5), indicated as Λ_1, others may use a less refined scale, Λ_2, based on three emoticons. In order to aggregate or compare opinions coming from different sources, sets of labels have to be put in correspondence with the common reference scale. A possible way of doing this is shown in the figure. At the third layer, entities to be assessed are labelled using the sets of labels from the second layer. In the figure we consider acquiring the indications of customers about the features they desire in a travel. The example considers four features and presents the indications of three hypotethical customers expressed as labellings, where L_1 and L_3 are based on Λ_1, while L_2 is based on Λ_2.

The model is general and applicable to any domain where labellings are used for assessment. Besides the illustrative example in Fig. 1, formal argumentation (see [3] for a comprehensive overview of the field), being a paradigmatic example of such domains, will be used as the main reference context to illustrate the approach and then to achieve specific results based on it.

Definition 2.1. A *set of assessment classes* is a set C equipped with a total order $\leq$ (*i.e.*, a reflexive, transitive, and antisymmetric relation such that any two elements are comparable) and including a maximum and a minimum element (*i.e.*, an element $c \in C$ such that $\forall c' \in C$ it holds that $c' \leq c$ or $c \leq c'$, respectively) which are assumed to be distinct.

In the following, we will abbreviate the term 'set(s) of assessment classes' as sac(s). Intuitively, the order is meant to capture an abstract distinction between different levels of positivity of the assessment, with $c_1 \leq c_2$ meaning that c_2 corresponds to an at least as positive assessment as c_1. We will mainly use the tripolar sac $C^3 = \{\mathsf{pos}, \mathsf{mid}, \mathsf{neg}\}$ shown in Fig. 1, with $\mathsf{neg} \leq \mathsf{mid} \leq \mathsf{pos}$. The basic idea, expressed by the following definition, is that a sac is used to classify the elements of a set of labels according to their level of positivity. Note that the elements of a sac are called classes because, in general, more than one label can be mapped to the same class.

Definition 2.2. Given a set of assessment classes C, a *C-classified set of assessment labels* is a set Λ equipped with a total function $C_\Lambda : \Lambda \to C$. The total preorder induced on Λ by C_Λ will be denoted by $\preceq$ where $\lambda_1 \preceq \lambda_2$ iff $C_\Lambda(\lambda_1) \leq C_\Lambda(\lambda_2)$. As usual, $\lambda_1 \prec \lambda_2$ will denote $\lambda_1 \preceq \lambda_2$ and $\lambda_2 \not\preceq \lambda_1$, while $\lambda_1 \approx \lambda_2$ will denote $\lambda_1 \preceq \lambda_2$ and $\lambda_2 \preceq \lambda_1$.

The fact that $\preceq$ is a total preorder is shown in the following proposition.

Proposition 2.1. *Given a set of assessment classes C and a C-classified set of assessment labels Λ, the relation $\preceq$ as introduced in Definition 2.2 is reflexive and transitive, and for any $\lambda_1, \lambda_2 \in \Lambda$, $\lambda_1 \preceq \lambda_2$ or $\lambda_2 \preceq \lambda_1$.*

Proof. Consider $\lambda_1, \lambda_2, \lambda_3 \in \Lambda$. By reflexivity of $\leq$ it holds that $C_\Lambda(\lambda_1) \leq C_\Lambda(\lambda_1)$, *i.e.*, $\lambda_1 \preceq \lambda_1$. Similarly, if $\lambda_1 \preceq \lambda_2$ and $\lambda_2 \preceq \lambda_3$, by transitivity of $\leq$ it holds that $C_\Lambda(\lambda_1) \leq C_\Lambda(\lambda_3)$, *i.e.*, $\lambda_1 \preceq \lambda_3$. Finally, since $\leq$ is total and reflexive it holds that $C_\Lambda(\lambda_1) \leq C_\Lambda(\lambda_2)$ or $C_\Lambda(\lambda_2) \leq C_\Lambda(\lambda_1)$, *i.e.*, $\lambda_1 \preceq \lambda_2$ or $\lambda_2 \preceq \lambda_1$. □

It is easy to see that $\preceq$ is not necessarily an order, since different labels can be classified with the same assessment class, thus antisymmetry does not hold.

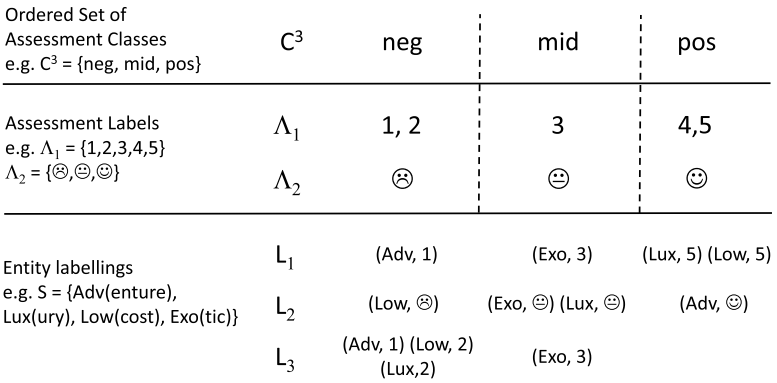


Fig. 1. An illustration of the three-layer model.

We will abbreviate the term 'set(s) of assessment labels' as sal(s) and omit 'C-classified', when C is not ambiguous. Also, to distinguish preorders referring to different sals, given a sal Λ we will denote the relevant preorder as $\preceq_\Lambda$.

Example 2.2. In Fig. 1 we show two sals, namely Λ_1 and Λ_2. In particular, the classification of the labels in Λ_1 is given by $C^3_{\Lambda_1} = \{(1,\mathsf{neg}),(2,\mathsf{neg}),(3,\mathsf{mid}),(4,\mathsf{pos}),(5,\mathsf{pos})\}$. The induced ordering is such that $1 \approx 2 \prec 3 \prec 4 \approx 5$. The function $C^3_{\Lambda_2}$ and the relevant induced ordering are obvious.

We can now introduce the notion of labelling based on a sal.

Definition 2.3. Given a sal Λ and a set S, a Λ-*labelling* of S is a function $L : S \to \Lambda$.

Example 2.3. Different sals can be used to express assessments in distinct, but possibly related, evaluation contexts. In Fig. 1, the set S consists of four features of a travel, namely adventure, luxury, low cost, exotic (denoted shortly as Adv, Lux, Low, Exo). L_1 and L_3 are Λ_1-labellings of S, while L_2 is a Λ_2-labelling of S. L_1 expresses the opinion of a customer who values luxury and low cost and definitely dislikes adventure in a travel, while L_2 provides the rather complementary view of a customer who appreciates adventure and is not really interested in low cost.

Turning to formal argumentation, in the context of argument acceptance evaluation based on the labelling-based version of Dung's semantics [7,8], the sal $\Lambda^{\mathrm{IOU}} = \{\mathsf{in},\mathsf{out},\mathsf{und}\}$ is used, while in Defeasible Logic Programming [12] arguments are marked as D(efeated) or U(ndefeated) corresponding to the use of the sal $\Lambda^{\mathrm{De}} = \{\mathsf{D},\mathsf{U}\}$, and in [13] an approach using the set of four labels $\Lambda^{\mathrm{JV}} = \{+,-,\pm,\emptyset\}$ is proposed. The sals mentioned above can be C^3-classified as follows: $C^3_{\Lambda^{\mathrm{IOU}}} = \{(\mathsf{in},\mathsf{pos}),(\mathsf{out},\mathsf{neg}),(\mathsf{und},\mathsf{mid})\}$; $C^3_{\Lambda^{\mathrm{De}}} = \{(\mathsf{D},\mathsf{neg}),(\mathsf{U},\mathsf{pos})\}$; $C^3_{\Lambda^{\mathrm{JV}}} = \{(-,\mathsf{neg}),(+,\mathsf{pos}),(\pm,\mathsf{mid}),(\emptyset,\mathsf{mid})\}$.

We are now ready to introduce the generalized notions of consistency and reinstatement in this formal context. Intuitively, they correspond to dual requirements aimed at satisfying somehow conflicting goals.

An inconsistency arises when two elements of a set that cannot stay together are assigned labels which are *too positive* overall. Correspondingly, consistency is satisfied whenever this situation does not hold for any couple of elements.

Reinstatement is violated when an element of a set is assigned a label which is *too negative*, *i.e.*, a negative label is assigned without a sufficient reason. A sufficient reason holds if another element that cannot stay together is assigned a sufficiently positive label. Correspondingly, reinstatement holds whenever a sufficiently positive label is assigned to any element such that all of its conflicting elements are negatively assessed.

It can be seen that consistency and reinstatement are dual properties. In particular, a skeptical assessment which assigns the most negative label to all elements trivially satisfies consistency but violates reinstatement. Conversely, assigning the most positive label to all elements trivially satisfies reinstatement, but violates consistency whenever two elements cannot stay together.

According to this informal introduction, both inconsistency and reinstatement violation can be understood as arising from two components: an intolerance relation at the level of the assessed elements, indicating who cannot stay with whom, and a relation at the level of the labels indicating which pairs of assessments correspond to a violation if ascribed to a pair of elements connected by the intolerance relation.

Definition 2.4. Given a set S, an *intolerance relation* on S is a binary relation $\mathsf{int} \subseteq S \times S$, where $(s_1,s_2) \in \mathsf{int}$ indicates that s_1 is intolerant of s_2 and will be denoted as $s_1 \odot s_2$, while $(s_1,s_2) \notin \mathsf{int}$ will be denoted as $s_1 \ominus s_2$. We will also denote as $snt(s_2)$ the set of elements intolerant of s_2: $snt(s_2) \triangleq \{s_1 \in S \mid s_1 \odot s_2\}$.

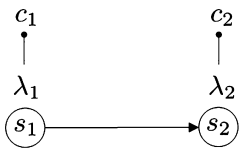


Fig. 2. Two elements s_1, s_2 such that $s_1 \odot s_2$. The intolerance relation is graphically represented by an arrow.

Note that we do not make any assumption on the intolerance relation; in particular, it needs not to be symmetric.

Example 2.4. In our running example, we can assume that some features cannot be present in the same travel, because they are inherently in conflict (or just because there are no available travel offers enjoying those features together). For instance, one can consider that there is a mutual intolerance between adventure and luxury, and between luxury and low cost, yielding for our travel example the intolerance relation $\mathsf{int}_{tr} = \{(\mathsf{Adv}, \mathsf{Lux}), (\mathsf{Lux}, \mathsf{Adv}), (\mathsf{Low}, \mathsf{Lux}), (\mathsf{Lux}, \mathsf{Low})\}$.

Turning to formal argumentation, in languages equipped with negation, typically intolerance between language elements coincides with negation (a symmetric relation where each element has exactly one opposite). However, more general forms of contrariness have been considered in argumentation contexts, where the corresponding intolerance relation may not be symmetric and allows the existence of multiple contraries for an element [14,15]. Moving from the language level to the argument level, the attack relation in Dung's argumentation frameworks (see Section 3) can be regarded as an example of intolerance relation.

Due to the dual nature of consistency w.r.t. reinstatement, violations at the level of the labellings are modelled by distinct relations, namely an incompatibility relation and a reinstatement violation relation on assessment labels, respectively. In the following, we will assume that each of these relations on assessment labels is always induced by a corresponding relation on assessment classes.

Definition 2.5. Given a sac C, an *incompatibility relation* on C is a relation $\mathsf{inc} \subseteq C \times C$, where $(c_1, c_2) \in \mathsf{inc}$ indicates that c_1 is incompatible with c_2 and will be denoted as $c_1 \underline{\boxdot} c_2$, while $(c_1, c_2) \notin \mathsf{inc}$ will be denoted as $c_1 \underline{\boxminus} c_2$. Given a C-classified sal Λ, we define the *induced incompatibility relation* $\mathsf{inc}' \subseteq \Lambda \times \Lambda$ as follows: for every $\lambda_1, \lambda_2 \in \Lambda$, $(\lambda_1, \lambda_2) \in \mathsf{inc}'$ iff $(C_\Lambda(\lambda_1), C_\Lambda(\lambda_2)) \in \mathsf{inc}$. With a little abuse of notation we will also denote $(\lambda_1, \lambda_2) \in \mathsf{inc}'$ as $\lambda_1 \underline{\boxdot} \lambda_2$, and analogously for $\lambda_1 \underline{\boxminus} \lambda_2$. Given a label λ, we define the set of labels which are *compatible* with λ according to inc' as $cc(\lambda) \triangleq \{\lambda' \in \Lambda \mid (\lambda, \lambda') \notin \mathsf{inc}'\}$.

Definition 2.6. Given a sac C, a *reinstatement violation relation* on C is a relation $\mathsf{rv} \subseteq C \times C$, where $(c_1, c_2) \in \mathsf{rv}$ indicates that c_1 is not sufficiently positive to justify c_2 and will be denoted as $c_1 \overline{\underline{\boxdot}} c_2$, while $(c_1, c_2) \notin \mathsf{rv}$ will be denoted as $c_1 \overline{\underline{\boxminus}} c_2$. Given a C-classified sal Λ, we define the *induced reinstatement violation relation* $\mathsf{rv}' \subseteq \Lambda \times \Lambda$ as follows: for every $\lambda_1, \lambda_2 \in \Lambda$, $(\lambda_1, \lambda_2) \in \mathsf{rv}'$ iff $(C_\Lambda(\lambda_1), C_\Lambda(\lambda_2)) \in \mathsf{rv}$. With a little abuse of notation we will also denote $(\lambda_1, \lambda_2) \in \mathsf{rv}'$ as $\lambda_1 \overline{\underline{\boxdot}} \lambda_2$, and analogously for $\lambda_1 \overline{\underline{\boxminus}} \lambda_2$. Given a label λ, we define the set of labels which are *compatible* with λ according to rv' as $rc(\lambda) \triangleq \{\lambda' \in \Lambda \mid (\lambda, \lambda') \notin \mathsf{rv}'\}$ and the set of labels which are *backwards compatible*[1] with λ according to rv' as $\overleftarrow{rc}(\lambda) \triangleq \{\lambda' \in \Lambda \mid (\lambda', \lambda) \notin \mathsf{rv}'\} = \{\lambda' \in \Lambda \mid \lambda \in rc(\lambda')\}$.

From the intuitions underlying the concepts of consistency and reinstatement, some rather natural properties can be identified for incompatibility and, in a dual manner, for reinstatement violation relations on C. The following definition introduces these properties for incompatibility relations.

Definition 2.7. Given a sac C, let inc be an incompatibility relation on C. We say that inc is *well-founded* if it satisfies the following properties:

- inc is *monotonic*, *i.e.*, given $c_1, c_2 \in C$ such that $c_1 \underline{\boxdot} c_2$, for every pair $c_1', c_2' \in C$ such that $c_1 \leq c_1'$ and $c_2 \leq c_2'$ it holds that $c_1' \underline{\boxdot} c_2'$
- inc is non empty, *i.e.*, $\mathsf{inc} \neq \emptyset$
- $\forall c_1 \in C$, $\exists c_2 \in C$ such that $c_1 \underline{\boxminus} c_2$ and $\exists c_3 \in C$ such that $c_3 \underline{\boxminus} c_1$

In order to discuss these properties, let us remark again that incompatibility refers to the situation where labels are assigned to entities which are linked by intolerance. For instance, in a context where statements are assessed and intolerance between them corresponds to contradiction, two (not necessarily distinct) positive labels expressing belief should be incompatible: they cannot be assigned to two contradictory statements, since you cannot believe both of them.

Let us then consider the simple case depicted in Fig. 2, involving two elements $s_1, s_2 \in S$ such that $s_1 \odot s_2$, and a Λ-labelling L such that $L(s_1) = \lambda_1$, $L(s_2) = \lambda_2$, $C_\Lambda(\lambda_1) = c_1$ and $C_\Lambda(\lambda_2) = c_2$.

The first property of Definition 2.7 relies on the idea that inconsistency arises from a sort of 'excess of simultaneous positiveness' in the assessment of some elements linked by intolerance. In particular, $c_1 \underline{\boxdot} c_2$ indicates that the simultaneous positiveness of the

[1] Backward compatibility concerns the case where the label λ is assigned to the target (rather than to the source) of an intolerance relation and λ' is assigned the source of the same relation. As it will be clear later, this notion is technically needed only for reinstatement violation.

labels λ_1 and λ_2 is not tolerated for two incompatible elements s_1 and s_2. Since the simultaneous positiveness expressed by c'_1 and c'_2 is not lesser that the one expressed by c_1 and c_2, then it must hold $c'_1 \underline{\boxdot} c'_2$.

The second property requires at least one labelling to yield inconsistency for two elements related by intolerance. Otherwise, an empty relation would completely neglect the intolerance relation between elements of S.

The third property requires each label to be attainable for s_1 and s_2 without necessarily generating inconsistencies, otherwise the role of the label would be too weak. For instance, if there was a class c_1 such that $\nexists c_2 \in C$ with $c_1 \underline{\boxminus} c_2$, this would mean that a label of class c_1 could not be assigned to any element which is the source of an intolerance relation without generating inconsistencies, implying that the class c_1 would have a very limited applicability (actually it could not be used at all in most cases). Similarly, if $\nexists c_3 \in C$ such that $c_3 \underline{\boxminus} c_1$, a label of class c_1 could not be assigned to any element which is the target of an intolerance relation, implying again a very limited applicability of labels of class c_1.

Two additional intuitive properties of well-founded incompatibility relations can be derived. First, two maximally positive labels cannot be ascribed together to conflicting elements. Second, the maximally negative label is compatible with any other label, in particular, $\min(C) \underline{\boxminus} \min(C)$, $\max(C) \underline{\boxminus} \min(C)$ and $\min(C) \underline{\boxminus} \max(C)$.

Proposition 2.2. *Given a* sac C*, let* inc *be a well-founded incompatibility relation on* C*. It then holds that:*

- $\max(C) \underline{\boxdot} \max(C)$
- $\nexists c \in C$ *such that* $c \underline{\boxdot} \min(C)$ *or* $\min(C) \underline{\boxdot} c$

Proof. As to the first property, since $\mathsf{inc} \neq \emptyset$ there are $c_1, c_2 \in C$ such that $c_1 \underline{\boxdot} c_2$. Taking into account that inc is monotonic, it obviously holds that $\max(C) \underline{\boxdot} \max(C)$.

As to the second property, assume by contradiction that $\exists c \in C$ such that $c \underline{\boxdot} \min(C)$. By the monotonicity property and the definition of $\min(C)$, $\forall c' \in C$ it holds that $c \underline{\boxdot} c'$, violating the third condition of Definition 2.7. The other condition can be proved in the same way. □

According to the above proposition, we can identify for any sac C the minimal well-founded incompatibility relation as $\underline{\mathsf{inc}}_C = \{(\max(C), \max(C))\}$.

Let us turn now to well-founded reinstatement violation relations.

Definition 2.8. Given a sac C, let rv be a reinstatement violation relation on C. We say that rv is *well-founded* iff it satisfies the following properties:

- rv is *dually monotonic*,[2] *i.e.*, given $c_1, c_2 \in C$ such that $c_1 \overline{\boxdot} c_2$, for every pair $c'_1, c'_2 \in C$ such that $c'_1 \leq c_1$ and $c'_2 \leq c_2$ it holds that $c'_1 \overline{\boxdot} c'_2$
- rv is non empty, *i.e.*, $\mathsf{rv} \neq \emptyset$
- $\forall c_1 \in C$, $\exists c_2 \in C$ such that $c_1 \overline{\boxminus} c_2$ and $\exists c_3 \in C$ such that $c_3 \overline{\boxminus} c_1$

In order to provide an explanation of these requirements, let us refer again to the simple case depicted in Fig. 2.

As to the first condition, we remark that reinstatement violation arises from a sort of 'excess of cautiousness' in assigning positive labels, *i.e.*, a too much negative label is assigned to an element even in the absence of a positively assessed element linked by intolerance. Let us then consider the case where, in Fig. 2, $c_1 \overline{\boxdot} c_2$. This situation can be interpreted in two equivalent ways:

1. The label λ_1 assigned to s_1 is too much negative to justify the label λ_2 assigned to s_2
2. The label λ_2 assigned to s_2 is too much negative w.r.t. the 'not so positive' label λ_1 assigned to s_1

Accordingly, if $c'_1 \leq c_1$ (*i.e.*, positiveness of λ_1 does not increase) and $c'_2 \leq c_2$ (*i.e.*, positiveness of λ_2 does not increase), then it must also hold $c'_1 \overline{\boxdot} c'_2$.

The second condition, *i.e.*, that rv is non empty, is required to avoid an overly skeptical assessment attitude such that assigning the most negative label to all elements is always allowed, independently of the labels of incompatible elements.

The third condition has an analogous rationale w.r.t. the analogous condition appearing in Definition 2.7.

Also in this case, two additional intuitive properties of well-founded reinstatement violation relations can be derived. First, two minimally positive labels cannot be ascribed together to conflicting elements.[3] Second, the maximally positive label is compatible with any other label, in particular $\max(C) \overline{\boxminus} \max(C)$, $\max(C) \overline{\boxminus} \min(C)$ and $\min(C) \overline{\boxminus} \max(C)$.

Proposition 2.3. *Given a* sac C*, let* rv *be a well-founded reinstatement violation relation on* C*. It then holds that:*

[2] We use this term since the property is preserved for decreasing, rather than increasing values in C.

[3] This refers to the simple case of Fig. 2, while the general case is handled according to Definition 2.10.

- $\min(C)\overline{\boxdot}\min(C)$
- $\nexists c \in C$ *such that* $c\overline{\boxdot}\max(C)$ *or* $\max(C)\overline{\boxdot}c$

Proof. As to the first property, since $\mathsf{rv} \neq \emptyset$ there are $c_1, c_2 \in C$ such that $c_1\overline{\boxdot}c_2$. Taking into account that inc is dually monotonic, it obviously holds that $\min(C)\overline{\boxdot}\min(C)$.

As to the second property, assume by contradiction that $\exists c \in C$ such that $c\overline{\boxdot}\max(C)$. By the dual monotonicity property and the definition of $\max(C)$, $\forall c' \in C$ it holds that $c\overline{\boxdot}c'$, violating the third condition of Definition 2.8. The other condition can be proved in the same way. □

According to the above proposition, we can identify for any sac C the minimal well-founded reinstatement violation relation as $\underline{\mathsf{rv}}_C = \{(\min(C), \min(C))\}$.

While we have considered above the particular case involving only a couple of elements of S, in order to introduce our generalized notions of inconsistency and reinstatement violation we have to consider the general case of labellings of generic sets.

Let us start with defining when a labelling is inconsistent.

Definition 2.9. Given a set S equipped with an intolerance relation int, a sac C equipped with an incompatibility relation inc, and a C-classified sal Λ, a Λ-labelling L of S is int-inc-*inconsistent* iff

$$\exists s_1, s_2 \in S \text{ such that } s_1 \odot s_2 \text{ and } L(s_1)\underline{\boxdot}L(s_2) \tag{1}$$

Conversely, we say that a labelling is int-inc-*consistent* if it is not int-inc-inconsistent, *i.e.*,

$$\forall s_1, s_2 \in S \text{ such that } s_1 \odot s_2, \text{ it holds that } L(s_1)\underline{\boxminus}L(s_2) \tag{2}$$

The above definition corresponds to the idea that consistency violation arises from an excess of simultaneous positivity between any couple of incompatible elements, *i.e.*, given $s_1 \in S$ a single s_2 satisfying the inc relation is sufficient to yield inconsistency.

The following proposition is obvious and will not be proved.

Proposition 2.4. *Given a set S equipped with an intolerance relation* int*, a* sac C *and a C-classified* sal Λ*, consider two incompatibility relations* inc *and* inc′ *such that* inc $\subseteq$ inc′*. Then, an* int-inc-*inconsistent Λ-labelling L of S is also* int-inc′*-inconsistent, and an* int-inc′*-consistent Λ-labelling L of S is also* int-inc-*consistent.*

Example 2.5. With reference to our running example, an inconsistent labelling can be intuitively regarded as corresponding to a customer who 'wants too much' with respect to what is possible (or actually offered) in a travel. Assume the minimal well-founded incompatibility relation for our case, namely $\underline{\mathsf{inc}}_{C^3} = \{(\mathsf{pos}, \mathsf{pos})\}$. Given the intolerance relation int_{tr} previously introduced, we get that L_1 is $\mathsf{int}_{tr} - \underline{\mathsf{inc}}_{C^3} - inconsistent$, since both Lux and Low, which are mutually intolerant, have a label of class pos. L_2 and L_3 are instead $\mathsf{int}_{tr} - \underline{\mathsf{inc}}_{C^3} - consistent$. One may however consider a stronger notion of consistency, where a positive label on an element is compatible only with a negative label on an intolerant element: this would correspond to the incompatibility relation $\underline{\mathsf{inc}}^+_{C^3} = \{(\mathsf{pos}, \mathsf{pos}), (\mathsf{pos}, \mathsf{mid}), (\mathsf{mid}, \mathsf{pos})\}$. By Proposition 2.4 (and as it can be obviously seen), L_1 is also $\mathsf{int}_{tr} - \underline{\mathsf{inc}}^+_{C^3} - inconsistent$. Moreover, L_2 is $\mathsf{int}_{tr} - \underline{\mathsf{inc}}^+_{C^3} - inconsistent$, since the mutually intolerant feature Adv and Lux have labels of the classes pos and mid respectively. Intuitively, under this stricter view, the labelling L_2 corresponds to a customer who values adventure a lot, but is not available enough to renounce luxury.

Turning to reinstatement violation, duality w.r.t. inconsistency is reflected also in the counterpart of Definition 2.9. In particular, given $s_2 \in S$, reinstatement is violated if *all* the elements s_1 that are intolerant w.r.t. s_2 do not provide a sufficient reason (*i.e.*, are not positive enough) to justify the 'not so positive' label assigned to s_2. Accordingly, a Λ-labelling L of S should violate reinstatement iff

$$\exists s_2 : \forall s_1 \in S \text{ such that } s_1 \odot s_2 \text{ it holds that } L(s_1)\overline{\boxdot}L(s_2) \tag{3}$$

while it should satisfy reinstatement iff

$$\forall s_2 \in S, \exists s_1 \in S \text{ such that } s_1 \odot s_2 \text{ and } L(s_1)\overline{\boxminus}L(s_2) \tag{4}$$

However, both conditions (3) and (4) are unsatisfactory for *initial*[4] elements of S, *i.e.*, elements s_2 of S such that there are no elements s_1 with $s_1 \odot s_2$. Such elements s_2 trivially satisfy condition (3) and never satisfy condition (4), entailing that no labelling is able to satisfy reinstatement whenever there are initial elements in S.

A suitable condition for initial elements is thus needed.

In this regard, a first option is to impose $\max(C)$ as the unique possible label for initial elements, on the grounds that there are no reasons against the acceptance of initial elements. However, this option looks somehow rigid, since a unique label is prescribed for

[4] We borrow the terminology from abstract argumentation, where initial nodes are those without attackers.

initial nodes, and, from a conceptual point of view, it looks strange that initial elements receive a special treatment which completely neglects the reinstatement violation relation.

Another option is to introduce a special relation for initial elements, which defines the set of their possible labels so that such a set can be tuned in the same way as for the reinstatement violation relation. While this solution would achieve the maximum flexibility, it is still characterized by the same conceptual problem concerning a special treatment for initial elements, which would be completely independent of how labels are selected for non-initial elements.

We are thus led to explore solutions where the set of possible labels for initial nodes is derived from the reinstatement violation relation. In this regard, the following two options for the allowed labels for initial elements can be considered:

1. $\{\lambda \in \Lambda \mid \min(C)\overline{\boxminus}C_\Lambda(\lambda)\}$
2. $\{\lambda \in \Lambda \mid \forall c \in C, c\overline{\boxminus}C_\Lambda(\lambda)\}$

Intuitively, according to the first option, initial elements are equated to non-initial elements where elements intolerant of them are all labelled with minimally positive labels. Accordingly, the labels that the reinstatement violation relation allows for initial elements are the same that are allowed for an element s_2 such that there is a unique element s_1 with $s_1 \odot s_2$, and the label assigned to s_1 is associated to $\min(C)$. In a sense, the absence of reasons against the acceptance of s_2 is equivalent to a contrary reason with a minimal acceptance degree.

The second option allows for initial nodes only those labels that would be allowed by all of the labels of intolerant nodes. The underlying idea is that the absence of reasons against the acceptance of a node s_2, *i.e.*, in case s_2 is an initial element, must only prevent any label for s_2 that would be prevented in case of presence of intolerant elements w.r.t. s_2, whatever the labels assigned to them.

Interestingly enough, the two options turn out to be equivalent if one adopts a well-founded reinstatement violation relation, as the following proposition shows.

Proposition 2.5. *Given a set S equipped with an intolerance relation* int*, a* sac *C equipped with a well-founded reinstatement violation relation* rv*, and a C-classified* sal *Λ, it turns out that*

$$\{\lambda \in \Lambda \mid \min(C)\overline{\boxminus}C_\Lambda(\lambda)\} = \{\lambda \in \Lambda \mid \forall c \in C, c\overline{\boxminus}C_\Lambda(\lambda)\}$$

Proof. Let us first prove the $\subseteq$ relation. Let $\lambda \in \Lambda$ be a label such that $\min(C)\overline{\boxminus}C_\Lambda(\lambda)$. By the definition of minimum, $\forall c \in C, \min(C) \leq c$. If by contradiction $c\overline{\boxdot}C_\Lambda(\lambda)$ then by dual monotonicity of rv it would be the case that $\min(C)\overline{\boxdot}C_\Lambda(\lambda)$, contradicting the hypothesis that $\min(C)\overline{\boxminus}C_\Lambda(\lambda)$.

As to the $\supseteq$ relation, obviously any λ such that $\forall c \in C, c\overline{\boxminus}C_\Lambda(\lambda)$ satisfies as a particular case $\min(C)\overline{\boxminus}C_\Lambda(\lambda)$. □

According to this result, we introduce our generalized notion of reinstatement violation of a labelling as follows.

Definition 2.10. Given a set S equipped with an intolerance relation int, a sac C equipped with a reinstatement violation relation rv, and a C-classified sal Λ, a Λ-labelling L of S is int-rv-*uncompliant* iff

$$\exists s_2 \in S : \begin{cases} \min(C)\overline{\boxdot}C_\Lambda(L(s_2)) & \text{if } s_2 \text{ is initial} \\ \forall s_1 \in S \text{ such that } s_1 \odot s_2 \text{ it holds that } L(s_1)\overline{\boxdot}L(s_2) & \text{otherwise} \end{cases} \tag{5}$$

Conversely, we say that a labelling is int-rv-*compliant* if it is not int-rv-uncompliant, *i.e.*,

$$\forall s_2 \in S \begin{cases} \min(C)\overline{\boxminus}C_\Lambda(L(s_2)) & \text{if } s_2 \text{ is initial} \\ \exists s_1 \in S \text{ such that } s_1 \odot s_2 \text{ and } L(s_1)\overline{\boxminus}L(s_2) & \text{otherwise} \end{cases} \tag{6}$$

A corresponding result w.r.t. Proposition 2.4 holds.

Proposition 2.6. *Given a set S equipped with an intolerance relation* int*, a* sac *C and a C-classified* sal *Λ, consider two reinstatement violation relations* rv *and* rv′ *such that* rv $\subseteq$ rv′*. Then, an* int-rv-*uncompliant Λ-labelling L of S is also* int-rv′-*uncompliant, and an* int-rv′-*compliant Λ-labelling L of S is also* int-rv-*compliant.*

Proof. If L is int-rv-uncompliant, there is an argument α which satisfies one of the two cases for s_2 of condition (5) w.r.t. rv. Since rv $\subseteq$ rv′, obviously this case would be satisfied also adopting rv′. The result concerning compliant labellings follows from the fact that a labelling is int-rv-compliant iff it is not int-rv-uncompliant. □

A final comment can be devoted to the constraints imposed by consistent labellings on the possible labels for initial elements. In particular, according to Definition 2.9 a labelling is int-inc-consistent if $\forall s_2 \in S$ the following condition holds:

$$\forall s_1 : s_1 \odot s_2,\ L(s_1) \underline{\boxminus} L(s_2) \tag{7}$$

If s_2 is initial, then condition (7) is trivially satisfied, *i.e.*, the possible labels for s_2 are unconstrained. However, one may wonder whether a different outcome would be obtained by modifying Definition 2.9 to enforce a specific treatment for initial elements, similarly to int-rv-compliant labellings (see Definition 2.10).

Taking into account the intuition behind consistency, the following two options for the possible labels of initial elements can be considered:

1. $\{\lambda \in \Lambda \mid \min(C) \underline{\boxminus} C_\Lambda(\lambda)\}$
2. $\{\lambda \in \Lambda \mid \exists c \in C : c \underline{\boxminus} C_\Lambda(\lambda)\}$

Similar to the counterpart condition in the case of reinstatement, the first option equates initial elements to non-initial elements where intolerant elements w.r.t. them are all labelled with minimally positive labels.

The second option allows for initial nodes all those labels that would be allowed by at least one label assigned to an intolerant node. The underlying idea is that the absence of simultaneous positivity must leave the maximal freedom in choosing the labels for initial elements, thus any label that can be allowed in case there is an intolerant element must also be allowed for initial elements.

It is easy to see that, in the case of a well-founded incompatibility relation, both options do not enforce any constraint on the possible labels of initial elements. As to the first option, by Proposition 2.2, there is no $c \in C$ such that $\min(C) \underline{\boxdot} c$, entailing that $\forall \lambda \in \Lambda\ \min(C) \underline{\boxminus} C_\Lambda(\lambda)$. Of course, the second option enforces a weaker constraint w.r.t. the first one, as it is evident by considering $c = \min(C)$, thus, it must allow all labels in Λ as the first option.

Summing up, explicitly considering initial elements would not bring any modification to Definition 2.9, which thus conceptually corresponds to the dual counterpart of Definition 2.10. From a theoretical perspective, these considerations support the well-foundedness and generality of both definitions.

Example 2.6. With reference to our running example, an uncompliant labelling can be intuitively regarded as corresponding to a customer who is 'too unopinionated' about travel features. While proposing a travel solution to an inconsistent customer is problematic because of 'too strong desiderata', here the problem consists in 'too weak desiderata', leaving in a sense too open the space of options. Assume the minimal well-founded incompatibility relation for our case, namely $\underline{\mathsf{rv}}_{C^3} = \{(\mathsf{neg}, \mathsf{neg})\}$. Given the intolerance relation int_{tr} previously introduced, we get that L_1 and L_2 are $\mathsf{int}_{tr} - \underline{\mathsf{rv}}_{C^3} - \textit{compliant}$. In particular, in L_1, the label of class neg assigned to Adv is justified by the label of class pos assigned to Lux, while in L_2 the label of class neg assigned to Low is justified by the label of class mid assigned to Lux. Note also that, in both cases, the initial element Exo is not assigned a label of class neg, which would lead to an uncompliance according to Definition 2.10. L_3 is instead $\mathsf{int}_{tr} - \underline{\mathsf{inc}}_{C^3} - \textit{uncompliant}$, since Lux, Low and Adv, all have a label of class neg, these very low assessments not being justified by higher assessments on their intolerant elements.

Also in this case one may consider some stronger requirement, for instance one may consider that an intermediate label is not enough to justify a negative label, which then requires a positive label on an intolerant element. This would correspond to the reinstatement violation relation $\underline{\mathsf{rv}}^+_{C^3} = \{(\mathsf{neg}, \mathsf{neg}), (\mathsf{mid}, \mathsf{neg})\}$.

By Proposition 2.6, L_3 is also $\mathsf{int}_{tr} - \underline{\mathsf{rv}}^+_{C^3} - \textit{uncompliant}$. Moreover, L_2 is $\mathsf{int}_{tr} - \underline{\mathsf{rv}}^+_{C^3} - \textit{uncompliant}$, since the label of class neg assigned to Low is no more justified by the label of class mid assigned to Lux. Intuitively, under this stricter view, the labelling L_2 corresponds to a customer who unjustifiedly disregards saving money, while not being strongly interested in luxury.

As also shown in the examples, the generic definitions of inconsistency and reinstatement violation we have introduced are *tunable* as their instances can be *adjusted* varying the incompatibility and reinstatement violation relations, and possibly also the underlying intolerance relation and C-classification, giving rise to a family of alternative (in)consistency and reinstatement (violation) notions. As an illustration and confirmation of the flexibility and expressiveness of the proposed approach in a more formal setting, in the next section we show that different combinations of (in)consistency and reinstatement (violation) notions are able to capture different argumentation semantics in Dung's argumentation frameworks.

3. Consistency and reinstatement properties in argumentation semantics

In abstract argumentation, a semantics [8] is a formal specification of a criterion to determine the possible outcomes of a situation of conflict, represented by a binary relation of attack (denoted as $\rightarrow$ in the following), defined on a set $\mathcal{A}$ of arguments. A set of arguments and the relevant attack relation are modelled by the traditional notion of argumentation framework [7].

Definition 3.1. An *argumentation framework* is a pair $AF = (\mathcal{A}, \rightarrow)$ where $\mathcal{A}$ is a set of arguments and $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation of attack between them. Given an argument $\alpha \in \mathcal{A}$, we denote as α^- the set $\{\beta \in \mathcal{A} \mid (\beta, \alpha) \in \rightarrow\}$. An argument α such that $\alpha^- = \emptyset$ is called initial.

In the *extension-based* approach to argumentation semantics, the conflict outcomes are expressed as sets of arguments called *extensions* and, in this context, two somewhat dual notions corresponding to those introduced in the paper have been exploited in the relevant definitions. On the one hand, a basic consistency notion called *conflict-freeness* has been traditionally considered: a set of arguments is conflict-free if it does not include any pair of arguments α, β such that $\alpha \in \beta^-$. On the other hand, the *reinstatement*

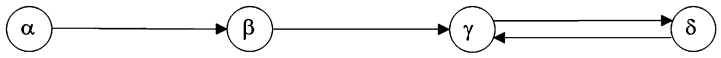


Fig. 3. An example of argumentation framework. Arrows indicate attacks.

criterion, as well as some of its variants, has been made explicit in [10]: a semantics satisfies this criterion if any extension includes all those arguments whose attackers are, in turn, attacked by the extension.

In this paper, we consider the *labelling-based* approach to argumentation semantics. In particular, the outcomes are expressed as argument labellings instead of extensions, *i.e.*, as assignments of labels, taken from a given set, to the set of arguments $\mathcal{A}$. Using the set of three labels Λ^{IOU}, a correspondence can be drawn between extensions and labellings (the reader can refer to [8] for more details), while, in general, the labelling-based approach is more expressive than the extension-based approach, since it can in principle adopt any set of labels (e.g. a set of four labels as in [13] or the $[0,1]$ real interval as in several approaches to gradual argumentation semantics [16,17]) thus yielding assessments which cannot be expressed in terms of extensions.

Combining the generalized notions of consistency and reinstatement with three-valued labellings enables us to identify correspondences between different instances of our generalized notions and different semantics. In particular, given an abstract argumentation framework, we assume that the intolerance relation coincides with the attack relation, *i.e.*, $\alpha \odot \beta$ iff $\alpha \in \beta^-$, and use the classification $C^3_{\Lambda^{\text{IOU}}}$ introduced in Section 2. Then, an analysis of labelling-based semantics in this perspective can be developed, as we do in the following, where we review the definitions of some fundamental labelling-based semantics [8], showing that they can be expressed as combinations of specific instances of generalized consistency and reinstatement properties. Some illustrative examples, concerning the argumentation framework shown in Fig. 3, are provided for the benefit of readers not already familiar with abstract argumentation.

The simplest semantics notion is conflict-freeness, recalled in Definition 3.2.

Definition 3.2. Let L be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. L is *conflict-free* iff for each $\alpha \in \mathcal{A}$ it holds that:

1. if $L(\alpha) = \texttt{in}$ then $\nexists \beta \in \alpha^- : L(\beta) = \texttt{in}$
2. if $L(\alpha) = \texttt{out}$ then $\exists \beta \in \alpha^- : L(\beta) = \texttt{in}$

Example 3.1. With the reference to the argumentation framework in Fig. 3, given the constraint in item 1 of Definition 3.2, in a conflict-free labelling the label `in` can be assigned to the arguments in the following sets: $\emptyset$, $\{\alpha\}$, $\{\beta\}$, $\{\gamma\}$, $\{\delta\}$, $\{\alpha,\gamma\}$, $\{\alpha,\delta\}$, $\{\beta,\delta\}$. In correspondence of each of the above listed sets of arguments labelled `in`, one or more conflict-free labellings can be identified by assigning the label `und` to the arguments not attacked by arguments labelled `in`, and assigning the label `und` or the label `out` to the arguments which are attacked by arguments labelled `in`. For instance, the only conflict-free labelling corresponding to the empty set of `in`-labelled arguments is $L_1 = \{(\alpha,\texttt{und}),(\beta,\texttt{und}),(\gamma,\texttt{und}),(\delta,\texttt{und})\}$. There are instead two conflict-free labellings corresponding to the set of `in`-labelled arguments $\{\beta,\delta\}$, namely $L_2 = \{(\alpha,\texttt{und}),(\beta,\texttt{in}),(\gamma,\texttt{und}),(\delta,\texttt{in})\}$ and $L_3 = \{(\alpha,\texttt{und}),(\beta,\texttt{in}),(\gamma,\texttt{out}),(\delta,\texttt{in})\}$, and four conflict-free labellings corresponding to the set of `in`-labelled arguments $\{\alpha,\delta\}$, namely $L_4 = \{(\alpha,\texttt{in}),(\beta,\texttt{und}),(\gamma,\texttt{und}),(\delta,\texttt{in})\}$, $L_5 = \{(\alpha,\texttt{in}),(\beta,\texttt{out}),(\gamma,\texttt{und}),(\delta,\texttt{in})\}$, $L_6 = \{(\alpha,\texttt{in}),(\beta,\texttt{und}),(\gamma,\texttt{out}),(\delta,\texttt{in})\}$, and $L_7 = \{(\alpha,\texttt{in}),(\beta,\texttt{out}),(\gamma,\texttt{out}),(\delta,\texttt{in})\}$. Continuing the enumeration it can be seen that in total the argumentation framework in Fig. 3 admits 19 conflict-free labellings.

Item 1 in Definition 3.2 corresponds precisely to the weakest form of consistency, *i.e.*, to the incompatibility relation $\underline{\text{inc}}_{C^3} = \{(\text{pos},\text{pos})\}$. The second item represents a requirement for assigning to an argument the least positive label, and corresponds to reinstatement when the following reinstatement violation relation is adopted: $\text{rv}^{cf}_{C^3} = \{(\text{neg},\text{neg}),(\text{mid},\text{neg})\}$.

Proposition 3.1. *Let L be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. Then, L is $\rightarrow$-$\text{rv}^{cf}_{C^3}$-compliant iff for each $\alpha \in \mathcal{A}$ it holds that if $L(\alpha) = \texttt{out}$ then $\exists \beta \in \alpha^- : L(\beta) = \texttt{in}$.*

Proof. Let L be $\rightarrow$-$\text{rv}^{cf}_{C^3}$-compliant and assume by contradiction that there is an argument $\alpha \in \mathcal{A}$ such that $L(\alpha) = \texttt{out}$ and $\forall \beta \in \alpha^- \; L(\beta) \neq \texttt{in}$. If α is initial, according to Definition 2.10 it must be the case that $\min(C) \overline{\boxminus} C_\Lambda(L(\alpha))$, *i.e.*, taking into account the definition of $\text{rv}^{cf}_{C^3}$ it holds that $C_\Lambda(L(\alpha)) \in \{\text{mid},\text{pos}\}$, which contradicts $L(\alpha) = \texttt{out}$. If α is not initial, according to Definition 2.10 it holds that $\exists \beta \in S : \beta \odot \alpha$ (*i.e.*, $\beta \in \alpha^-$) and $L(\beta) \overline{\boxminus} L(\alpha)$, *i.e.*, taking again into account the definition of $\text{rv}^{cf}_{C^3}$ and the fact that $L(\alpha) = \texttt{out}$ we have that $\exists \beta \in S : \beta \in \alpha^-$ and $L(\beta) \in \{\texttt{in}\}$, contradicting the initial assumption that $\forall \beta \in \alpha^- \; L(\beta) \neq \texttt{in}$.

As to the reverse direction, assume that for each $\alpha \in \mathcal{A}$ if $L(\alpha) = \texttt{out}$ then $\exists \beta \in \alpha^- : L(\beta) = \texttt{in}$, and assume by contradiction that L is $\rightarrow$-$\text{rv}^{cf}_{C^3}$-uncompliant. According to Definition 2.10, at least one of the following two cases holds.

1. There is $\alpha \in \mathcal{A}$ such that α is initial and $\min(C) \overline{\boxdot} C_\Lambda(L(\alpha))$. Taking into account the definition of $\text{rv}^{cf}_{C^3}$, this entails $L(\alpha) = \texttt{out}$. By the initial assumption $\exists \beta \in \alpha^- : L(\beta) = \texttt{in}$, contradicting the fact that α is initial.
2. There is a non initial argument $\alpha \in \mathcal{A}$ such that $\forall \beta \in S$ such that $\beta \in \alpha^-$ it holds that $L(\beta) \overline{\boxdot} L(\alpha)$. Taking into account the definition of $\text{rv}^{cf}_{C^3}$, it must be the case that $L(\alpha) = \texttt{out}$ and $\forall \beta \in S$ such that $\beta \in \alpha^-$, $L(\beta) \in \{\texttt{out},\texttt{und}\}$. This contradicts the assumption that $\exists \beta \in \alpha^- : L(\beta) = \texttt{in}$. □

It is then immediate to characterize conflict-free labellings in terms of our generalized notions.

Proposition 3.2. *The set of conflict-free labellings coincides with the set of labellings which are* $\rightarrow$-$\underline{\mathsf{inc}}_{C^3}$*-consistent and* $\rightarrow$-$\mathsf{rv}^{cf}_{C^3}$*-compliant.*

Proof. The proof is immediate taking into account the correspondence between Item 1 in Definition 3.2 and consistency under $\underline{\mathsf{inc}}_{C^3}$, as well as correspondence between Item 2 and reinstatement compliance under $\mathsf{rv}^{cf}_{C^3}$ ensured by Proposition 3.1. □

Admissibility of a set of arguments was introduced in [7] with reference to the notion of defense, *i.e.*, the ability of a conflict-free set to defend its members by counterattacking their attackers. The labelling-based counterpart of this idea is given in Definition 3.3.

Definition 3.3. Let L be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. L is *admissible* iff for each $\alpha \in \mathcal{A}$ it holds that:

1. if $L(\alpha) = \mathtt{in}$ then $\forall \beta \in \alpha^-, L(\beta) = \mathtt{out}$
2. if $L(\alpha) = \mathtt{out}$ then $\exists \beta \in \alpha^- : L(\beta) = \mathtt{in}$

Item 1 in Definition 3.3 is a strengthening of item 1 of Definition 3.2, while item 2 is the same in both Definition 3.2 and 3.3. It follows that admissible labellings are a (tipycally strict) subset of conflict-free labellings.

It can be observed that the labelling assigning und to every argument, besides being conflict-free, is admissible in any argumentation framework.

Example 3.2. Focusing on the example of Fig. 3, it can be noted that the argument β cannot be assigned the label in in any admissible labelling, since its attacker α cannot be assigned the label out. Moreover, it can be noted that the argument γ can be assigned the label in only if both its attackers β and δ are labelled out, with β labelled out implying in turn that α is labelled in. It follows that in an admissible labelling the label in can be assigned to the arguments in the following sets: $\emptyset$, $\{\alpha\}$, $\{\delta\}$, $\{\alpha, \gamma\}$, $\{\alpha, \delta\}$. In correspondence of each of the above listed sets of arguments labelled in, one or more admissible labellings can be identified by assigning first the label out to the arguments attacking arguments labelled in, then assigning the label und to arguments not attacked by arguments labelled in, and finally assigning the label und or the label out to the remaining arguments which are attacked by arguments labelled in. For instance, there are two admissible labellings corresponding to the case where only α is labelled in, namely $L_1 = \{(\alpha, \mathtt{in}), (\beta, \mathtt{und}), (\gamma, \mathtt{und}), (\delta, \mathtt{und})\}$ and $L_2 = \{(\alpha, \mathtt{in}), (\beta, \mathtt{out}), (\gamma, \mathtt{und}), (\delta, \mathtt{und})\}$ and there are two admissible labellings corresponding to the case where α and δ are labelled in, namely $L_3 = \{(\alpha, \mathtt{in}), (\beta, \mathtt{und}), (\gamma, \mathtt{out}), (\delta, \mathtt{in})\}$ and $L_4 = \{(\alpha, \mathtt{in}), (\beta, \mathtt{out}), (\gamma, \mathtt{out}), (\delta, \mathtt{in})\}$. Continuing the enumeration it can be seen that in total the argumentation framework in Fig. 3 has 7 admissible labellings.

Interestingly, the strengthening encompassed by Definition 3.3 corresponds to the choice of a stronger form of consistency: having an attacker labelled und is forbidden for an argument labelled in, while having an attacker labelled in is allowed for an argument labelled und. This coincides with adopting the following asymmetric incompatibility relation $\mathsf{inc}^a_{C^3} = \{(\mathsf{pos}, \mathsf{pos}), (\mathsf{mid}, \mathsf{pos})\}$.

Proposition 3.3. *The set of admissible labellings coincides with the set of labellings which are* $\rightarrow$-$\mathsf{inc}^a_{C^3}$*-consistent and* $\rightarrow$-$\mathsf{rv}^{cf}_{C^3}$*-compliant.*

Proof. We show below that admissible labellings correspond to the set of conflict-free labellings that are $\rightarrow$-$\mathsf{inc}^a_{C^3}$-consistent. Then the conclusion easily follows from Proposition 3.2.

For a labelling L let us first assume that L is admissible. Then L is conflict-free and by item 1 of Definition 3.3 $\nexists \alpha, \beta \in \mathcal{A}$ such that $\beta \in \alpha^-$ (*i.e.*, $\beta \odot \alpha$) and $(L(\beta), L(\alpha)) \in \mathsf{inc}^a_{C^3}$ (*i.e.*, $L(\beta) \underline{\boxdot} L(\alpha)$). Hence L is $\rightarrow$-$\mathsf{inc}^a_{C^3}$-consistent. Let now assume L is conflict-free and $\rightarrow$-$\mathsf{inc}^a_{C^3}$-consistent. To complete the proof we have to show that item 1 of Definition 3.3 holds: assume by contradiction that $\exists \alpha$ such that $L(\alpha) = \mathtt{in}$ and $\exists \beta \in \alpha^- : L(\beta) \neq \mathtt{out}$. It follows that $(L(\beta), L(\alpha)) \in \mathsf{inc}^a_{C^3}$ which contradicts the hypothesis that L is $\rightarrow$-$\mathsf{inc}^a_{C^3}$-consistent. □

Completeness of a set of arguments was introduced in [7] and is based on the idea that if an argument is defended by an admissible set of arguments, it should be accepted together with its defenders. The labelling-based counterpart of this idea is given in Definition 3.4.

Definition 3.4. Let L be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. L is *complete* if it is admissible and for each $\alpha \in \mathcal{A}$ it holds that if $L(\alpha) = \mathtt{und}$ then $\nexists \beta \in \alpha^- : L(\beta) = \mathtt{in}$ and $\exists \beta \in \alpha^- : L(\beta) = \mathtt{und}$.

Example 3.3. Complete labellings are in turn a (typically strict) subset of admissible labellings. First of all, differently from the previous ones, Definition 3.4 implies that unattacked arguments are labelled in (the label und being no more possible since it requires the existence of an attacker). With reference to the example of Fig. 3, α must be labelled in in any complete labelling. Moreover, arguments attacked by an argument labelled in must now be labelled out (the label und being excluded by the last condition in Definition 3.4).

In the example, β must be labelled out. As to the arguments γ and δ it can be seen that, given their mutual attack relation, they can be both labelled und or one in and the other out. In total, the argumentation framework in Fig. 3 admits 3 complete labellings: $L_1 = \{(\alpha, \texttt{in}), (\beta, \texttt{out}), (\gamma, \texttt{und}), (\delta, \texttt{und})\}$, $L_2 = \{(\alpha, \texttt{in}), (\beta, \texttt{out}), (\gamma, \texttt{in}), (\delta, \texttt{out})\}$, and $L_3 = \{(\alpha, \texttt{in}), (\beta, \texttt{out}), (\gamma, \texttt{out}), (\delta, \texttt{in})\}$.

In words, a complete labelling is an admissible labelling with the additional requirement that an argument that is labelled und must have an und-labelled attacker and no in-labelled attackers. In particular, the last condition can be enforced by further strengthening the notion of consistency by means of the incompatibility relation $\mathsf{inc}^c_{C^3} = \{(\mathsf{pos}, \mathsf{pos}), (\mathsf{pos}, \mathsf{mid}), (\mathsf{mid}, \mathsf{pos})\}$, while the first condition is verified if the following reinstatement property is enforced.

Definition 3.5. A labelling L satisfies the *reinstatement property* if $\forall \alpha \in \mathcal{A}$ it holds that if $\forall \beta \in \alpha^-\ L(\beta) = \texttt{out}$ then $L(\alpha) = \texttt{in}$.

The following proposition shows that the reinstatement property can be captured by the reinstatement violation relation $\mathsf{rv}^c_{C^3} = \{(\mathsf{neg}, \mathsf{neg}), (\mathsf{neg}, \mathsf{mid}), (\mathsf{mid}, \mathsf{neg})\}$.

Proposition 3.4. *Let L be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. Then, L is $\rightarrow$-$\mathsf{rv}^c_{C^3}$-compliant iff it satisfies the reinstatement property and for each $\alpha \in \mathcal{A}$ it holds that if $L(\alpha) = \texttt{out}$ then $\exists \beta \in \alpha^- : L(\beta) = \texttt{in}$.*

Proof. Assume that L is $\rightarrow$-$\mathsf{rv}^c_{C^3}$-compliant. By Proposition 2.6, L is $\rightarrow$-$\mathsf{rv}^{cf}_{C^3}$-compliant, and by Proposition 3.1 the second condition of the thesis holds. To prove the reinstatement property, let us consider an argument $\alpha \in \mathcal{A}$ such that $\forall \beta \in \alpha^-$, $L(\beta) = \texttt{out}$, and let us show that $L(\alpha) = \texttt{in}$. If α is initial, by Definition 2.10 it must be the case that $\min(C) \overline{\boxminus} C_\Lambda(L(\alpha))$, *i.e.*, taking into account the definition of $\mathsf{rv}^c_{C^3}$ we have that $C_\Lambda(L(\alpha)) = \mathsf{pos}$, which holds iff $L(\alpha) = \texttt{in}$. If α is non initial, by Definition 2.10 there is an argument $\beta \in \alpha^-$ such that $L(\beta) \overline{\boxminus} L(\alpha)$. Since by the hypothesis $L(\beta) = \texttt{out}$, according to the definition of $\mathsf{rv}^c_{C^3}$ it must be the case that $L(\alpha) = \texttt{in}$.

As to the reverse direction of the proof, assume that L satisfies the reinstatement property and the second condition of the hypothesis, and let us prove that L is $\rightarrow$-$\mathsf{rv}^c_{C^3}$-compliant. According to Definition 2.10, we can consider an argument α and distinguish two cases for it. If α is initial, by the reinstatement property $L(\alpha) = \texttt{in}$, thus $C_\Lambda(L(\alpha)) = \mathsf{pos}$ which satisfies the required condition $\min(C) \overline{\boxminus} C_\Lambda(L(\alpha))$. If α is not initial, referring to Definition 2.10 assume by contradiction that there is no $\beta \in \alpha^-$ such that $L(\beta) \overline{\boxminus} L(\alpha)$. This means that $\forall \beta \in \alpha^-$, $L(\beta) \overline{\boxdot} L(\alpha)$, *i.e.*, $(L(\beta), L(\alpha)) \in \{(\texttt{out}, \texttt{out}), (\texttt{out}, \texttt{und}), (\texttt{und}, \texttt{out})\}$. According to the second condition of the hypothesis if $L(\alpha) = \texttt{out}$ then $\exists \beta \in \alpha^- : L(\beta) = \texttt{in}$, which entails that $L(\alpha) \neq \texttt{out}$. Then $L(\alpha) = \texttt{und}$ and $\forall \beta \in \alpha^-\ L(\beta) = \texttt{out}$, contradicting the reinstatement property. □

We can now characterize complete labellings in terms of generalized consistency and reinstatement.

Proposition 3.5. *The set of complete labellings coincides with the set of admissible labellings which are $\rightarrow$-$\mathsf{inc}^c_{C^3}$-consistent and satisfy the reinstatement property.*

Proof. For a labelling L let us first assume that L is complete, hence admissible. From Proposition 3.3 we have that $\nexists \alpha, \beta$ such that $\beta \in \alpha^-$ and $(L(\beta), L(\alpha)) \in \{(\mathsf{pos}, \mathsf{pos}), (\mathsf{mid}, \mathsf{pos})\}$. From Definition 3.4 we have also that if $L(\alpha) = \texttt{und}$ then $\nexists \beta \in \alpha^- : L(\beta) = \texttt{in}$, *i.e.*, $\nexists \alpha, \beta$ such that $\beta \in \alpha^-$ and $(L(\beta), L(\alpha)) \in \{(\mathsf{pos}, \mathsf{mid})\}$. It follows that L is $\rightarrow$-$\mathsf{inc}^c_{C^3}$-consistent. Moreover, it is well known that complete labellings satisfy the reinstatement property [8].

Let us now assume L is admissible, $\rightarrow$-$\mathsf{inc}^c_{C^3}$-consistent and satisfies the reinstatement property. Given an argument α such that $L(\alpha) = \texttt{und}$ it follows (from consistency) that $\nexists \beta \in \alpha^- : L(\beta) = \texttt{in}$ and (from the reinstatement property) that $\exists \beta \in \alpha^- : L(\beta) \neq \texttt{out}$, hence $\exists \beta \in \alpha^- : L(\beta) = \texttt{und}$ and L is a complete labelling. □

Proposition 3.6. *The set of complete labellings coincides with the set of labellings which are $\rightarrow$-$\mathsf{inc}^c_{C^3}$-consistent and $\rightarrow$-$\mathsf{rv}^c_{C^3}$-compliant.*

Proof. It is shown in Proposition 3.5 that complete labellings coincide with admissible labellings that are $\rightarrow$-$\mathsf{inc}^c_{C^3}$-consistent and satisfy the reinstatement property. Then, according to the definition of admissible labellings if a labelling is complete it satisfies the condition that if $L(\alpha) = \texttt{out}$ then $\exists \beta \in \alpha^- : L(\beta) = \texttt{in}$, entailing by Proposition 3.4 that it is $\rightarrow$-$\mathsf{rv}^c_{C^3}$-compliant. Conversely, if a labelling is $\rightarrow$-$\mathsf{inc}^c_{C^3}$-consistent and $\rightarrow$-$\mathsf{rv}^c_{C^3}$-compliant then by Proposition 2.4 and Proposition 2.6 it is also $\rightarrow$-$\mathsf{inc}^a_{C^3}$-consistent and $\rightarrow$-$\mathsf{rv}^{cf}_{C^3}$-compliant, and thus admissible by Proposition 3.3. Moreover, by Proposition 3.4 it satisfies the reinstatement property. As a consequence, taking into account the aforementioned result of Proposition 3.5 the labelling is complete. □

The notion of a stable set of arguments can be characterized in several ways, its key feature being that no room is left for undecidedness (an argument is either accepted or attacked by an accepted argument) as indicated by Definition 3.6.

Definition 3.6. Let L be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. L is *stable* if it is complete and $\nexists \alpha \in \mathcal{A} : L(\alpha) = \texttt{und}$.

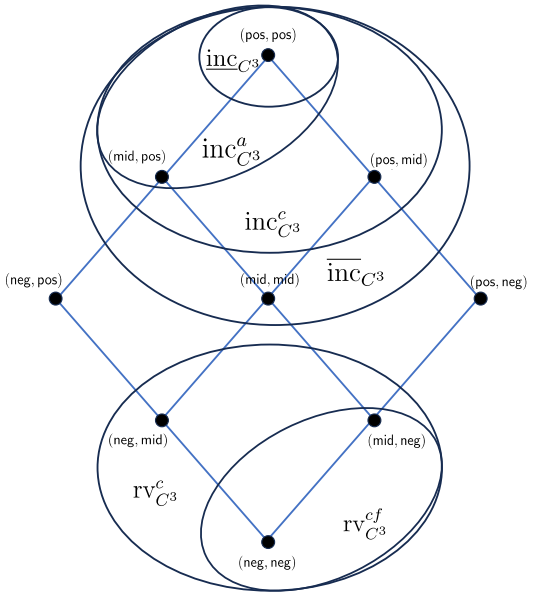


Fig. 4. The incompatibility and reinstatement violation relations introduced in Section 3.

Example 3.4. Of the three complete labellings listed in Example 3.3 clearly L_2 and L_3 are stable, while L_1 is not, having some arguments labelled und.

This constraint can be put in correspondence with the adoption of the strongest notion of consistency, namely with the choice of the incompatibility relation $\overline{\text{inc}}_{C^3} = \{(\text{pos}, \text{pos}), (\text{pos}, \text{mid}), (\text{mid}, \text{pos}), (\text{mid}, \text{mid})\}$.

Proposition 3.7. *The set of stable labellings coincides with the set of labellings that are* $\rightarrow$-$\overline{\text{inc}}_{C^3}$*-consistent and* $\rightarrow$-$\text{rv}^c_{C^3}$*-compliant.*

Proof. We show below that stable labellings coincide with complete labellings which are $\rightarrow$-$\overline{\text{inc}}_{C^3}$-consistent. Then the conclusion follows from Proposition 3.6.

For a labelling L let us first assume that L is stable. By Definition 3.6, L is complete. Moroever, by the same definition no argument is labelled und hence $\nexists \alpha, \beta$ such that $\beta \in \alpha^-$ and $(L(\beta), L(\alpha)) \in \{(\text{pos}, \text{mid}), (\text{mid}, \underline{\text{pos}}), (\text{mid}, \text{mid})\}$ and from conflict-freeness we have also that $\nexists \alpha, \beta$ such that $\beta \in \alpha^-$ and $(L(\beta), L(\alpha)) = (\text{pos}, \text{pos})$. Therefore L is $\rightarrow$-$\overline{\text{inc}}_{C^3}$-consistent.

Assume now that L is complete and $\rightarrow$-$\overline{\text{inc}}_{C^3}$-consistent and suppose by contradiction that $\exists \alpha$ such that $L(\alpha) = \texttt{und}$. It follows that $\alpha^- \neq \emptyset$, otherwise by the reinstatement property it would hold that $L(\alpha) = \texttt{in}$. For every $\beta \in \alpha^-$ we have that $L(\beta) \notin \{\texttt{in}, \texttt{und}\}$ otherwise L would not be $\rightarrow$-$\overline{\text{inc}}_{C^3}$-consistent. But then $\forall \beta \in \alpha^-$ we get $L(\beta) = \texttt{out}$ which, by the reinstatement property, contradicts $L(\alpha) = \texttt{und}$. □

The incompatibility and reinstatement violation relations introduced in this section are depicted in Fig. 4 with reference to the Hasse diagram of the elements of $C^3 \times C^3$. In particular, in the diagram the total order in C^3 is extended to a partial order in $C^3 \times C^3$ by considering $(c_1, c_2) \leq (c'_1, c'_2)$ iff $c_1 \leq c'_1$ and $c_2 \leq c'_2$. It is easy to check that each incompatibility (reinstatement violation) relation satisfies the conditions of Definition 2.7 (Definition 2.8), *i.e.*, all relations are well-founded.

To summarize, conflict-free labellings can be characterized in terms of generalized consistency and reinstatement, admissible labellings can be characterized in terms of strengthening consistency with respect to conflict-freeness without resorting to the traditional notion of defense, while characterizations of complete and stable labellings are achieved by further strengthening generalized consistency and reinstatement.

To complete the picture, we recall that the traditional *grounded* and *preferred* semantics [7] as well as the *semi-stable* semantics [9] correspond to applying minimality or maximality constraints to the set of the outcomes of other semantics. Hence, our characterization also carries over to these semantics by applying the same global constraints. In particular, as shown in [9], the (provably unique) grounded labelling corresponds to the complete labelling where the set of arguments labelled in is minimal with respect to set inclusion, while the preferred labellings correspond to the complete labellings where the set of arguments labelled in is maximal, and the semi-stable labellings to the complete labellings where the set of arguments labelled und is minimal.

In the next section, we move beyond the evaluation of acceptability of arguments based on argumentation semantics and consider further evaluations that can be derived from it and raise the issue of preserving consistency across the derivation.

4. Consistency and reinstatement preservation in labelling derivation mechanisms

The production of a set of labellings, in general, is just the first step of a reasoning process, where the set of labellings is used as the starting point for the derivation of further evaluations.

In formal argumentation, the outcomes prescribed by argumentation semantics are typically used as a basis for subsequent evaluations, in particular concerning the justification status of arguments.

It is then interesting to consider whether and how the consistency and reinstatement properties of the original evaluation are preserved in the derived evaluation and the requirements that can be posed on the derivation mechanism to ensure this preservation.

We focus here on what we call *pure synthesis labelling derivation*, namely a mechanism where a labelling of a set S is generated from a set of labellings of the same set S, deriving the label of each element from the labels assigned to the same element by the original labellings.

As an instance of this general pattern, we refer to the evaluation of the argument justification status according to a given semantics, which is derived from the set of argument extensions/labellings prescribed by the same semantics.

The simplest notion of argument justification, which we will use as a running example, is based on three possible states.

Definition 4.1. Given a set $\mathcal{L}$ of Λ^{IOU}-labellings of a set of arguments $\mathcal{A}$, an argument $\alpha \in \mathcal{A}$ is:

- *skeptically justified* (shortly, SkJ) iff $\forall L \in \mathcal{L}\ L(\alpha) = \texttt{in}$;
- *credulously justified* (CrJ) iff it is not skeptically justified[5] and $\exists L \in \mathcal{L} : L(\alpha) = \texttt{in}$;
- *not justified* (NoJ) iff $\nexists L \in \mathcal{L} : L(\alpha) = \texttt{in}$

Example 4.1. Considering the set of complete labellings $\mathcal{L}_c = \{L_1, L_2, L_3\}$ introduced in Example 3.3, argument α is skeptically justified, argument β is not justified, arguments γ and δ are credulously justified.

Considering a sal $\Lambda^{\text{AJ}} = \{\text{SkJ}, \text{CrJ}, \text{NoJ}\}$, the evaluation of argument justification can be modelled as the generation of a Λ^{AJ}-labelling from a set of Λ^{IOU}-labellings. Concerning Λ^{AJ} it is intuitive to assume the classification $C^3_{\Lambda^{\text{AJ}}} = \{(\text{SkJ}, \text{pos}), (\text{NoJ}, \text{neg}), (\text{CrJ}, \text{mid})\}$.

Example 4.2. The argument justification labelling corresponding to Example 3.3 is $(\alpha, \text{SkJ}), (\beta, \text{NoJ}), (\gamma, \text{CrJ}), (\delta, \text{CrJ})$.

At a general level, pure synthesis labelling derivations, like the one of argument justification, can be formalized through a simple synthesis function.

Definition 4.2. Given two sals Λ_1 and Λ_2, a *simple synthesis function* (ssf) from Λ_1 to Λ_2 is a mapping $\text{syn} : 2^{\Lambda_1} \setminus \{\emptyset\} \to \Lambda_2$.

The idea is that, given a set of Λ_1-labellings of a set S, a Λ_2-labelling of S can be derived by applying a ssf to the set of labels relevant to each element of S.

Definition 4.3. Let S be a set, Λ_1 and Λ_2 two sals, syn a ssf from Λ_1 to Λ_2, and $\mathcal{L}_1$ a non-empty set of Λ_1-labellings of S. *The Λ_2-labelling derived from $\mathcal{L}_1$ through* syn, denoted as $DL^{\text{syn}}_{\mathcal{L}_1}$, is defined, for every $s \in S$ as:

$$DL^{\text{syn}}_{\mathcal{L}_1}(s) = \text{syn}(\mathcal{L}_1^{\downarrow}(s))$$

where for any sal Λ, any set $\mathcal{L}$ of Λ-labellings of S and any $s \in S$, $\mathcal{L}^{\downarrow}(s) \triangleq \{L(s) \mid L \in \mathcal{L}\}$.

Example 4.3. Referring again to Example 3.3, we have that $\mathcal{L}_c^{\downarrow}(\alpha) = \{\texttt{in}\}$, $\mathcal{L}_c^{\downarrow}(\beta) = \{\texttt{out}\}$, $\mathcal{L}_c^{\downarrow}(\gamma) = \mathcal{L}_c^{\downarrow}(\delta) = \{\texttt{in}, \texttt{und}, \texttt{out}\}$. Then the idea is that the justification labelling presented in Example 4.2 (and, in general, any justification labelling) can be obtained by applying a proper ssf syn to the set of labels relevant to each argument. For instance since α is skeptically justified, it must be the case that $\text{syn}(\{\texttt{in}\}) = \text{SkJ}$. A proper ssf for argument justification is directly identifiable, as discussed next.

It is rather easy to see that the argument justification evaluation of Definition 4.1 corresponds to the use of a ssf syn_{AJ} from Λ^{IOU} to Λ^{AJ} defined, for every $\Lambda \subseteq \Lambda^{\text{IOU}}$ as follows:

- $\text{syn}_{\text{AJ}}(\Lambda) = \text{SkJ}$ if $\Lambda = \{\texttt{in}\}$;
- $\text{syn}_{\text{AJ}}(\Lambda) = \text{CrJ}$ if $\Lambda \supsetneq \{\texttt{in}\}$;
- $\text{syn}_{\text{AJ}}(\Lambda) = \text{NoJ}$ otherwise.

[5] While traditionally credulous justification is regarded as including skeptical justification, we enforce this distinction so that argument justification can be properly modelled as a labelling.

Assuming that the labellings used for derivation satisfy some consistency and reinstatement properties, preserving these properties in the derived labelling appears desirable. We address this issue in a domain-independent way by defining general preservation properties for ssfs.

Definition 4.4. Let C be a sac equipped with an incompatibility relation inc, and Λ_1 and Λ_2 be two C-classifed sets of labels. A ssf syn from Λ_1 to Λ_2 is *consistency preserving* according to[6] inc iff for any set S equipped with an intolerance relation int and any non-empty int-inc-consistent set[7] $\mathcal{L}_1$ of Λ_1-labellings of S it holds that the labelling $DL^{\text{syn}}_{\mathcal{L}_1}$ is int-inc-consistent.

Definition 4.5. Let C be a sac equipped with a reinstatement violation relation rv, and Λ_1 and Λ_2 be two C-classifed sets of labels. A ssf syn from Λ_1 to Λ_2 is *reinstatement preserving* according to[8] rv iff for any set S equipped with an intolerance relation int and any non-empty int-rv-compliant set $\mathcal{L}_1$ of Λ_1-labellings of S it holds that the labelling $DL^{\text{syn}}_{\mathcal{L}_1}$ is int-rv-compliant.

This raises the issue of analyzing at a general level some properties of the ssf that can ensure consistency and reinstatement preservation.

To start, we introduce a notion of well-behaved ssf, which intuitively means that the function is monotonic with respect to an ordering of sets of labels, induced by the ordering on labels introduced in Definition 2.2, as specified in the next definition.

Definition 4.6. Let C be a sac, and Λ a C-classifed sal. Given $\Lambda_1, \Lambda_2 \subseteq \Lambda$, we say that Λ_2 is *at least as positive* as Λ_1, denoted as $\Lambda_1 \preceq_P \Lambda_2$, iff $\forall \lambda \in \Lambda_1\ \exists \lambda' \in \Lambda_2$ such that $\lambda \preceq_\Lambda \lambda'$ and $\forall \lambda' \in \Lambda_2\ \exists \lambda \in \Lambda_1$ such that $\lambda \preceq_\Lambda \lambda'$.

The idea of the $\preceq_P$ relation is that every element of Λ_1 can be mapped into an at least as positive element of Λ_2 and, at the same time, every element of Λ_2 can be mapped into a no more positive element of Λ_1. To exemplify, for every non-empty $\Lambda \subseteq \Lambda^{\text{IOU}}$ it holds that $\Lambda \preceq_P \{\texttt{in}\}$ and $\{\texttt{out}\} \preceq_P \Lambda$. Also $\{\texttt{in}, \texttt{out}\} \preceq_P \{\texttt{in}, \texttt{und}, \texttt{out}\}$ and $\{\texttt{in}, \texttt{und}, \texttt{out}\} \preceq_P \{\texttt{in}, \texttt{out}\}$ while $\{\texttt{in}, \texttt{out}\} \not\preceq_P \{\texttt{und}\}$ and $\{\texttt{und}\} \not\preceq_P \{\texttt{in}, \texttt{out}\}$.

It is easy to see that $\preceq_P$ is a preorder.

Proposition 4.1. *For any* sac C *and* C*-classifed* sal Λ*, the relation* $\preceq_P$ *is reflexive and transitive.*

Proof. By Proposition 2.1, the relation $\preceq_\Lambda$ is reflexive and transitive. The conclusion then easily follows from Definition 4.6. □

We can now introduce the notion of well-behaved ssf.

Definition 4.7. Given a sac C and two C-classifed sals Λ_1 and Λ_2, a ssf syn from Λ_1 to Λ_2 is *well-behaved* iff for any non-empty $\Lambda'_1, \Lambda'_2 \subseteq \Lambda_1$ such that $\Lambda'_1 \preceq_P \Lambda'_2$, it holds that $\text{syn}(\Lambda'_1) \preceq \text{syn}(\Lambda'_2)$.

Using these basic concepts, we will investigate the problems of ensuring consistency and reinstatement preservation.

4.1. Characterising consistency preserving synthesis functions

Towards characterizing well-behaved and consistency preserving ssfs, we assume that for every label $\lambda \in \Lambda$ the set $cc(\lambda)$, which is non-empty by the third condition of Definition 2.7, includes a non-empty set of maximal elements with respect to the $\preceq$ relation, *i.e.*, $\forall \lambda \in \Lambda$ the set $\{\lambda' \in cc(\lambda) \mid \nexists \lambda'' \in cc(\lambda) \text{ with } \lambda' \prec \lambda''\}$ is non-empty. This is clearly true if Λ (and hence $cc(\lambda)$) is finite, while it corresponds to a relatively mild requirement if Λ is infinite (and hence $cc(\lambda)$ can be infinite too). Investigating contexts where this assumption is not satisfied is left to future work.

As a first step, given a set of labels Λ_1, we consider whether a set of labels Λ_2 is a compatible dual of Λ_1, with respect to an inconsistency relation. This means that, given an int-inc-consistent set of labellings $\mathcal{L}$, if $\Lambda_1 = \mathcal{L}^{\downarrow}(s)$ for some element s, then it is possible that $\Lambda_2 = \mathcal{L}^{\downarrow}(s')$ for some s' such that $s \odot s'$.

Definition 4.8. Let C be a sac, Λ be a C-classifed set of labels, and $\Lambda_1 \subseteq \Lambda$. Given an incompatibility relation inc on C, we say that $\Lambda_2 \subseteq \Lambda$ is an inc-*compatible dual* of Λ_1, denoted as $\Lambda_2 \in CCD(\Lambda_1)$, iff $\forall \lambda \in \Lambda_1\ \exists \lambda' \in \Lambda_2$ such that $\lambda' \in cc(\lambda)$, and $\forall \lambda' \in \Lambda_2\ \exists \lambda \in \Lambda_1$ such that $\lambda' \in cc(\lambda)$.

[6] More precisely, consistency preservation should be defined w.r.t. a tuple $(C, C_{\Lambda_1}, C_{\Lambda_2}, \text{inc})$, since it also depends on how the labels of Λ_1 and Λ_2 are mapped on assessment classes. However, for ease of notation, we focus on the incompatibility relation inc, since the mappings C_{Λ_1} and C_{Λ_2} are usually clear from the context.

[7] With a little abuse of language we say that a set of labellings is int-inc-consistent if all its elements are int-inc-consistent. We will say that a set is int-rv-compliant with the analogous meaning.

[8] Similarly to the case of consistency preservation, reinstatement preservation should be defined w.r.t. $(C, C_{\Lambda_1}, C_{\Lambda_2}, \text{rv})$.

Example 4.4. Let us consider Λ^{IOU} and the incompatibility relation $\text{inc}^c_{C^3} = \{(\text{pos}, \text{pos}), (\text{pos}, \text{mid}), (\text{mid}, \text{pos})\}$. Letting $\Lambda_1 = \{\texttt{in}, \texttt{und}\}$ we have that $cc(\texttt{in}) = \{\texttt{out}\}$ and $cc(\texttt{und}) = \{\texttt{und}, \texttt{out}\}$. Hence $CCD(\Lambda_1) = \{\{\texttt{out}\}, \{\texttt{und}, \texttt{out}\}\}$: in fact, out must be included in every element of $CCD(\Lambda_1)$ since it is the only compatible label with in, while both out and und are compatible with und. Considering the weaker incompatibility relation $\text{inc}^a_{C^3} = \{(\text{pos}, \text{pos}), (\text{mid}, \text{pos})\}$ we would have $cc(\texttt{in}) = cc(\texttt{und}) = \{\texttt{und}, \texttt{out}\}$, and hence $CCD(\Lambda_1) = \{\{\texttt{out}\}, \{\texttt{und}\}, \{\texttt{und}, \texttt{out}\}\}$.

The following proposition confirms the intended meaning of Definition 4.8.

Proposition 4.2. *Let C be a* sac*, Λ be a C-classifed set of labels, and* inc *an incompatibility relation on C. For any set S equipped with an intolerance relation* int*, any* int-inc*-consistent set $\mathcal{L}$ of Λ-labellings of S, and any $s_1, s_2 \in S$ such that $(s_1, s_2) \in$ int*, it holds that $\mathcal{L}^{\downarrow}(s_2) \in CCD(\mathcal{L}^{\downarrow}(s_1))$.*

Proof. Under the hypothesis that every $L \in \mathcal{L}$ is int-inc-consistent it holds that $L(s_2) \in cc(L(s_1))$, hence $\forall \lambda \in \mathcal{L}^{\downarrow}(s_1)\ \exists \lambda' \in \mathcal{L}^{\downarrow}(s_2)$ such that $\lambda' \in cc(\lambda)$ and also $\forall \lambda' \in \mathcal{L}^{\downarrow}(s_2)\ \exists \lambda \in \mathcal{L}^{\downarrow}(s_1)$ such that $\lambda' \in cc(\lambda)$, hence $\mathcal{L}^{\downarrow}(s_2) \in CCD(\mathcal{L}^{\downarrow}(s_1))$. □

We need then to consider the 'extreme cases' of compatible dual of a finite set of labels with respect to an inconsistency relation.

Definition 4.9. Let C be a sac, Λ a C-classifed set of labels, and inc an incompatibility relation on C. Given $\Lambda_1 \subseteq \Lambda$ we define $MCCD(\Lambda_1) \triangleq \bigcup_{\lambda \in \Lambda_1} \widehat{MCC}(\lambda)$, where $\widehat{MCC}(\lambda) \triangleq \{\lambda' \in cc(\lambda) \mid \nexists \lambda'' \in cc(\lambda) : \lambda' \prec \lambda''\}$.

Example 4.5. Continuing Example 4.4, with reference to $\text{inc}^c_{C^3}$, we have $\widehat{MCC}(\texttt{in}) = \{\texttt{out}\}$, $\widehat{MCC}(\texttt{und}) = \{\texttt{und}\}$ hence $MCCD(\Lambda_1) = \{\texttt{out}, \texttt{und}\}$. Referring instead to $\text{inc}^a_{C^3}$ we get $\widehat{MCC}(\texttt{in}) = \widehat{MCC}(\texttt{und}) = \{\texttt{und}\}$, hence $MCCD(\Lambda_1) = \{\texttt{und}\}$.

Note that if $\Lambda_1 \neq \emptyset$ then non emptiness of $MCCD(\Lambda_1)$ follows from the assumption that for every λ the set $cc(\lambda)$ includes some maximal elements.

The following propositions provide two interesting properties of $MCCD(\Lambda_1)$: it belongs to $CCD(\Lambda_1)$ and is maximal with respect to $\preceq_P$.

Proposition 4.3. *Let C be a* sac*, Λ a C-classifed set of labels, and* inc *an incompatibility relation on C. Then, for every non-empty $\Lambda_1 \subseteq \Lambda$ it holds that $MCCD(\Lambda_1) \in CCD(\Lambda_1)$.*

Proof. From the definition of $MCCD(\Lambda_1)$ and the above mentioned assumption it is immediate to see that for every $\lambda \in \Lambda_1\ \exists \lambda' \in MCCD(\Lambda_1)$ such that $\lambda' \in cc(\lambda)$ and that for every $\lambda' \in MCCD(\Lambda_1)\ \exists \lambda \in \Lambda_1$ such that $\lambda' \in cc(\lambda)$. □

Proposition 4.4. *Let C be a* sac*, Λ a C-classified set of labels, and* inc *an incompatibility relation on C. Given $\Lambda_1 \subseteq \Lambda$ with $\Lambda_1 \neq \emptyset$, it holds that $\forall D \in CCD(\Lambda_1)$, $D \preceq_P MCCD(\Lambda_1)$.*

Proof. For any $\lambda' \in D$, from Definition 4.8 it holds that $\exists \lambda \in \Lambda_1$ such that $\lambda' \in cc(\lambda)$. Then, by Definition 4.9 $\exists \lambda'' \in MCCD(\Lambda_1)$ such that $\lambda'' \in cc(\lambda)$ and $\nexists \lambda''' \in cc(\lambda) : \lambda'' \prec \lambda'''$, which implies that $\lambda' \preceq \lambda''$ given that, by Proposition 2.1, $\preceq$ is total.

Consider now any $\lambda'' \in MCCD(\Lambda_1)$. By Definition 4.9 it holds that $\exists \lambda \in \Lambda_1$ such that $\lambda'' \in cc(\lambda)$. Moreover by Definition 4.8 $\exists \lambda' \in D$ such that $\lambda' \in cc(\lambda)$. Now by Definition 4.9 we have again that $\nexists \lambda''' \in cc(\lambda) : \lambda'' \prec \lambda'''$ and hence, taking again into account that $\preceq$ is a total preorder, $\lambda' \preceq \lambda''$. □

On this basis, we can now first derive a necessary and sufficient condition for consistency preservation by a well-behaved ssf.

Proposition 4.5. *Let C be a* sac*, Λ_1 and Λ_2 two C-classified* sals*, and* inc *a well-founded incompatibility relation on C. A well-behaved* ssf syn *from Λ_1 to Λ_2 is consistency preserving if and only if for every non-empty set $\Lambda'_1 \subseteq \Lambda_1$ it holds that $\text{syn}(MCCD(\Lambda'_1)) \in cc(\text{syn}(\Lambda'_1))$.*

Proof. Let syn be a ssf satisfying the hypotheses and assume by contradiction that syn is not consistency preserving. This means that there are two elements $s_1, s_2 \in S$ such that $s_1 \odot s_2$ and an int-inc-consistent set $\mathcal{L}$ of Λ-labellings of S such that $DL^{\text{syn}}_{\mathcal{L}}(s_1) \underline{\boxdot} DL^{\text{syn}}_{\mathcal{L}}(s_2)$. Now $DL^{\text{syn}}_{\mathcal{L}}(s_1) = \text{syn}(\mathcal{L}^{\downarrow}(s_1))$ and similarly $DL^{\text{syn}}_{\mathcal{L}}(s_2) = \text{syn}(\mathcal{L}^{\downarrow}(s_2))$. Let $\Lambda'_1 = \mathcal{L}^{\downarrow}(s_1)$. From Proposition 4.2 we have $\mathcal{L}^{\downarrow}(s_2) \in CCD(\Lambda'_1)$ and hence from Proposition 4.4 $\mathcal{L}^{\downarrow}(s_2) \preceq_P MCCD(\Lambda'_1)$. Since syn is well-behaved $\text{syn}(\mathcal{L}^{\downarrow}(s_2)) \preceq \text{syn}(MCCD(\Lambda'_1))$. By the hypothesis, $\text{syn}(MCCD(\Lambda'_1)) \in cc(\text{syn}(\Lambda'_1))$, *i.e.*, $\text{syn}(\Lambda'_1) \underline{\boxminus} \text{syn}(MCCD(\Lambda'_1))$. Since $\text{syn}(\mathcal{L}^{\downarrow}(s_2)) \preceq \text{syn}(MCCD(\Lambda'_1))$ and inc is well-founded, according to Definition 2.7 (point 1) $\text{syn}(\Lambda'_1) \underline{\boxminus} \text{syn}(\mathcal{L}^{\downarrow}(s_2))$, contradicting $\text{syn}(\mathcal{L}^{\downarrow}(s_1)) \underline{\boxdot} \text{syn}(\mathcal{L}^{\downarrow}(s_2))$.

As to the other direction of the proof, assume now that syn is consistency preserving according to inc. Consider a set $S = \{s_1, s_2\}$ with an intolerance relation such that $s_1 \odot s_2$. Since by Proposition 4.3 for every set $\Lambda'_1 \subseteq \Lambda_1$ it holds that $MCCD(\Lambda'_1) \in CCD(\Lambda'_1)$, we can identify a consistent set $\mathcal{L}$ of Λ_1-labellings of S such that $\mathcal{L}^{\downarrow}(s_1) = \Lambda'_1$ and $\mathcal{L}^{\downarrow}(s_2) = MCCD(\Lambda'_1)$. Then by consistency preservation it must also hold that $\text{syn}(MCCD(\Lambda'_1)) \in cc(\text{syn}(\Lambda'_1))$. □

Table 1
Illustration of the proof of Proposition 4.7.

Λ_1 $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1)$	$\underline{\mathsf{inc}}_{C^3}$	$\mathsf{inc}^a_{C^3}$	$\mathsf{inc}^c_{C^3}$
{in} SkJ	{und} NoJ	{und} NoJ	{out} NoJ
{out} NoJ	{in} SkJ	{in} SkJ	{in} SkJ
{und} NoJ	{in} SkJ	{und} NoJ	{und} NoJ
{in, out} CrJ	{in, und} CrJ	{in, und} CrJ	{in, out} CrJ
{in, und} CrJ	{in, und} CrJ	{und} NoJ	{und, out} NoJ
{und, out} NoJ	{in} SkJ	{in, und} CrJ	{in, und} CrJ
{in, und, out} CrJ	{in, und} CrJ	{in, und} CrJ	{in, und, out} CrJ

As an example of application of the above result, we show that the function $\mathsf{syn}_{\mathrm{AJ}}$ is consistency preserving according to the incompatibility relations $\underline{\mathsf{inc}}_{C^3}$, $\mathsf{inc}^a_{C^3}$, and $\mathsf{inc}^c_{C^3}$ while it is not according to $\overline{\mathsf{inc}}_{C^3}$.

First of all, we need to show that $\mathsf{syn}_{\mathrm{AJ}}$ is well-behaved.

Proposition 4.6. *The* ssf syn_{AJ} *is well-behaved.*

Proof. We have to prove that for any two non-empty sets $\Lambda_1, \Lambda_2 \subseteq \Lambda^{\mathrm{IOU}}$ such that $\Lambda_1 \preceq_P \Lambda_2$, $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1) \preceq \mathsf{syn}_{\mathrm{AJ}}(\Lambda_2)$. Since the strict order $\prec$ induced on Λ^{AJ} is total, it is sufficient to show that for any two non-empty sets $\Lambda_1, \Lambda_2 \subseteq \Lambda^{\mathrm{IOU}}$, whenever $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_2) \prec \mathsf{syn}_{\mathrm{AJ}}(\Lambda_1)$ it does not hold that $\Lambda_1 \preceq_P \Lambda_2$. First, consider the case that $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1) = \mathsf{SkJ}$, thus $\Lambda_1 = \{\mathtt{in}\}$. It is easy to see that $\{\mathtt{in}\} \not\preceq_P \Lambda_2$ for any $\Lambda_2 \subseteq \Lambda^{\mathrm{IOU}}$, with $\Lambda_2 \notin \{\emptyset, \{\mathtt{in}\}\}$, which is impossible due to $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_2) \prec \mathsf{SkJ}$. Consider then the case $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1) = \mathsf{CrJ}$, thus $\{\mathtt{in}\} \subsetneq \Lambda_1$ and $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_2) = \mathsf{NoJ}$ entailing $\mathtt{in} \notin \Lambda_2$. It is then easy to see that for any non-empty set Λ_2 such that $\mathtt{in} \notin \Lambda_2$ and any set Λ_1 such that $\{\mathtt{in}\} \subsetneq \Lambda_1$, $\Lambda_1 \not\preceq_P \Lambda_2$. Finally, if $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1) = \mathsf{NoJ}$ then the thesis trivially holds, since there is no label strictly lower than NoJ. □

Proposition 4.7. *The* ssf syn_{AJ} *is consistency preserving according to the incompatibility relations* $\underline{\mathsf{inc}}_{C^3}$, $\mathsf{inc}^a_{C^3}$, *and* $\mathsf{inc}^c_{C^3}$ *while it is not according to* $\overline{\mathsf{inc}}_{C^3}$.

Proof. We need to show that for every non-empty set $\Lambda_1 \subseteq \Lambda^{\mathrm{IOU}}$ it holds that $\mathsf{syn}_{\mathrm{AJ}}(MCCD(\Lambda_1)) \in cc(\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1))$. For $\underline{\mathsf{inc}}_{C^3}$, $\mathsf{inc}^a_{C^3}$, and $\mathsf{inc}^c_{C^3}$ this is illustrated in Table 1, where the first column presents the various possible cases for Λ_1 with the relevant value of $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1)$ and the following columns (illustrating $\underline{\mathsf{inc}}_{C^3}$, $\mathsf{inc}^a_{C^3}$, and $\mathsf{inc}^c_{C^3}$ respectively) show the corresponding $MCCD(\Lambda_1)$ and the relevant value of $\mathsf{syn}_{\mathrm{AJ}}(MCCD(\Lambda_1))$. By inspection, it can be checked that, as desired, for every pair $(\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1), \mathsf{syn}_{\mathrm{AJ}}(MCCD(\Lambda_1)))$ obtained by taking the first element from a row of the first column, and the second element from any other cell (say the i-th with $i \in \{2, 3, 4\}$) of the same row it holds that $(\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1), \mathsf{syn}_{\mathrm{AJ}}(MCCD(\Lambda_1))) \notin \mathsf{inc}'$ where inc' is the incompatibility relation induced by the inc relation specified at the top of the i-th column (from which the second element of the pair was taken). For instance, considering the fifth row, with $\Lambda_1 = \{\mathtt{in}, \mathtt{out}\}$ and $(\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1)) = \mathsf{CrJ}$ and its second cell where (according to $\underline{\mathsf{inc}}_{C^3}$) $MCCD(\Lambda_1) = \{\mathtt{in}, \mathtt{und}\}$ we have $(\mathsf{syn}_{\mathrm{AJ}}(MCCD(\Lambda_1))) = \mathsf{CrJ}$ and then $(\mathsf{CrJ}, \mathsf{CrJ}) \notin \mathsf{inc}'$ since $(\mathsf{mid}, \mathsf{mid}) \notin \underline{\mathsf{inc}}_{C^3}$.

Concerning $\overline{\mathsf{inc}}_{C^3}$ a counterexample is provided by $\Lambda_1 = \{\mathtt{in}, \mathtt{out}\}$ with $MCCD(\Lambda_1) = \{\mathtt{in}, \mathtt{out}\}$ and $\mathsf{syn}_{\mathrm{AJ}}(\Lambda_1) = \mathsf{syn}_{\mathrm{AJ}}(MCCD(\Lambda_1)) = \mathsf{CrJ}$ while $(\mathsf{mid}, \mathsf{mid}) \in \overline{\mathsf{inc}}_{C^3}$. □

The fact that $\mathsf{syn}_{\mathrm{AJ}}$ is not consistency preserving according to $\overline{\mathsf{inc}}_{C^3}$ is not surprising, given that $\overline{\mathsf{inc}}_{C^3}$ essentially reflects the fully bipolar nature of stable semantics, while $\mathsf{syn}_{\mathrm{AJ}}$ admits tripolar assessments.

As stable labellings are a special case of complete labellings, it can be remarked however that, when applied to stable labellings, $\mathsf{syn}_{\mathrm{AJ}}$ partially preserves consistency, which is, in a sense, weakened from $\overline{\mathsf{inc}}_{C^3}$ to $\mathsf{inc}^c_{C^3}$. We leave the investigation of some formal notion of partial preservation to future work.

In contrast with the essentially positive result concerning consistency preservation properties of $\mathsf{syn}_{\mathrm{AJ}}$, reinstatement preservation turns out to be problematic, as discussed in next section.

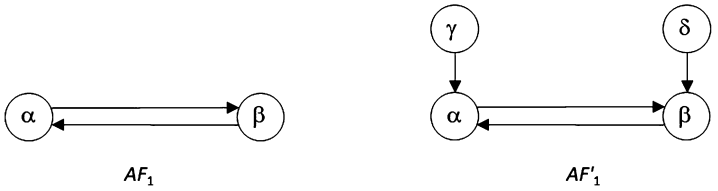


Fig. 5. The argumentation frameworks AF_1 and AF_1' used in Example 4.6.

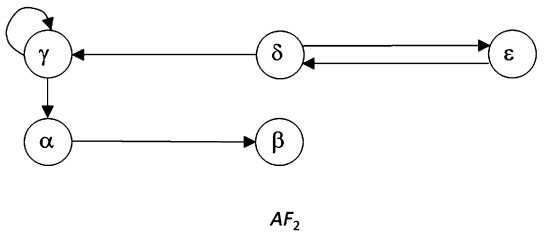


Fig. 6. The argumentation framework AF_2 used in Example 4.6.

4.2. An impossibility result for reinstatement preservation

Addressing reinstatement preservation turns out to require more articulated considerations: in particular basic counterexamples to its satisfaction can easily be identified.

Example 4.6 *(Mutual rejection).* Consider the argumentation framework $AF_1 = \langle\{\alpha,\beta\},\{(\alpha,\beta),(\beta,\alpha)\}\rangle$ *i.e.*, a pair of mutually attacking arguments (see Fig. 5). In this case, under grounded semantics the unique prescribed labelling L is such that $L(\alpha) = L(\beta) = \texttt{und}$. Then the synthesis produced by $\mathsf{syn}_{\mathsf{AJ}}$ for both arguments is NoJ, corresponding to the pair (neg, neg) which is 'forbidden' by both $\mathsf{rv}^{cf}_{C^3}$ and $\mathsf{rv}^{c}_{C^3}$. In fact, for arguments α and β, the outcome produced by $\mathsf{syn}_{\mathsf{AJ}}$ is the same as for the argumentation framework $AF_1' = \langle\{\alpha,\beta,\gamma,\delta\},\{(\alpha,\beta),(\beta,\alpha),(\gamma,\alpha),(\delta,\beta)\}\rangle$, *i.e.*, a case where both arguments are definitely rejected. Equating these two different situations is obviously debatable and, in a sense, the use of Λ^{AJ} and $\mathsf{syn}_{\mathsf{AJ}}$ involves a significant information loss.

In this example, the problem would not arise with the use of a multiple status semantics like stable, semi-stable, or preferred semantics. In fact, we would get two prescribed labellings $L_1 = \{(\alpha,\texttt{in}),(\beta,\texttt{out})\}$ and $L_2 = \{(\alpha,\texttt{out}),(\beta,\texttt{in})\}$. Then the synthesis produced by $\mathsf{syn}_{\mathsf{AJ}}$ for both arguments is CrJ, corresponding to the pair (mid, mid) which is allowed by both $\mathsf{rv}^{cf}_{C^3}$ and $\mathsf{rv}^{c}_{C^3}$.

Reinstatement preservation failures can occur also with multiple status semantics, though. Consider the argumentation framework depicted in Fig. 6: $AF_2 = \langle\{\alpha,\beta,\gamma,\delta,\epsilon\},\{(\alpha,\beta),(\gamma,\alpha),(\gamma,\gamma),(\delta,\gamma),(\delta,\epsilon),(\epsilon,\delta)\}\rangle$. Here preferred semantics prescribes the labellings $L_1 = \{(\alpha,\texttt{und}),(\beta,\texttt{und}),(\gamma,\texttt{und}),(\delta,\texttt{out}),(\epsilon,\texttt{in})\}$ and $L_2 = \{(\alpha,\texttt{in}),(\beta,\texttt{out}),(\gamma,\texttt{out}),(\delta,\texttt{in}),(\epsilon,\texttt{out})\}$. The synthesis produced by $\mathsf{syn}_{\mathsf{AJ}}$ for α is CrJ and for β is NoJ corresponding to the pair (mid, neg) which is forbidden by both $\mathsf{rv}^{cf}_{C^3}$ and $\mathsf{rv}^{c}_{C^3}$.

To address the issues evidenced in Example 4.6, one may observe that the difficulties with reinstatement preservation derive from the asymmetry of Λ^{AJ} (and hence of $\mathsf{syn}_{\mathsf{AJ}}$), which encompasses two levels of positive justification but considers only one 'flat' level of negative justification, thus showing a sort of bias in terms of higher attention paid to positive statuses and consistency issues with respect to their dual notions.

In particular, one may observe that $\mathsf{syn}_{\mathsf{AJ}}$ fails to satisfy a simple desirable property, that we call faithfulness. Intuitively, we consider a ssf syn faithful with respect to the set of labels which are aggregated, if the classification of the label produced by syn is not 'surprising' with respect to the set of classifications of the labels which are aggregated.

Definition 4.10. Let C be a sac, Λ_1 and Λ_2 two C-classified sals. A ssf syn from Λ_1 to Λ_2 is *faithful* iff for every $\emptyset \subsetneq \Lambda \subseteq \Lambda_1$ $\exists c \in \{C_{\Lambda_1}(\lambda) \mid \lambda \in \Lambda\} : C_{\Lambda_2}(\mathsf{syn}(\Lambda)) \preceq c$ and $\exists c' \in \{C_{\Lambda_1}(\lambda) \mid \lambda \in \Lambda\} : c' \preceq C_{\Lambda_2}(\mathsf{syn}(\Lambda))$.

In words, the result produced by syn is neither strictly greater nor strictly lower than all labels which are aggregated. In particular, the aggregation of a singleton $\{\lambda_1\}$ must produce as result an element λ_2 with $C_{\Lambda_1}(\lambda_1) = C_{\Lambda_2}(\lambda_2)$ *i.e.*, belonging to the same class as the only element of the singleton. This basic requirement is violated by $\mathsf{syn}_{\mathsf{AJ}}$, as $\mathsf{syn}_{\mathsf{AJ}}(\{\texttt{und}\}) = \mathsf{NoJ}$, with $C^3_{\Lambda^{\mathsf{IOU}}}(\texttt{und}) = \mathsf{mid}$ while $C^3_{\Lambda^{\mathsf{AJ}}}(\mathsf{NoJ}) = \mathsf{neg}$.

To overcome this limitation, one can consider an alternative notion of justification status which distinguishes non-positive justifications statuses into strongly not justified, denoted as SNoJ, which occurs when the argument is labelled out in all extensions, and weakly not justified, denoted as WNoJ, which occurs when the argument is never labelled in and is labelled und at least once.

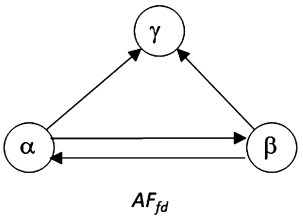


Fig. 7. The argumentation framework AF_{fd} used in Example 4.7.

Accordingly, we consider a sal $\Lambda'^{\mathrm{AJ}} = \{\mathsf{SkJ}, \mathsf{CrJ}, \mathsf{WNoJ}, \mathsf{SNoJ}\}$, with classification $C^3_{\Lambda'^{\mathrm{AJ}}} = \{(\mathsf{SkJ}, \mathsf{pos}), (\mathsf{SNoJ}, \mathsf{neg}), (\mathsf{CrJ}, \mathsf{mid}), (\mathsf{WNoJ}, \mathsf{mid})\}$, and we introduce the ssf $\mathsf{syn}'_{\mathrm{AJ}}$ from Λ^{IOU} to Λ'^{AJ} defined, for every $\Lambda \subseteq \Lambda^{\mathrm{IOU}}$ as follows:

- $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda) = \mathsf{SkJ}$ if $\Lambda = \{\mathtt{in}\}$;
- $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda) = \mathsf{CrJ}$ if $\Lambda \supsetneq \{\mathtt{in}\}$;
- $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda) = \mathsf{SNoJ}$ if $\Lambda = \{\mathtt{out}\}$;
- $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda) = \mathsf{WNoJ}$ otherwise.

Intuitively, the adoption of Λ'^{AJ} and $\mathsf{syn}'_{\mathrm{AJ}}$ fills the gap related to the inherent asymmetry of Λ^{AJ} and, in fact, it is easy to see that $\mathsf{syn}'_{\mathrm{AJ}}$ is faithful.

It is immediate to see that faithfulness is a sufficient condition for both consistency and reinstatement preservation when the set of labellings is a singleton (which covers in particular the case of grounded semantics).

Proposition 4.8. *Let C be a* sac *equipped with an incompatibility relation* inc *(a reinstatement violation relation* rv*), let Λ_1 and Λ_2 be two C-classified* sals*, and* syn *a faithful* ssf *from Λ_1 to Λ_2. For any set S equipped with an intolerance relation* int *and any non-empty* int-inc-*consistent (*int-rv-*compliant) set $\mathcal{L}_1$ of Λ_1-labellings of S such that $|\mathcal{L}_1| = 1$, it holds that the labelling $DL^{\mathsf{syn}}_{\mathcal{L}_1}$ is* int-inc-*consistent (*int-rv-*compliant).*

Proof. Let $\mathcal{L}_1 = \{L\}$. Then, for any $s \in S$ it holds that $\mathcal{L}_1^{\downarrow}(s) = \{L(s)\}$ and then, from faithfulness of syn we get $C_{\Lambda_1}(L(s)) = C_{\Lambda_2}(\mathsf{syn}(\mathcal{L}_1^{\downarrow}(s)))$. It follows that every consistency or reinstatement property satisfied by $L(s)$ is also satisfied by $DL^{\mathsf{syn}}_{\mathcal{L}_1}$. □

While the simple result above confirms that faithfulness is a desirable property, another counterexample turns out to require further considerations.

Example 4.7 *(Floating defeat).* Consider the argumentation framework: $AF_{fd} = \langle\{\alpha, \beta, \gamma\}, \{(\alpha, \beta), (\beta, \alpha), (\alpha, \gamma), (\beta, \gamma)\}\rangle$ (see Fig. 7). Here preferred and stable semantics prescribe two labellings: $L_1 = \{(\alpha, \mathtt{in}), (\beta, \mathtt{out}), (\gamma, \mathtt{out})\}$ and $L_2 = \{(\alpha, \mathtt{out}), (\beta, \mathtt{in}), (\gamma, \mathtt{out})\}$. The synthesis produced by $\mathsf{syn}'_{\mathrm{AJ}}$ for α and β is CrJ and for γ is SNoJ corresponding to the pair (mid, neg) which is forbidden by both $\mathsf{rv}^{cf}_{C^3}$ and $\mathsf{rv}^{c}_{C^3}$.

Example 4.7 has to do with the debated question of floating defeat and, in a sense, provides a further viewpoint on its thorny nature. Here an argument, namely γ, is labelled out in all labellings. As a consequence, any faithful synthesis can only produce an outcome whose classification is neg. However, the label out is motivated by different arguments in different labellings, which in a sense play in turn the role of 'effective attacker': no single argument is labelled in in all labellings and reinstatement preservation turns out to be incompatible with consistency preservation under the weak requirements imposed by conflict-freeness, as formalized by the following result.

Proposition 4.9. *Given the argumentation framework and relevant labellings presented in Example 4.7, under the use of the* sac $C^3 = \{\mathsf{pos}, \mathsf{mid}, \mathsf{neg}\}$ *and of the* sal Λ^{IOU} *for the labellings C^3-classified according to $C^3_{\Lambda^{IOU}}$, for any* sal Λ *there is no simple synthesis function* syn *from Λ^{IOU} to Λ which is faithful, consistency-preserving for $\underline{\mathsf{inc}}_{C^3}$ and reinstatement-preserving for $\mathsf{rv}^{cf}_{C^3}$.*

Proof. With reference to Example 4.7, letting $\mathcal{L} = \{L_1, L_2\}$ we get $\mathcal{L}^{\downarrow}(\alpha) = \mathcal{L}^{\downarrow}(\beta) = \{\mathtt{in}, \mathtt{out}\}$ and $\mathcal{L}^{\downarrow}(\gamma) = \{\mathtt{out}\}$. Given that syn is faithful it follows that for any set of labels Λ it holds that $C^3_{\Lambda}(\mathsf{syn}(\mathcal{L}^{\downarrow}(\gamma))) = \mathsf{neg}$. Let us now consider the possible options for $c = C^3_{\Lambda}(\mathsf{syn}(\mathcal{L}^{\downarrow}(\alpha))) = C^3_{\Lambda}(\mathsf{syn}(\mathcal{L}^{\downarrow}(\beta))) = C^3_{\Lambda}(\mathsf{syn}(\{\mathtt{in}, \mathtt{out}\}))$. If $c = \mathsf{pos}$ the outcome is not $\rightarrow$-$\underline{\mathsf{inc}}_{C^3}$-consistent with respect to the mutually attacking arguments α and β, while if $c = \mathsf{mid}$ or $c = \mathsf{neg}$ the outcome is not $\rightarrow$-$\mathsf{rv}^{cf}_{C^3}$-compliant with respect to γ. □

The problem evidenced by Proposition 4.9 is rooted in a difficulty intrinsic to the notion of reinstatement preservation and not specific to the argumentation domain. Letting $\mathcal{L} = \{L_1, \ldots, L_n\}$ be a set of labellings of a set S and focusing on a given element s

of S, in every labelling $L_i \in \mathcal{L}$ a single element s' (which may differ from labelling to labelling) is enough to ensure reinstatement compliance as far as the value of $L_i(s)$ is concerned. In the example, α and β ensure in turn reinstatement compliance for the label `out` assigned in both labellings to γ. However, when a synthesis is drawn, it may happen that while the synthesised label of s is, in a sense, reminiscent of the effect of the various individual compliance-ensuring elements s', none of them keeps a suitable aggregated value, due to a sort of 'dilution'. In the example, the synthesised label of γ remains `out` but the information that at least one of its attackers was `in` in each labelling is somehow lost.

This suggests that in order to achieve a viable notion of reinstatement preservation some additional technical device is needed, which we propose next.

4.3. A viable notion of reinstatement preservation

As evidenced in the previous section, a formal mechanism to manage the information which can justify, from a reinstatement perspective, a given synthesised labelling is necessary. To this purpose we introduce the notion of r-enhanced sets and r-enhanced intolerance relations. In words, for each element $s \in S$ which has at least two intolerant elements (formally $|snt(s)| > 1$), an additional virtual element, denoted as $\widehat{s}$, is introduced, which is meant to keep track of the most effective (from a reinstatement perspective) labels of the intolerant elements.

Definition 4.11. Given a set S equipped with an intolerance relation int, the corresponding *r-enhanced set and relation*, denoted respectively as $\widetilde{S}$ and $\widetilde{\mathsf{int}}$, are defined as follows:

- $\widetilde{S} \triangleq S \cup \{\widehat{s} \mid s \in S \text{ and } |snt(s)| > 1\}$;
- $\widetilde{\mathsf{int}} \triangleq \mathsf{int} \cup \{(\widehat{s}, s) \mid s \in S \text{ and } |snt(s)| > 1\}$.

A labelling on a set S is extended to its r-enhanced version $\widetilde{S}$ by assigning to each additional element $\widehat{s}$ a label with the highest class among those of the elements of $snt(s)$. In case more elements have labels with the same highest class the choice is arbitrary among them. As we will see, this non deterministic choice is inessential to the achievement of the desired results.

Definition 4.12. Given a sac C and a C-classifed sal Λ, a set S, and a Λ-labelling L of S, a corresponding *r-enhanced labelling* $\widetilde{L}$ on $\widetilde{S}$ is defined for every $s \in \widetilde{S}$ as follows:

- $\widetilde{L}(s) = L(s)$ if $s \in S$;
- $\widetilde{L}(\widehat{s}) = choice\left(\underset{\lambda \in \{L(s') \mid s' \in snt(s)\}}{\arg\max} C_\Lambda(\lambda)\right)$ if $\widehat{s} \in \widetilde{S} \setminus S$

where $choice(C)$ is an operator returning an arbitrary element[9] of the set C. For a labelling L the set of corresponding r-enhanced labellings is denoted as $\widehat{L}$.

Example 4.8. To illustrate the above notions, consider again the argumentation framework presented in Example 4.7, namely AF_{fd} with $S = \{\alpha, \beta, \gamma\}$ and $\mathsf{int} = \{(\alpha, \beta), (\beta, \alpha), (\alpha, \gamma), (\beta, \gamma)\}$.

We get $\widetilde{S} = \{\alpha, \beta, \gamma, \widehat{\gamma}\}$ and $\widetilde{\mathsf{int}} = \{(\alpha, \beta), (\beta, \alpha), (\alpha, \gamma), (\beta, \gamma), (\widehat{\gamma}, \gamma)\}$. The only corresponding r-enhanced labelling $\widetilde{L_1}$ coincides with L_1 with the addition of $\widetilde{L_1}(\widehat{\gamma}) = \mathtt{in}$ and similarly $\widetilde{L_2}$ coincides with L_2 with the addition of $\widetilde{L_2}(\widehat{\gamma}) = \mathtt{in}$.

The properties of consistency and reinstatement compliance need small adaptations for r-enhanced labellings, since the additional virtual elements need a suitable treatment. On the one hand, the labels they are assigned are not subjected to any constraint, on the other hand they affect the evaluation of the elements they are intolerant of. This is reflected in the following definitions.

Definition 4.13. Given a set S equipped with an intolerance relation int, a sac C equipped with an incompatibility relation inc, and a C-classified sal Λ, for any Λ-labelling L of S a corresponding r-enhanced labelling $\widetilde{L}$ is int-inc-*inconsistent* iff

$$\exists s_1 \in \widetilde{S}, s_2 \in S \text{ such that } (s_1, s_2) \in \widetilde{\mathsf{int}} \text{ and } \widetilde{L}(s_1) \underline{\dot{\square}} \widetilde{L}(s_2) \tag{8}$$

Conversely, we say that $\widetilde{L}$ is int-inc-*consistent* if it is not int-inc-inconsistent, *i.e.*,

$$\forall s_1 \in \widetilde{S}, s_2 \in S \text{ such that } (s_1, s_2) \in \widetilde{\mathsf{int}} \text{ it holds that } \widetilde{L}(s_1) \underline{\boxminus} \widetilde{L}(s_2). \tag{9}$$

Definition 4.14. Given a set S equipped with an intolerance relation int, a sac C equipped with a reinstatement violation relation rv, and a C-classified sal Λ, for any Λ-labelling L of S a corresponding r-enhanced labelling $\widetilde{L}$ is int-rv-*uncompliant* iff

[9] In case the set $snt(s)$ is infinite this requires adopting the axiom of choice.

$$\exists s_2 \in S : \begin{cases} \min(C)\overline{\boxdot}C_\Lambda(\widetilde{L}(s_2)) & \text{if } s_2 \text{ is initial} \\ \forall s_1 \in \widetilde{S} \text{ such that } (s_1, s_2) \in \widetilde{\mathsf{int}} \text{ it holds that } \widetilde{L}(s_1)\overline{\boxdot}\widetilde{L}(s_2) & \text{otherwise} \end{cases} \tag{10}$$

Conversely, we say that $\widetilde{L}$ is int-rv-*compliant* if it is not int-rv-uncompliant, *i.e.*,

$$\forall s_2 \in S \begin{cases} \min(C)\overline{\boxminus}C_\Lambda(\widetilde{L}(s_2)) & \text{if } s_2 \text{ is initial} \\ \exists s_1 \in \widetilde{S} \text{ such that } (s_1, s_2) \in \widetilde{\mathsf{int}} \text{ and } \widetilde{L}(s_1)\overline{\boxminus}\widetilde{L}(s_2) & \text{otherwise} \end{cases} \tag{11}$$

In words, consistency and reinstatement are evaluated by taking into account also the additional elements $\hat{s}$ included in $\widetilde{S}$, whose labels are taken for granted (*i.e.*, are not required to satisfy the condition on initial elements).

It is easy to see that an r-enhanced labelling $\widetilde{L}$ inherits the consistency and reinstatement compliance properties from L.

Proposition 4.10. *Given a* sac C *equipped with an incompatibility relation* inc*, a C-classified* sal Λ *and a set S equipped with an intolerance relation* int*, if a Λ-labelling L of S is (not)* int-inc-*inconsistent then any corresponding r-enhanced labelling $\widetilde{L} \in \widehat{L}$ is (not)* int-inc-*inconsistent.*

Proof. If L is int-inc-inconsistent then $\exists s_1, s_2 \in S$ such that $s_1 \odot s_2$ and $L(s_1)\underline{\boxdot}L(s_2)$ and, since $s_1, s_2 \in \widetilde{S}$, also $\widetilde{L}$ is int-inc-inconsistent.

If L is not int-inc-inconsistent then $\forall s_1, s_2 \in S$ such that $s_1 \odot s_2$ it holds that $L(s_1)\underline{\boxminus}L(s_2)$. Assume by contradiction that $\widetilde{L}$ is int-inc-inconsistent. Then, it must be the case that $\exists \hat{s} \in \widetilde{S} \setminus S$ and $s \in S$ such that $\widetilde{L}(\hat{s})\underline{\boxdot}\widetilde{L}(s)$. However, by Definition 4.12, $\exists s' \in S$ such that $s' \odot s$ and $\widetilde{L}(\hat{s}) = L(s')$, while $\widetilde{L}(s) = L(s)$, thus getting $L(s')\underline{\boxdot}L(s)$ which contradicts the hypothesis that L is not int-inc-inconsistent. □

Proposition 4.11. *Given a* sac C *equipped with a reinstatement violation relation* rv*, a C-classified* sal Λ *and a set S equipped with an intolerance relation* int*, if a Λ-labelling L of S is (not)* int-rv-*uncompliant then any corresponding r-enhanced labelling $\widetilde{L} \in \widehat{L}$ is (not)* int-rv-*uncompliant.*

Proof. If L is int-rv-uncompliant then one of the two cases of (5) in Definition 2.10 applies. If the first case holds, then obviously also the first case of (10) in Definition 4.14 holds. As to the second case, given the hypothesis that $\forall s_1 \in S$ such that $s_1 \odot s_2$ it holds that $L(s_1)\overline{\boxdot}L(s_2)$, assume by contradiction that $\widetilde{L}$ is not int-rv-uncompliant. It must then be the case that $\exists s \in S$ such that $\widetilde{L}(\hat{s})\overline{\boxminus}\widetilde{L}(s_2)$. However, by Definition 4.12, $\exists s' \in S$ such that $s' \odot s$ and $\widetilde{L}(\hat{s}) = L(s')$, thus getting $L(s')\overline{\boxminus}L(s)$ which contradicts the hypothesis.

If L is not int-rv-uncompliant then we observe that both conditions specified in (6) of Definition 2.10 are not affected by the additional elements in $\widetilde{S}$ and hence continue to hold for $\widetilde{L}$, implying that the corresponding conditions in (11) of Definition 4.14 hold in turn. □

We can now consider an adjusted version of the notion of reinstatement preservation.

Definition 4.15. Let C be a sac equipped with a reinstatement violation relation rv, and Λ_1 and Λ_2 be two C-classified sets of labels. A ssf syn from Λ_1 to Λ_2 is *weakly reinstatement preserving* iff for any set S equipped with an intolerance relation int and any non-empty int-rv-compliant set $\mathcal{L}_1$ of Λ_1-labellings of S it holds that any labelling $DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}_1}}$ is int-rv-compliant (see Definition 4.14), where, with a little abuse of notation, for a set of labellings $\mathcal{L}$ we define $\widetilde{\mathcal{L}} = \{choice\left(\widehat{L}\right) \mid L \in \mathcal{L}\}$.

In words, the requirement of reinstatement preservation is verified by applying the synthesis function on sets of r-enhanced labellings and checking that the outcome is still int-rv-compliant.

It is easy to see that a reinstatement preserving ssf is also weakly reinstatement preserving.

Proposition 4.12. *Let C be a* sac *equipped with a reinstatement violation relation* rv*, and Λ_1 and Λ_2 be two C-classified sets of labels. If a* ssf syn *from Λ_1 to Λ_2 is reinstatement preserving then it is also weakly reinstatement preserving.*

Proof. Consider an int-rv-compliant set $\mathcal{L}_1$ of Λ_1-labellings. Since syn is reinstatement preserving, $DL^{\mathsf{syn}}_{\mathcal{L}_1}$ is int-rv-compliant. According to Definition 4.12, it is easy to see that any $DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}_1}}$ coincides in S with $DL^{\mathsf{syn}}_{\mathcal{L}_1}$, thus by Definition 4.14 the fact that the latter is int-rv-compliant entails that the first is int-rv-compliant too. □

As we will see, the weakening of the notion of reinstatement preservation avoids impossibility and provides anyway a useful soundness check for synthesis functions.

4.4. Characterizing weakly reinstatement preserving synthesis functions

In order to provide a characterization of weakly reinstatement preserving synthesis functions, we follow a path similar to the one followed in Section 4.1 for characterizing consistency preserving synthesis functions.

First, we introduce a notion of r-compatible dual for reinstatement. Differently from the case of inconsistency, the notion here takes into account the presence of additional virtual elements. The idea is that, given a set of labels Λ_1 and a reinstatement violation relation, a set of labels Λ_2 is a compatible dual of Λ_1 if, given an inst-rv-compliant set of labellings $\mathcal{L}_1$, assuming that $\Lambda_1 = \mathcal{L}_1^{\downarrow}(s_1)$ for some element $s_1 \in S$, it is possible that $\Lambda_2 = \widetilde{\mathcal{L}_1}^{\downarrow}(s_2)$ for some element $s_2 \in \widetilde{S}$ such that $(s_2, s_1) \in \widetilde{\mathsf{inst}}$. This gives rise to the following definition.

Definition 4.16. Let C be a sac, Λ be a C-classifed set of labels, and $\Lambda_1 \subseteq \Lambda$. Given a reinstatement violation relation rv on C, we say that $\Lambda_2 \subseteq \Lambda$ is a rv-*r-compatible dual* of Λ_1, denoted as $\Lambda_2 \in RCD(\Lambda_1)$, iff $\forall \lambda \in \Lambda_1\ \exists \lambda' \in \Lambda_2$ such that $\lambda' \in \overline{rc}(\lambda)$, and $\forall \lambda' \in \Lambda_2$ $\exists \lambda \in \Lambda_1$ such that $\lambda' \in \overline{rc}(\lambda)$.

Example 4.9. Consider Λ^{IOU} and the reinstatement violation relation $\mathsf{rv}^{c}_{C^3} = \{(\mathsf{neg}, \mathsf{neg}), (\mathsf{neg}, \mathsf{mid}), (\mathsf{mid}, \mathsf{neg})\}$. Letting $\Lambda_1 = \{\mathtt{out}, \mathtt{und}\}$ we have that $\overline{rc}(\mathtt{out}) = \{\mathtt{in}\}$, $\overline{rc}(\mathtt{und}) = \{\mathtt{in}, \mathtt{und}\}$. Hence $RCD(\Lambda_1) = \{\{\mathtt{in}\}, \{\mathtt{und}, \mathtt{in}\}\}$: in must be included in every element of $RCD(\Lambda_1)$ since it is the only backward compatible label with out, while both in and und are backward compatible with und. Considering the weaker reinstatement violation relation $\mathsf{rv}^{cf}_{C^3} = \{(\mathsf{neg}, \mathsf{neg}), (\mathsf{mid}, \mathsf{neg})\}$ we would have $\overline{rc}(\mathtt{out}) = \{\mathtt{in}\}$, $\overline{rc}(\mathtt{und}) = \{\mathtt{in}, \mathtt{und}, \mathtt{out}\}$, and hence $RCD(\Lambda_1) = \{\{\mathtt{in}\}, \{\mathtt{in}, \mathtt{und}\}, \{\mathtt{in}, \mathtt{out}\}, \{\mathtt{in}, \mathtt{und}, \mathtt{out}\}\}$.

Note that, though Definition 4.8 and Definition 4.16 are structurally similar, they differ significantly since in Definition 4.16 compatibility is meant to be evaluated 'backwards' with respect to the intolerance relation, as shown by Proposition 4.13, which confirms the intended meaning and soundness of Definition 4.16.

Proposition 4.13. *Let C be a* sac, *Λ be a C-classifed set of labels, and* rv *a well-founded reinstatement violation relation on C. For any set S equipped with an intolerance relation* inst, *any* inst-rv-*compliant set $\mathcal{L}_1$ of Λ-labellings of S and any non-initial element s_1 of S there exists $s_2 \in \widetilde{S}$ such that $(s_2, s_1) \in \widetilde{\mathsf{inst}}$ and $\widetilde{\mathcal{L}_1}^{\downarrow}(s_2) \in RCD(\widetilde{\mathcal{L}_1}^{\downarrow}(s_1))$.*[10]

Proof. Consider first the case where $|snt(s_1)| = 1$, *i.e.*, let $snt(s_1) = \{s_2\}$. Under the hypothesis that every $L \in \mathcal{L}_1$ is inst-rv-compliant it holds that $\forall L \in \mathcal{L}_1\ L(s_2) \in \overline{rc}(L(s_1))$, hence $\forall \lambda \in \mathcal{L}_1^{\downarrow}(s_1)\ \exists \lambda' \in \mathcal{L}_1^{\downarrow}(s_2)$ such that $\lambda' \in \overline{rc}(\lambda)$ and also $\forall \lambda' \in \mathcal{L}_1^{\downarrow}(s_2)\ \exists \lambda \in \mathcal{L}_1^{\downarrow}(s_1)$ such that $\lambda' \in \overline{rc}(\lambda)$, hence $\mathcal{L}_1^{\downarrow}(s_2) \in RCD(\mathcal{L}_1^{\downarrow}(s_1))$ which implies $\widetilde{\mathcal{L}_1}^{\downarrow}(s_2) \in RCD(\widetilde{\mathcal{L}_1}^{\downarrow}(s_1))$ given that $\widetilde{\mathcal{L}_1}^{\downarrow}(s_1) = \mathcal{L}_1^{\downarrow}(s_1)$ and $\widetilde{\mathcal{L}_1}^{\downarrow}(s_2) = \mathcal{L}_1^{\downarrow}(s_2)$.

Consider now the case where $|snt(s_1)| > 1$ and let in this case $s_2 = \widehat{s_1}$. Under the hypothesis that every $L \in \mathcal{L}_1$ is inst-rv-compliant it holds that $\forall L \in \mathcal{L}_1\ \exists s^* \in S$ such that $(s^*, s_1) \in \mathsf{inst}$ and $L(s^*) \in \overline{rc}(L(s_1))$ (note that s^* can vary across different labellings). By Definition 4.12, $L(s^*) \preceq \widetilde{L}(\widehat{s_1})$ and hence by dual monotonicity of rv (see the first point of Definition 2.8) also $\widetilde{L}(\widehat{s_1}) \in \overline{rc}(L(s_1))$. Since this holds for every labelling L it follows as desired that $\forall \lambda \in \widetilde{\mathcal{L}_1}^{\downarrow}(s_1)\ \exists \lambda' \in \widetilde{\mathcal{L}_1}^{\downarrow}(\widehat{s_1})$ such that $\lambda' \in \overline{rc}(\lambda)$ and also $\forall \lambda' \in \widetilde{\mathcal{L}_1}^{\downarrow}(\widehat{s_1})$ $\exists \lambda \in \widetilde{\mathcal{L}_1}^{\downarrow}(s_1)$ such that $\lambda' \in \overline{rc}(\lambda)$. □

In the following definition we then consider the 'extreme case' of r-compatible dual of a set of labels with respect to a reinstatement violation relation.

Definition 4.17. Let C be a sac, Λ a C-classified set of labels, and rv a reinstatement violation relation on C. Given $\Lambda_1 \subseteq \Lambda$ we define $MRCD(\Lambda_1) \triangleq \bigcup_{\lambda \in \Lambda_1} \widehat{MRC}(\lambda)$, where $\widehat{MRC}(\lambda) \triangleq \{\lambda' \in \overline{rc}(\lambda) \mid \nexists \lambda'' \in \overline{rc}(\lambda) : \lambda'' \prec \lambda'\}$.

Note that to ensure non-emptiness of $MRCD(\Lambda_1)$ we need to assume that for every label $\lambda \in \Lambda$ the set $\overline{rc}(\lambda)$ (which is non-empty by the third condition of Definition 2.8) includes a non-empty set of minimal elements with respect to the $\preceq$ relation. As indicated for the analogous assumption in Section 4.1, this obviously holds if Λ is finite and we regard it as a mild requirement otherwise.

Example 4.10. Continuing Example 4.9, with reference to $\mathsf{rv}^{c}_{C^3}$, we have $\widehat{MRC}(\mathtt{out}) = \{\mathtt{in}\}$ and $\widehat{MRC}(\mathtt{und}) = \{\mathtt{und}\}$. Hence $MRCD(\Lambda_1) = \{\mathtt{in}, \mathtt{und}\}$. With reference to $\mathsf{rv}^{cf}_{C^3}$ we have $\widehat{MRC}(\mathtt{out}) = \{\mathtt{in}\}$, $\widehat{MRC}(\mathtt{und}) = \{\mathtt{out}\}$, hence $MRCD(\Lambda_1) = \{\mathtt{in}, \mathtt{out}\}$.

The following propositions show two basic properties of $MRCD(\Lambda_1)$: it belongs to $RCD(\Lambda_1)$ and is minimal with respect to $\preceq_P$.

Proposition 4.14. *Let C be a* sac, *Λ a C-classified set of labels, and* rv *a reinstatement violation relation on C. Then, for any non-empty $\Lambda_1 \subseteq \Lambda$ it holds that $MRCD(\Lambda_1) \in RCD(\Lambda_1)$.*

Proof. From the definition and the above mentioned assumption it is immediate to see that for every $\lambda \in \Lambda_1\ \exists \lambda' \in MRCD(\Lambda_1)$ such that $\lambda' \in \overline{rc}(\lambda)$ and that for every $\lambda' \in MRCD(\Lambda_1)\ \exists \lambda \in \Lambda_1$ such that $\lambda' \in \overline{rc}(\lambda)$. □

[10] We assume here a fixed choice operator for $\widetilde{\mathcal{L}_1}$. The result holds for any actual instance of the operator.

Proposition 4.15. *Let C be a* sac*, Λ a C-classified set of labels, and* rv *a reinstatement violation relation on C. Given a non-empty $\Lambda_1 \subseteq \Lambda$, it holds that $\forall D \in RCD(\Lambda_1), MRCD(\Lambda_1) \preceq_P D$.*

Proof. For any $\lambda' \in D$, from Definition 4.16 it holds that $\exists \lambda \in \Lambda_1$ such that $\lambda' \in \overline{rc}(\lambda)$. Then, by Definition 4.17 $\exists \lambda'' \in MCCD(\Lambda_1)$ such that $\lambda'' \in \overline{rc}(\lambda)$ and $\nexists \lambda''' \in \overline{rc}(\lambda) : \lambda''' \prec \lambda''$ which implies that $\lambda'' \preceq \lambda'$ since $\preceq$ is a total preorder by Proposition 2.1.

Consider now any $\lambda'' \in MRCD(\Lambda_1)$. By Definition 4.17 it holds that $\exists \lambda \in \Lambda_1$ such that $\lambda'' \in \overline{rc}(\lambda)$. Moreover by Definition 4.16 $\exists \lambda' \in D$ such that $\lambda' \in \overline{rc}(\lambda)$. Now by Definition 4.17 we have again that $\nexists \lambda''' \in \overline{rc}(\lambda) : \lambda''' \prec \lambda''$ and hence $\lambda'' \preceq \lambda'$. □

On this basis, we can now derive a necessary and sufficient condition for weak reinstatement preservation by a well-behaved ssf.

Proposition 4.16. *Let C be a* sac*, Λ_1 and Λ_2 two C-classified* sals*, and* rv *a well-founded reinstatement violation relation on C. A well-behaved and faithful* ssf syn *from Λ_1 to Λ_2 is weakly reinstatement preserving if and only if for every non-empty set $\Lambda'_1 \subseteq \Lambda_1$ it holds that* $\mathsf{syn}(\Lambda'_1) \in rc(\mathsf{syn}(MRCD(\Lambda'_1)))$.

Proof. Let syn be a ssf satisfying the hypotheses and assume by contradiction that syn is not weakly reinstatement preserving. This means that there is a set S equipped with an intolerance relation int and a non-empty set $\mathcal{L}$ of int-rv-compliant Λ_1-labellings of S such that some labelling $DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}}}$ is int-rv-uncompliant. According to Definition 4.14 and Definition 4.12, this in turn leads to consider two cases: (i) $\exists s_2 \in S$ such that s_2 is initial and $\min(C)\overline{\boxdot}C_{\Lambda_2}(DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}}}(s_2))$; (ii) $\exists s_2 \in S$ such that s_2 is not initial and $\forall s_1 \in \widetilde{S}$ such that $(s_1, s_2) \in \widetilde{\mathsf{int}}$ it holds that $DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}}}(s_1)\overline{\boxdot}DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}}}(s_2)$. Since, for any s, $DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}}}(s) = \mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s))$, this condition is equivalent to $\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_1))\overline{\boxdot}\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_2))$, *i.e.*, $\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_2)) \notin rc(\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_1)))$.

As to (i), since syn is faithful we have that for any initial $s_2 \in S$ it holds that $\exists L \in \mathcal{L}$ such that $C_{\Lambda_1}(L(s_2)) \preceq C_{\Lambda_2}(DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}}}(s_2))$. Then, from $\min(C)\overline{\boxdot}C_{\Lambda_2}(DL^{\mathsf{syn}}_{\widetilde{\mathcal{L}}}(s_2))$ and the dual monotonicity of rv (*i.e.*, the first condition of Definition 2.8) it follows $\min(C)\overline{\boxdot}C_{\Lambda_1}(L(s_2))$, which contradicts the hypothesis that every $L \in \mathcal{L}$ is int-rv-compliant.

As to (ii), from Proposition 4.13 we have that there exists $s_1 \in \widetilde{S}$ such that $(s_1, s_2) \in \widetilde{\mathsf{int}}$ and $\widetilde{\mathcal{L}}^{\downarrow}(s_1) \in RCD(\widetilde{\mathcal{L}}^{\downarrow}(s_2))$ and hence from Proposition 4.15 it holds that $MRCD(\widetilde{\mathcal{L}}^{\downarrow}(s_2)) \preceq_P \widetilde{\mathcal{L}}^{\downarrow}(s_1)$. Since syn is well-behaved, $\mathsf{syn}(MRCD(\widetilde{\mathcal{L}}^{\downarrow}(s_2))) \preceq \mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_1))$. Moreover, by the hypothesis of this proposition (with $\Lambda'_1 = \widetilde{\mathcal{L}}^{\downarrow}(s_2)$, which is non-empty) we have $\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_2)) \in rc(\mathsf{syn}(MRCD(\widetilde{\mathcal{L}}^{\downarrow}(s_2))))$. However, by the dual monotonicity of rv (*i.e.*, the first condition of Definition 2.8) it must be the case that $\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_2)) \in rc(\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_1)))$, contradicting $\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_2)) \notin rc(\mathsf{syn}(\widetilde{\mathcal{L}}^{\downarrow}(s_1)))$.

As to the other direction of the proof, assume now that syn is weakly reinstatement preserving. Since by Proposition 4.14 for every non-empty set $\Lambda'_1 \subseteq \Lambda_1$ it holds that $MRCD(\Lambda'_1) \in RCD(\Lambda'_1)$, considering a simple situation where $S = \{s_1, s_2\}$ and $\mathsf{int} = \{(s_1, s_2)\}$ (note that in this case $\widetilde{S} = S$ and $\widetilde{\mathsf{int}} = \mathsf{int}$) we can identify an int-rv-compliant set $\mathcal{L}_1$ of Λ_1-labellings such that $\mathcal{L}_1^{\downarrow}(s_2) = \Lambda'_1$ and $\mathcal{L}_1^{\downarrow}(s_1) = MRCD(\Lambda'_1)$. Then since syn is weakly reinstatement preserving it must also hold that $\mathsf{syn}(\Lambda'_1) \in rc(\mathsf{syn}(MRCD(\Lambda'_1)))$. □

4.5. Comparing synthesis functions for argument justification

Exploiting the general results obtained above, we can now turn back to the analysis of synthesis functions for argument justification and, in particular, compare $\mathsf{syn}_{\mathrm{AJ}}$, corresponding to the traditional notion of argument justification, with the novel proposal of $\mathsf{syn}'_{\mathrm{AJ}}$ introduced in Section 4.2.

We already know from Proposition 4.6 that $\mathsf{syn}_{\mathrm{AJ}}$ is well-behaved and from the simple counterexample provided in Section 4.2 that it is not faithful. The following propositions show that $\mathsf{syn}'_{\mathrm{AJ}}$ satisfies both properties.

Proposition 4.17. *The* ssf syn'_{AJ} *is well-behaved.*

Proof. As in the proof of Proposition 4.6, since the strict order $\prec$ induced on Λ'^{AJ} is total, it is sufficient to show that for any two non-empty sets $\Lambda_1, \Lambda_2 \subseteq \Lambda^{\mathrm{IOU}}$ whenever $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda_1) \prec \mathsf{syn}'_{\mathrm{AJ}}(\Lambda_2)$ it does not hold that $\Lambda_2 \preceq_P \Lambda_1$. First, consider the case $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda_2) = \mathsf{SkJ}$, entailing $\Lambda_2 = \{\mathtt{in}\}$ and, since $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda_1) \prec \mathsf{syn}'_{\mathrm{AJ}}(\Lambda_2)$, $\Lambda_1 \neq \{\mathtt{in}\}$. Then, it is easy to see that $\{\mathtt{in}\} \not\preceq_P \Lambda_1$ for any $\Lambda_1 \subseteq \Lambda^{\mathrm{IOU}}$, with $\Lambda_1 \notin \{\emptyset, \{\mathtt{in}\}\}$. Consider now the case $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda_2) \in \{\mathsf{WNoJ}, \mathsf{CrJ}\}$, entailing $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda_1) = \mathsf{SNoJ}$ and thus $\Lambda_1 = \{\mathtt{out}\}$. It is also easy to see that for any $\Lambda_2 \subseteq \Lambda^{\mathrm{IOU}}$, with $\Lambda_2 \notin \{\emptyset, \{\mathtt{out}\}\}$ it holds that $\Lambda_2 \not\preceq_P \{\mathtt{out}\}$. The conclusion follows from the fact that Λ_2 is non-empty and different from $\{\mathtt{out}\}$, since $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda_2) \neq \mathsf{SNoJ}$. Finally, the case $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda_2) = \mathsf{SNoJ}$ is impossible if $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda_1) \prec \mathsf{syn}'_{\mathrm{AJ}}(\Lambda_2)$, since there is no label strictly lower than SNoJ. □

Proposition 4.18. *The* ssf syn'_{AJ} *is faithful.*

Proof. We need to show that for every $\emptyset \subsetneq \Lambda \subseteq \Lambda^{\mathrm{IOU}}$ $\exists c \in \{C^3_{\Lambda^{\mathrm{IOU}}}(\lambda) \mid \lambda \in \Lambda\} : C^3_{\Lambda'^{\mathrm{AJ}}}(\mathsf{syn}'(\Lambda)) \leq c$ and $\exists c' \in \{C^3_{\Lambda^{\mathrm{IOU}}}(\lambda) \mid \lambda \in \Lambda\} : c' \leq C^3_{\Lambda'^{\mathrm{AJ}}}(\mathsf{syn}'(\Lambda))$. For $\Lambda = \{\mathtt{in}\}$ we have $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda) = \mathsf{SkJ}$ and $C^3_{\Lambda'^{\mathrm{AJ}}}(\mathsf{SkJ}) = \mathsf{pos} = C^3_{\Lambda^{\mathrm{IOU}}}(\mathtt{in})$. For $\Lambda = \{\mathtt{out}\}$ we have $\mathsf{syn}'_{\mathrm{AJ}}(\Lambda) = \mathsf{SNoJ}$ and $C^3_{\Lambda'^{\mathrm{AJ}}}(\mathsf{SNoJ}) = \mathsf{neg} = C^3_{\Lambda^{\mathrm{IOU}}}(\mathtt{out})$. For any other Λ it holds that $C^3_{\Lambda'^{\mathrm{AJ}}}(\mathsf{syn}'_{\mathrm{AJ}}(\Lambda)) = \mathsf{mid}$ and it is easy to see that either $\mathtt{und} \in \Lambda$ with $C^3_{\Lambda^{\mathrm{IOU}}}(\mathtt{und}) = \mathsf{mid}$ or $\Lambda = \{\mathtt{in}, \mathtt{out}\}$ with $C^3_{\Lambda^{\mathrm{IOU}}}(\mathtt{in}) = \mathsf{pos} > \mathsf{mid}$ and $C^3_{\Lambda^{\mathrm{IOU}}}(\mathtt{out}) = \mathsf{neg} < \mathsf{mid}$. □

Table 2
Illustration of the proof of Proposition 4.19.

Λ_1 $\mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1)$	$\underline{\mathsf{inc}}_{C^3}$	$\mathsf{inc}^a_{C^3}$	$\mathsf{inc}^c_{C^3}$
{in} SkJ	{und} WNoJ	{und} WNoJ	{out} SNoJ
{out} SNoJ	{in} SkJ	{in} SkJ	{in} SkJ
{und} WNoJ	{in} SkJ	{und} WNoJ	{und} WNoJ
{in, out} CrJ	{in, und} CrJ	{in, und} CrJ	{in, out} CrJ
{in, und} CrJ	{in, und} CrJ	{und} WNoJ	{und, out} WNoJ
{und, out} WNoJ	{in} SkJ	{in, und} CrJ	{in, und} CrJ
{in, und, out} CrJ	{in, und} CrJ	{in, und} CrJ	{in, und, out} CrJ

We can now see that, like $\mathsf{syn}_{\mathrm{AJ}}$, also $\mathsf{syn'}_{\mathrm{AJ}}$ is consistency preserving.

Proposition 4.19. *The* ssf $\mathsf{syn'}_{AJ}$ *is consistency preserving according to the incompatibility relations* $\underline{\mathsf{inc}}_{C^3}$, $\mathsf{inc}^a_{C^3}$, *and* $\mathsf{inc}^c_{C^3}$ *while it is not according to* $\overline{\mathsf{inc}}_{C^3}$.

Proof. We need to show that for every non-empty set $\Lambda_1 \subseteq \Lambda^{\mathrm{IOU}}$ it holds that $\mathsf{syn'}_{\mathrm{AJ}}(MCCD(\Lambda_1)) \in cc(\mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1))$. Similarly to the proof of Proposition 4.7, for $\underline{\mathsf{inc}}_{C^3}$, $\mathsf{inc}^a_{C^3}$, and $\mathsf{inc}^c_{C^3}$ this is illustrated in Table 2, where the first column presents the various possible cases for Λ_1 with the relevant value of $\mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1)$ and the following columns (illustrating $\underline{\mathsf{inc}}_{C^3}$, $\mathsf{inc}^a_{C^3}$, and $\mathsf{inc}^c_{C^3}$ respectively) show the corresponding $MCCD(\Lambda_1)$ and the relevant value of $\mathsf{syn'}_{\mathrm{AJ}}(MCCD(\Lambda_1))$.

By inspection of the table, it can be checked that, as desired, for every pair $(\mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1), \mathsf{syn'}_{\mathrm{AJ}}(MCCD(\Lambda_1)))$ obtained by taking the first element from a row of the first column, and the second element from any other cell (say the i-th with $i \in \{2,3,4\}$) of the same row it holds that $(\mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1), \mathsf{syn'}_{\mathrm{AJ}}(MCCD(\Lambda_1))) \notin \mathsf{inc'}$ where $\mathsf{inc'}$ is the incompatibility relation induced by the inc relation specified at the top of the i-th column (from which the second element of the pair was taken).

Concerning $\overline{\mathsf{inc}}_{C^3}$, exactly as in the case of $\mathsf{syn}_{\mathrm{AJ}}$ a counterexample is provided by $\Lambda_1 = \{\mathtt{in}, \mathtt{out}\}$ with $MCCD(\Lambda_1) = \{\mathtt{in}, \mathtt{out}\}$ and $\mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1) = \mathsf{syn'}_{\mathrm{AJ}}(MCCD(\Lambda_1)) = \mathsf{CrJ}$ while $(\mathsf{mid}, \mathsf{mid}) \in \overline{\mathsf{inc}}_{C^3}$. □

Turning now to weak reinstatement preservation, we already know from Example 4.6 that $\mathsf{syn}_{\mathrm{AJ}}$ is not reinstatement preserving for the reinstatement violation relations $\mathsf{rv}^{cf}_{C^3}$ and $\mathsf{rv}^{c}_{C^3}$. Since in the argumentation framework used in the example all arguments have exactly one attacker, in this case reinstatement preservation coincides with weak reinstatement preservation and hence we get that $\mathsf{syn}_{\mathrm{AJ}}$ is also not weakly reinstatement preserving.

Instead $\mathsf{syn'}_{\mathrm{AJ}}$ turns out to be satisfactory from this viewpoint.

Proposition 4.20. *The* ssf $\mathsf{syn'}_{AJ}$ *is weakly reinstatement preserving for the reinstatement violation relations* $\mathsf{rv}^{cf}_{C^3}$ *and* $\mathsf{rv}^{c}_{C^3}$.

Proof. We need to show that for every non-empty set $\Lambda_1 \subseteq \Lambda^{\mathrm{IOU}}$ it holds that $\mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1) \in rc(\mathsf{syn'}_{\mathrm{AJ}}(MRCD(\Lambda_1)))$. This is illustrated in Table 3, where the first column presents the various possible cases for Λ_1 with the relevant value of $\mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1)$ and the following columns (illustrating $\mathsf{rv}^{cf}_{C^3}$ and $\mathsf{rv}^{c}_{C^3}$ respectively) show the corresponding $MRCD(\Lambda_1)$ and the relevant value of $\mathsf{syn'}_{\mathrm{AJ}}(MRCD(\Lambda_1))$.

By inspection of Table 3, it can be checked that, as desired, for every pair $(\mathsf{syn'}_{\mathrm{AJ}}(MRCD(\Lambda_1)), \mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1))$ obtained by taking the second element from a row of the first column, and the first element from any other cell of the same row it holds that $(\mathsf{syn'}_{\mathrm{AJ}}(MRCD(\Lambda_1)), \mathsf{syn'}_{\mathrm{AJ}}(\Lambda_1)) \notin \mathsf{rv'}$ where $\mathsf{rv'}$ is the reinstatement violation relation induced by the rv relation specified at the top of the column from which the first element of the pair was taken. For instance, considering the sixth row, the second column indicates that, with respect to $\mathsf{rv}^{cf}_{C^3}$, $MRCD(\{\mathtt{in}, \mathtt{und}\}) = \{\mathtt{out}\}$. Then $\mathsf{syn'}_{\mathrm{AJ}}(MRCD(\{\mathtt{in}, \mathtt{und}\})) = \mathsf{SNoJ}$ while $\mathsf{syn'}_{\mathrm{AJ}}(\{\mathtt{in}, \mathtt{und}\}) = \mathsf{CrJ}$, and $(\mathsf{SNoJ}, \mathsf{CrJ}) \notin \mathsf{rv}^{cf}{}'_{C^3}$ since $(\mathsf{neg}, \mathsf{mid}) \notin \mathsf{rv}^{cf}_{C^3}$. The third column of the same row indicates that, with respect to $\mathsf{rv}^{c}_{C^3}$, $MRCD(\{\mathtt{in}, \mathtt{und}\}) = \{\mathtt{und}, \mathtt{out}\}$. Then in this case $\mathsf{syn'}_{\mathrm{AJ}}(MRCD(\{\mathtt{in}, \mathtt{und}\})) = \mathsf{WNoJ}$ and $(\mathsf{WNoJ}, \mathsf{CrJ}) \notin \mathsf{rv}^{c}{}'_{C^3}$ since $(\mathsf{mid}, \mathsf{mid}) \notin \mathsf{rv}^{c}_{C^3}$. □

Table 3
Illustration of the proof of Proposition 4.20.

Λ_1 $\mathsf{syn}'_{\mathbf{AJ}}(\Lambda_1)$	$\mathsf{rv}^{cf}_{C^3}$	$\mathsf{rv}^{c}_{C^3}$
{in} SkJ	{out} SNoJ	{out} SNoJ
{out} SNoJ	{in} SkJ	{in} SkJ
{und} WNoJ	{out} SNoJ	{und} WNoJ
{in, out} CrJ	{in, out} CrJ	{in, out} CrJ
{in, und} CrJ	{out} SNoJ	{und, out} WNoJ
{und, out} WNoJ	{in, out} CrJ	{in, und} CrJ
{in, und, out} CrJ	{in, out} CrJ	{in, und, out} CrJ

Table 4
Illustration of weak reinstatement preservation in the case of AF_2.

	α	β	γ	δ	ϵ	$\widehat{\gamma}$
$\widetilde{L_1}$	und	und	und	out	in	und
$\widetilde{L_2}$	in	out	out	in	out	in
$DL^{\mathsf{syn}'}_{\widetilde{\mathcal{L}}}$	CrJ	WNoJ	WNoJ	CrJ	CrJ	CrJ

Example 4.11. To exemplify the above result, referring again to the case of AF_{fd} from Example 4.7 and in particular to the set of labellings $\widetilde{\mathcal{L}} = \{\widetilde{L_1}, \widetilde{L_2}\}$ as introduced in Section 4.3, letting $L' = DL^{\mathsf{syn}'}_{\widetilde{\mathcal{L}}}$, we have $L'(\alpha) = L'(\beta) = \mathsf{CrJ}$, $L'(\gamma) = \mathsf{SNoJ}$, $L'(\widehat{\gamma}) = \mathsf{SkJ}$ with the label of $L'(\widehat{\gamma})$ ensuring reinstatement compliance.

Turning to $AF_2 = \langle \mathcal{A}, \rightarrow \rangle$ as defined in Example 4.6, the only argument with more than one attacker is γ, leading to $\widetilde{\mathcal{A}} = \{\alpha, \beta, \gamma, \delta, \epsilon, \widehat{\gamma}\}$ and $\widetilde{\rightarrow} = \{(\alpha, \beta), (\gamma, \alpha), (\gamma, \gamma), (\delta, \gamma), (\delta, \epsilon), (\epsilon, \delta), (\widehat{\gamma}, \gamma)\}$. We have then the set of labellings $\widetilde{\mathcal{L}} = \{\widetilde{L_1}, \widetilde{L_2}\}$ and the argument justification labelling $DL^{\mathsf{syn}'}_{\widetilde{\mathcal{L}}}$ as illustrated in Table 4, which shows that the problem of reinstatement preservation pointed out for AF_2 in Example 4.6 is solved.

5. Discussion and related works

We have presented a generalized treatment of the notions of consistency and reinstatement and shown its application in the context of formal argumentation. Accordingly, the discussion of the related works includes both general considerations and more specific aspects concerning the argumentation field, and is then complemented by some preliminary perspectives on further applications of the proposed formalism.

5.1. Labelling-based assessments

Our proposal is based on the generic notion of an ordered assessment, virtually of any kind, based on labellings. Labellings play a specific role in formal argumentation where they are at the heart of the labelling-based approach to abstract argumentation semantics [8] and, more generally, can be used to capture various phases of an argumentative reasoning process [18].

Concerning argumentation semantics, we focused on tripolar labellings, which are the most commonly adopted ones for argument acceptability assessments, as discussed in Section 3. Quadripolar labellings have also been introduced in the argumentation literature (see, for instance, [13,19]) and will be considered in future work.

As to argument justification, we considered the three traditional labels recalled in Definition 4.1 as a starting point and proposed, as an improvement, a set of four labels in Section 4.2. In the literature alternative (and more articulated) notions of argument justification have been considered too. For instance, in [20], a classification encompassing four states is proposed, namely: uni-accepted (corresponding to traditional skeptical justification), not-accepted (corresponding to traditional no-justification), cleanly-accepted and only-exi-accepted. The two latter states refine the traditional notion of credulous justification, while our proposal refines the notion of no-justification. In [21], in the context of the study of skepticism relations between argumentation semantics, seven possible argument justification states were identified, corresponding to all non-empty subsets of Λ^{IOU}, namely to the possible

alternatives of $\mathcal{L}^{\downarrow}(\alpha)$ given an argument α and a set $\mathcal{L}$ of Λ^{IOU}-labellings. Essentially the same kind of argument justification labelling has been independently proposed in [22]. Encompassing these more articulated proposals into our approach is left to future work.

Throughout the paper, we mainly referred to discrete sets of labels, traditionally used in argumentation. The use of different sets of labels for different assessments (*e.g.*, argument acceptability *vs.* argument justification) motivated the need for a set of assessment classes on top of the various sets of labels.

We remark that the proposed notions are not limited to this context. In particular, nothing prevents real-valued labels and the use of infinite sets of labels, such as the $[0, 1]$ real interval. Putting in correspondence such a set with a tripolar evaluation has been considered in some works in the literature (see, for instance, the notion of epistemic labelling in [23] or the labellings considered in the equational approach of [24]). However, we remark that our approach is not restricted to the use of finite sets of assessment classes, nor imposes that the sets of labels and the set of assessment classes are different if this is not necessary. In contexts where the $[0, 1]$ real interval is uniformly used as the set of labels Λ for various assessments, it can also be used as the sac C with C_{Λ} being the identity function. Further developments in this direction are left to future work, while some related comments can be found in Section 5.4.

5.2. Notions of consistency and reinstatement

Our proposal encompasses a combination of the notions of consistency and reinstatement. Consistency is a ubiquitous term used in a variety of technical and non-technical contexts to indicate a desirable property, with various possible and somehow interrelated intuitive meanings like coherence, stability, firmness, harmony.

Within this broad landscape, we focus on consistency of labellings, which, borrowing terminology from [25,14], can be regarded as a form of *direct* consistency, in the sense that, given a set of elements, their (in)consistency can be verified by a direct check on the elements themselves and their attributes. For instance, given a language equipped with negation, a set of language elements is consistent if it does not include two elements such that one negates the other, which corresponds to the traditional view of consistency as explicit non-contradiction.

In addition to the traditional notion of negation, a more general, not necessarily symmetric, notion of contrariness is adopted in some approaches. For instance, in the ASPIC+ formalism [14], a language equipped with contrariness is considered, and a set of language elements is directly consistent if it does not include two elements such that one is a contrary of the other.

Our approach provides an abstract formalism to represent arbitrary forms of direct consistency and analyzes some of their properties (in particular preservation) in general terms.

It can be observed that the notions of consistency mentioned above are special cases of our approach, where a binary labelling is implicitly considered (with set membership corresponding to a positive label, and non-membership to a negative one) and the intolerance relation corresponds to negation or contrariness. The binary assessments, as the ones above, correspond to a bipolar sac C, which then necessarily consists of its minimal and maximal elements only: $C = \{\min(C), \max(C)\}$. Then, the requirement of well-foundedness leaves no room for alternatives in the definition of the incompatibility relation inc to be adopted, since $(\max(C), \max(C))$ must belong to inc and no other pair can belong to inc.

This observation explains why the notions of intolerance and incompatibility considered in this paper as essential ingredients of (in)consistency need not an explicit treatment in contexts based on bipolar labellings, where the only possible choice for these notions is, so to say, hardwired in the adopted formalisms. In turn, the habit of having these notions implicit in traditional bipolar contexts may explain why they did not receive explicit attention even in contexts, like formal argumentation, where they can play a useful and possibly enlightening role, as discussed in this paper.

We suggest that the ideas underlying our proposal have the potential to capture, as special cases, a large spectrum of direct consistency notions in a variety of domains, enabling the study of alternative options and revealing relationships with other notions, as we have done in Sections 3 for abstract argumentation semantics. Investigating the application of the approach in other contexts is a direction of future work, for which we provide some perspectives in Section 5.7.

Turning to *indirect* consistency notions, they can be regarded, in general terms, as involving some inferential activity carried out on a set of elements, whose outcomes reveal whether the set is consistent. Trivialization and unfeasibility, in addition to contradiction, are typical examples of inconsistency-revealing properties. For instance, in classical logic, a set S of formulas is inconsistent if every formula is a consequence of S. In contrast, in constraint-based reasoning, a set of constraints is inconsistent (or unfeasible) if no variable assignment satisfying all the constraints exists.

Indirect consistency is, in principle, more general than direct consistency and, as in the case of constraints, may refer to cases where no assessment labellings are involved and hence out of the scope of the present work. There are however many significant cases where indirect and direct consistency are related. For instance, deriving any formula is regarded as pathological because it is assumed that the set of all formulas is directly inconsistent, *i.e.*, includes elements that cannot stay together (in particular, because the language includes negation). If this were not the case, deriving all formulas might not be considered problematic per se. Even more explicitly, in [14], indirect consistency is defined in terms of direct consistency: a set of language elements is indirectly consistent if its closure under strict rules is directly consistent. While an extensive discussion of the spectrum of the notions of (in)consistency is beyond the limits of the present paper (see, for instance, [26] for a broad analysis), we regard the investigation of relations between generalized direct and indirect notions of consistency as a promising direction of future work.

Compared with consistency, the notion of reinstatement has a narrower focus: it arises in the context of defeasible reasoning [27] based on the idea that provisional *prima facie* reasons to believe some conclusion are subject to be retracted in presence of defeaters, namely of other reasons attacking them in some way. Defeaters may be defeated since they may be based in turn on *prima facie*

reasons. Then, the idea is that a provisional reason whose defeaters are defeated is *reinstated*, namely, recovers its ability to support some belief.

The idea of reinstatement is embedded in several abstract argumentation semantics, and, in this context, it has been formalized as the *reinstatement principle* in [10], which in the labelling-based formulation corresponds to the requirement that if all the attackers of an argument α are labelled `out` then the argument α should be labelled `in`. In the context of formal argumentation, the notion of reinstatement has received significant attention, including debates on its appropriateness [28,29] and the definition of variants like weak reinstatement and CF-reinstatement in [10], in any case focusing on the notion of defense, *i.e.*, of attackers being attacked. In the context of labelling-based semantics, the idea was first developed in [9], where the term reinstatement labelling refers to an equivalent formulation of the notion of complete labelling. Our proposal provides a related but complementary perspective on the notion of reinstatement, connected to the need not to be too conservative in deriving conclusions. This perspective, besides supporting the definition of different and tunable reinstatement requirements, enables the use of this notion outside its native context and suggests a potential reappraisal of its role as a basic ingredient for the characterization of reasoning activities of various natures.

Regarding the problem of assessing the quality of the outcomes of some reasoning process, consistency and reinstatement can be regarded as two sides of the same coin. On the one hand, consistency can be related to the ability of the process to avoid the presence of defects in its outcomes; on the other hand, reinstatement can be associated with the productivity of the process itself. At one extreme, the empty set of outcomes cannot include defects and is consistent, but it is not a very productive outcome. On the other hand, a process that outputs all possible outcomes corresponds to the maximum productivity in quantitative terms but is also guaranteed to include all possible defects and hence be inconsistent. Both extremes are equally uninformative, and a suitable tradeoff between them has to be found. In this respect, the generalized notion of reinstatement we propose needs not to be confined to nonmonotonic reasoning but is applicable in any context where a reasoning process admits a range of outcomes.

As anticipated in the introduction, consistency establishes a sort of upper bound of the range, while reinstatement is a sort of lower bound, and their combination determines the subrange of the outcomes considered appropriate in a given situation or domain. While this paper has illustrated this idea in the context of formal argumentation, a stimulating direction of future work consists in exploring its use in other settings encompassing some forms of non-bipolar labellings, like, for instance, many-valued logics [30].

5.3. Argumentation semantics and argument justification

We have shown in Section 3 that the generalized notions of consistency and reinstatement provide a novel characterization of some well-known argumentation semantics, and we commented that the characterization carries over to semantics which correspond to imposing some minimality or maximality constraints on complete labellings.

Other argumentation semantics in the literature are defined by referring to some topological features and modifications of the argumentation framework. For instance, CF2 [31] and stage2 [32] semantics are based on the topological notion of strongly connected components of the graph corresponding to the argumentation framework, while the recent notion of weak admissibility [33] (which provides the basis for reformulating other semantics notions too) is based on the notion of *reduct* of an argumentation framework with respect to a set of arguments S (this amounts to consider a modified framework where S and the arguments attacked by S are deleted). While these more articulated schemes of semantics definition appear to be beyond the expressive capabilities of the generalized notions of consistency and reinstatement, it will be interesting to investigate how they can be integrated within such schemes to capture some existing semantics or possibly devise new ones.

We focused in this paper on Dung's argumentation frameworks based on the relation of attack. It is worth recalling, however, that forms of argumentation frameworks, including other kinds of relations, have been considered in the literature. For instance, bipolar argumentation frameworks [34] also involve a relation of support, while Abstract Dialectical Frameworks [35] allow to express a variety of kinds of influences between arguments. In our approach, consistency and reinstatement refer to an intolerance relation, which corresponds intuitively to some form of conflict; exploring the definition of dual notions referring to some form of support, and possibly to more heterogeneous relations, appears to be a further research line worth pursuing.

5.4. Gradual notions of consistency in argumentation

In argumentation literature, several works have considered variations of the notion of consistency. In Weighted Argument Systems [36], the idea is to assign to each attack in an argumentation framework a positive real number representing its weight. This weight indicates intuitively the strength of the attack: a stronger attack suggests a higher level of inconsistency between the attacker and the attacked arguments.

This gives rise to the notion of *inconsistency budget*, namely the amount of inconsistency one is ready to tolerate in a set of arguments that are accepted altogether. Traditional Dung's semantics, requiring conflict-freeness, correspond to an inconsistency budget of zero, while this proposal allows the inclusion in the same extension, *i.e.*, a set of arguments labelled `in`, of some pairs of conflicting arguments as far the sum of the relevant attack weights does not exceed the inconsistency budget. This proposal differs from ours in two main respects. First, it refers to evaluating inconsistency in the attack weights rather than in the labels assigned to arguments. Second, the inconsistency budget represents a global constraint at the level of the argumentation framework, while our proposal focuses on local constraints at the level of pairs of arguments.

The approach in [36] can be regarded as orthogonal to ours and the study of possible integration between local and global inconsistency constraints represents an interesting direction for future work. A similar comment applies to fuzzy argumentation frameworks as defined in [37], where a degree is assigned to each attack and these degrees play a role in the definitions of parametric

semantics notions (like conflict-freeness and admissibility) based on fuzzy sets of arguments. In our context, this idea would correspond to considering a fuzzy intolerance relation, which is beyond the scope of the present paper and reserved for future work.

Forms of argumentation where gradual assessments are attached to arguments only can be put in more direct connection with our approach. For instance, in several literature proposals, arguments are labelled with a value in the $[0, 1]$ interval, with different possible interpretations like fuzzy acceptability degrees [38], epistemic probabilities [23], or strength [39]. Denoting as $v(\alpha)$ the value assigned to an argument α and letting α and β be two arguments such that α attacks β, a common consistency constraint considered in this context (called coherence in [23]) is that $v(\beta) \leq 1 - v(\alpha)$. In our terminology, this corresponds to a specific incompatibility relation on $[0, 1]$ where values x and y are incompatible whenever $x + y > 1$. As a small note, we observe that expressing this kind of constraint through numerical inequalities enforces the corresponding incompatibility relation to be symmetric, while our generic formulation also allows for asymmetric incompatibility in numerical settings, where appropriate.

Properties related to the idea of reinstatement are present in gradual settings too. In particular, in [39] it is shown that if a gradual semantics satisfies a set of basic principles (called Independence, Circumscription, Neutrality, and Maximality) then it satisfies a numerical counterpart (let us call it *n-reinstatement*) of the property stated in Definition 3.4, namely if for every attacker α of an argument β it holds that $v(\alpha) = 0$ then it must hold that $v(\beta) = 1$. The same property is also considered in [23]: an epistemic probability is said to be optimistic if for every argument β it holds that $v(\beta) \geq 1 - \sum_{\alpha \in \beta^-} v(\alpha)$. This condition, which directly implies n-reinstatement, refers collectively to the set of attackers.

As the above examples indicate, applying our approach to gradual argumentation shows promise of high potential and deserves to be investigated in future work. In particular, a study of the relationships with the rich corpus of principles for gradual argumentation [39,16,40,17] proposed in the literature is envisaged. Moreover, it can be remarked that the notions introduced in Section 2 do not make any commitment on the nature of the labelled entities and of the intolerance relation, hence applying the approach at the level of sets of arguments, as it may be required to capture properties like optimism in [23], does not pose specific problems and appears a natural development to pursue.

5.5. Consistency and reinstatement preservation

The issue of preserving desirable properties (like consistency or reinstatement) of some reasoning outcomes arises when a reasoning process consists of multiple stages where the outcomes of later stages are derived from those of the earlier ones. In particular, argumentative processes naturally lend themselves to a description in terms of reasoning stages or layers (see, for instance, [8,6]), and a relevant formal model called multi-labelling systems has been introduced in [18]. We have discussed consistency and reinstatement preservation concerning the stage of derivation of argument justification status, and, to the best of our knowledge, this kind of analysis has not been considered before in the argumentation literature.

It can be remarked, however, that the rationality postulates introduced in [25] can be regarded as introducing consistency preservation requirements concerning the assessment of the conclusions of arguments. In particular, the postulate of *direct consistency* requires that (i) for each labelling[11] the set of supported conclusions, namely the conclusions of arguments labelled `in`, is consistent and (ii) the set of justified conclusions, namely the intersection of the sets of supported conclusions of all labellings, is consistent. It is proved that, under suitable conditions, (i) implies (ii), which can be considered a form of consistency preservation when moving from individual labellings to skeptically accepted conclusions. The postulate of *indirect consistency* concerns the closure under strict rules of a set of conclusions S, namely the set obtained adding to S all the conclusions that can be further derived by applying rules which are certain and do not admit exceptions, *i.e.*, all the conclusions which follow necessarily from S. The postulate, analogously to the above one, requires that (i) for each labelling, the closure under strict rules of the set of supported conclusions is consistent, and (ii) the closure under strict rules of the set of justified conclusions is consistent. This requirement can be regarded as a form of consistency preservation concerning the stage of closure under strict rules. Moreover, also in this case it is proved that (i) implies (ii).

Our approach makes explicit and generalizes the notion of consistency preservation in abstract terms and complements it with the one of reinstatement preservation. In this context, we have provided some novel general results and analyzed the case of preservation in argument justification. Extending in general terms the study of property preservation to argument conclusions in the spirit of [25] requires further analyses and is an important direction of future work, which we plan to develop leveraging the multi-labelling approach introduced in [18] and taking into account the recent studies on claim-augmented argumentation frameworks [41–43].

5.6. Impossibility of reinstatement preservation

We have shown that reinstatement preservation is generally impossible, providing a counterexample related to the case of floating defeat. Conceptually this can be regarded as related to the impossibility result discussed in [44,45] where, rephrasing the result in our context, it is shown that equating skeptical acceptance of a conclusion to skeptical acceptance of at least an argument supporting it fails to provide the desired outcomes in the presence of floating defeat. Our result shows that problems appear at the argument justification level too, using any synthesis function satisfying some basic requirements.

Floating defeat and floating conclusions have been the subject of significant debates [46,29] concerning controversial examples from an intuitive point of view and then involving the underlying reasoning principles and modelling assumptions. In [46], it is

[11] The rationality postulates are defined in [25] with reference to the extension-based formulation of abstract argumentation semantics. We provide here an equivalent description in terms of labellings.

suggested that a possible way out of the controversies would be 'placing statements into several categories, depending on the degree to which they are supported by a set of premises, with floating conclusions then classified, not necessarily as unsupported, but perhaps only as less firmly supported than statements that are justified by the same argument in every extension.' In a similar vein, in [47], it is observed that 'some of the different consequence notions are not mutually exclusive but can be used in parallel, as capturing different senses in which belief in a proposition can be supported by a body of information' and it is remarked that 'in general the existence of different definitions is not a problem for, but a feature of the field of defeasible argumentation'. Considering explicitly a spectrum of notions of consistency and reinstatement and revising the notion of justification status, as proposed in this paper, is coherent with these indications and provides a basis for further developments.

While only weak reinstatement preservation can be achieved in general, there are cases where 'full' reinstatement preservation holds, and it will be interesting to investigate in future work some conditions under which the stronger property can be ensured. It is interesting to note that these conditions may either refer to the intolerance relation or the evaluation method. As to the former, we already commented that if the cardinality of the set of intolerant elements can be at most 1 (a very restrictive condition) 'full' and weak reinstatement preservation coincide. As to the latter, Proposition 4.8 has shown that faithfulness is a sufficient condition for assessments consisting of precisely one labelling (as in the case of grounded semantics for abstract argumentation). The identification of further less restrictive conditions on either side will be pursued in future work.

To introduce a weakened, generally viable notion of reinstatement preservation, we resorted to a transformation where an additional virtual element is included as a proxy of the set of intolerant elements when this is not a singleton. Transformations involving the addition of virtual arguments and relevant attacks have been considered in abstract argumentation (*e.g.*, to convert frameworks with higher-order attacks into conventional ones, as in [48,49]); however, to the best of our knowledge, the transformation we propose has no direct counterpart in the literature. In particular, while the virtual element corresponds to a set of attackers, its meaning is distinct from the notions of joint attack from a set of arguments [50], which, differently from ours, is based on a conjunctive interpretation, *i.e.*, requires that all elements of the set are accepted for the set-attack to be valid.

While conceived as a technical device for a specific problem, we suggest that virtual elements may also play a role in explaining the outcomes of argumentative assessments, a subject that is receiving increasing attention in the literature [51]. In argumentation frameworks, virtual elements provide a synthetic view of the status of the set of attackers across different labellings, thus providing a simple but effective justification of the status of the attacked argument. The investigation of the use of this kind of virtual explanatory gadgets would represent a novel complementary approach with respect to the various explanation means previously considered in the literature, like subgraphs [52,53], dialogues [54], and suitable framework changes [55].

5.7. Perspectives on further application domains

Abstract argumentation having been the main reference context to illustrate the use of our proposal in the paper, in this section we discuss some perspectives concerning its applicability in other domains. Exploring in detail these investigation directions is beyond the scope of this paper and is left to future work.

5.7.1. Qualitative decision making with conflicting attributes

In multiple attribute decision making [56] a decision maker has to make a choice in a set of options on the basis of the values of a set of attributes, which are weighted by the decision maker in terms of their importance. The illustrative example in Section 2 can be regarded as a (simplified and partial) instance of this family of problems.

In [57] a bipolar qualitative approach to decision making is considered, where each attribute may play a positive, negative, or neutral role with respect to an option and attribute importance is assessed on a qualitative scale. Assuming a scale with three levels of importance, like 'Not Important', 'Important', 'Very Important', the following set of labels $\Lambda = \{++, +, 0, -, --\}$ can be adopted, where the label $++$ ($--$) indicates a very important positive (negative) attribute, $+$ ($-$) indicates an important positive (negative) attribute, and 0 indicates a neutral attribute.

Letting A be a set of attributes the assessment expressed by a decision maker corresponds to a Λ-labelling of the set A. In an example taken directly from [57], the attributes considered in the selection for a travel destination are *Landscape*, *Tennis court*, *Swimming pool*, *Disco*, *Price*, *Airline reputation*, *Non democratic governance*. For instance the assessment of a travel option o_1 might consists of the labelling $L_1 = \{(L,0),(T,++),(S,0),(D,++),(P,-),(A,-)(N,0)\}$. In words, in the view of the decision maker, the option has two very important positive features (tennis and disco) and two mildly important negative features (price and airline reputation). Alternative options would be characterized by other labellings with the same structure. A variety of methods for deriving a decision from these labellings are discussed in [57].

Our approach could be applied to extend such a formal context in order to take into account explicitly the possible existence of conflicts between attributes, not encompassed by the proposal in [57]. In particular, one might however observe (as anticipated also in the example in Section 2) that, in presence of some kind of conflict between attributes, some labelling may have a dubious status. For instance, in the above example, one might argue that fully enjoying disco and tennis court in the same vacation is hardly possible. As a consequence, the labelling L_1 might be regarded as affected by some inconsistency and this might be taken into account in the decision process. Consider for instance another option o_2 whose assessment labelling is $L_2 = \{(L,0),(T,0),(S,++),(D,++),(P,-),(A,-)(N,0)\}$. The methods based on the order of magnitude of the importance of positive and negative features proposed in [57] would regard these options as equivalent, however, assuming that there is no conflict between enjoying disco and swimming pool, the second option should be regarded as preferable, which could be derived, for instance,

by assigning priority to more consistent labellings. We suggest that our approach, allowing a fine grained definition of various levels of inconsistency, would be suitable to support an articulated and flexible representation of this kind of enhanced decision criteria.

5.7.2. Multiwinner voting methods with constraints

A multiwinner election, also called committee voting, is an electoral system in which a collection of voters aims to elect multiple candidates, called a committee, from a larger set of available candidates [58]. A variety of voting methods can be adopted to this purpose, which may differ both in the way expressing votes and in the way the set winners is derived from the sef of expressed votes. To exemplify some alternatives concerning the former aspect, in multiwinner approval voting, each voter may select any number[12] of candidates, thus just distinguishing between approved and not approved candidates, while in ranked voting each voter expresses a linear preference on the set of candidates [59], and in cardinal voting each voter assigns independently to each candidate a grade on a given scale. In any case, each vote can be expressed as a labelling of the set of candidates, with a suitable set of labels.

In some contexts, it is desired that the election outcome satisfies some general principles, like gender equality or proper representation of minorities in the committee. To this purpose, some constraints can be imposed to voters. For instance, with reference to the goal of ensuring equality between two groups in a ranked voting system, in [58] the following constraints are considered: imposing that the rank is an alternation of the members of the two groups, imposing that the top k elements (where k is the number of seats in the committee) include a given proportion of the two groups, imposing that the top half of the rank include exactly the half of both groups. It can be observed that constraints of this kind are rather rigid, and do not seem easily generalizable to the case of cardinal voting. Our approach could be used to express diversity constraints in terms of consistency and reinstatement requirements in this context. In particular, assuming that the candidates belonging to the same group are related by the intolerance relation, a requirement of consistency may impose, for instance, that if a member of a given group receives the top rate by a voter, no other member of the same group can receive the top rate by the same voter. On the other hand, a reinstatement requirement may impose that at least a member of each group receives a rate at a given level.

The preservation of consistency and reinstatement properties from single votes to the final election outcome would be then an interesting subject, in line with the vast amount of studies concerning the properties of voting system, like e.g. [59,60]. In this respect it is worth noting that the class of synthesis functions considered in this paper, which is appropriate for the argumentation domain, does not appear to be suitable for the voting domain, as pure synthesis cannot encompass notions like counting and majority. Investigating other families of synthesis function, possibly borrowing ideas from social choice theory, is a very interesting direction of future work.

5.7.3. Legal reasoning

Legal reasoning appears to be another promising domain of application for the notions proposed in this paper. In particular, it is interesting to note that the evaluation criteria adopted in this domain may follow specific principles and may vary not only between the legal systems of different countries but also depending on the kind of legal procedure, e.g. a criminal trial vs. a civil trial.

For instance, the *in dubio pro reo* principle can be regarded as a specific form of reinstatement in favour of the presumption of innocence, which can be discarded only if there are sufficiently strong evidences against it. This shows that, in some reasoning contexts, different criteria can be applied to different entities (e.g. because some are favoured with respect to others), which suggests an interesting investigation direction and confirms the importance of having a tunable representation.

Proof standards [61] provide another significant example of potential connection with our proposal. Intuitively, a proof standard is a requirement to be met in order to accept an issue at stake, given a set of evidences. Different proof standards correspond to more or less demanding requirements and, consequently, to different consistency and reinstatement properties. In particular, the following four proof standards are described in [61], on the basis of [62], with reference to common law jurisdiction:

- *scintilla of evidence*: 'any evidence at all in a case, even a scintilla, tending to support a material issue';
- *preponderance of evidence*: 'evidence which as a whole shows that the fact sought to be proved is more credible and convincing to the mind';
- *clear and convincing evidence*: 'measure or degree of proof which will produce in mind of trier of facts a firm belief or conviction as to allegations sought to be established; it is intermediate, being more than preponderance, but not to extent of such certainty as is required beyond reasonable doubt';
- *beyond reasonable doubt*: 'requires evidence which leaves the trier of fact fully satisfied, entirely convinced, to a moral certainty'.

These informal descriptions can be given a formal counterpart in terms of arguments pro and con a given proposition, as done in [61]. From our perspective, it is interesting to note that they can be put in correspondence with different consistency and reinstatement properties. The *scintilla of evidence* standard gives up any consistency requirement: both a claim and its opposite can be evaluated positively by this standard, which does not weight or balance in any way conflicting reasons. In a sense, this corresponds to an extreme notion of reinstatement: there is no way to dismiss a conclusion, provided that there is some evidence (even very weak) supporting it. The *preponderance of evidence* standard instead, indicates that the decision is based on the evidence considered more credible. Thus, a claim can be dismissed in favour of its opposite if the evidence supporting the latter is considered to be stronger, even by a little amount. This can be regarded as pursuing a strong form of consistency, since a positive outcome on one side corresponds to a negative

[12] Variants of this scheme may impose constraints on this number, e.g. to be lesser or equal than the number of seats in the committee.

outcome on the other side, and to a weak form of reinstatement, since a conclusion can be dismissed by a small prevailing evidence on the opposite side. The *clear and convincing evidence* and *beyond reasonable doubt* can then be regarded as ways of trading off consistency and reinstatement, up to a certain degree, in the spirit of the *in dubio pro reo* principle. In both cases (but at different degrees) even if a claim is regarded as more credible than its opposite, the latter cannot be dismissed, thus imposing stronger reinstatement violation constraints (its rejection is more and more demanding) while partially sacrificing consistency.

Providing a detailed formal counterpart to the intuitions described above, along with investigating notions of (possibly partial) property preservation across different proof standards, represents another attractive line of future research, which may also stimulate the study of extensions and variations of the proposal presented in this paper.

5.7.4. Belief revision

In general terms, *Belief Revision* [63,64] studies the process of changing the content of a belief repository when considering the arrival of a new piece of information, possibly inconsistent with the previous beliefs. Assuming that the new piece of information has the priority, this calls for the retraction of some of the previous belief to restore consistency. Together with consistency, another basic principle informs the belief revision process, namely *minimal change*, which requires that as much as possible of the original information is preserved in the process. While our notion of reinstatement does not refer to a dynamic context and hence does not encompass a notion of initial state to be preserved, the underlying intuition of avoiding an unnecessarily restricted set of derivations is basically the same. Exploring the potential bridges between our proposal and the field of belief revision at large is far beyond the scope of the present paper. Here we limit us to some essential considerations.

In [63,64], two paradigmatic extreme cases of revision operator are considered, which are regarded as unsuitable for opposite reasons. Let K be the belief set to be revised and A be the new evidence acquired. At one extreme the revision operator based on *maxichoice contraction*[13] is 'too productive' since, after the revision, for every proposition B the agent either believes B or its negation, even if B was not included in K and is not related to A. At the other extreme, the operator based on *full meet contraction* is 'too restrictive' since, after the revision, the belief set of the agent coincides with the logical consequences of A: all the elements of K not derivable from A are no more believed, even those which were compatible with A. To avoid these extreme undesirable situations an intermediate form of contraction, called *partial meet contraction* has been identified and characterized. In this scenario, the need of tuning the outcomes of the reasoning process and the existence of a set of alternatives have a clear correspondence with the basic intuitions motivating our approach.

There are however also some significant differences. First, as already mentioned, in belief revision there is an initial state of belief that is the reference for minimal change. In particular, the postulate of *recovery* concerns the fact that the initial state is recovered if a proposition is added after having been retracted. The underlying intuition is rather different from our notion of reinstatement, in particular no intolerance relation is involved. Another remark concerns the fact that the 'tuning' of the outcomes of the belief revision process is driven by an ordering of importance of beliefs called *epistemic entrenchment* [64]. Epistemic entrenchment has intuitively to do with the explanatory power of initially held beliefs, a notion which has no direct counterpart in our model. Furthermore, it can be observed that in belief revision the focus is on indirect inconsistency, which suggests that a suitable notion of intolerance in this context should probably refer to sets of propositions.

In our opinion, these dissimilarities make particularly challenging, but also particularly interesting, the investigation of connections between our proposal and the field of belief revision. As possible starting points, we mention approaches to belief revision with multiple levels of credibility, like e.g. [65], and the formalism encompassing attacks at the level of sets of arguments proposed in [66].

6. Conclusions

In this paper we have introduced a novel domain-independent formalization of consistency and reinstatement as general properties of any labelling-based assessment produced by a reasoning process. As a demonstration of the approach's potential, we have illustrated its capability to provide an original characterization of several abstract argumentation semantics.

We have then investigated the issue of preserving these properties when a synthetic labelling is derived from other labellings. Using the synthesis of argument justification as an illustrative instance, we have obtained a general characterization of consistency preservation synthesis functions and provided an impossibility result for reinstatement preservation. This led us to investigate a weaker reinstatement preservation notion. Along this journey, we evidenced a limitation of the traditional notion of argument justification and proposed an improved version.

As discussed in Section 5, this work has multifaceted connections both with various investigation topics in formal argumentation and, more generally, with the study of various forms of reasoning whose outcomes can be modelled in terms of production of labellings, possibly in a multi-stage setting. A wide variety of potential future developments is therefore envisaged. In addition to those discussed in Section 5, we mention that one can consider complementary bounds also for the outcomes of learning processes (*e.g.*, a learned model should adhere to the training data but not be overfitting) and that different types of uncertainty have to be taken into account for them (like aleatoric and epistemic uncertainty [67,68]). It will then be interesting to explore the connections between these notions and our approach.

[13] A belief revision operator can be defined in terms of a contraction and an expansion operator, according to the Levi's identity. Recalling the relevant technical details is beyond the scope of this paper.

CRediT authorship contribution statement

Pietro Baroni: Conceptualization, Formal analysis, Funding acquisition, Investigation, Writing – original draft, Writing – review & editing. **Federico Cerutti:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Writing – original draft, Writing – review & editing. **Massimiliano Giacomin:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Writing – original draft, 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.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors are grateful to the associate editor and the anonymous reviewers for their helpful comments. This work was supported by MUR project PRIN 2022 EPICA 'Enhancing Public Interest Communication with Argumentation' (CUP D53D23008860006) funded by the European Union - Next Generation EU, mission 4, component 2, investment 1.1, and by the project Argumentation for Informed Decisions with Applications to Energy Consumption in Computing – AIDECC (CUP D53C24000530001), within the PNRR Future Artificial Intelligence – FAIR project (PE0000013, CUP H23C22000860006), Objective 10: Abstract Argumentation for Knowledge Representation and Reasoning, funded by the European Union - Next Generation EU.

References

[1] P. Baroni, F. Cerutti, M. Giacomin, A generalized notion of consistency with applications to formal argumentation, in: Proc. of the 9th Int. Conf. on Computational Models of Argument (COMMA 2022), 2022, pp. 56–67.
[2] P. Baroni, F. Cerutti, M. Giacomin, Generalizing consistency and reinstatement in abstract argumentation, in: R. Confalonieri, D. Porello (Eds.), Proc. of the 6th Workshop on Advances in Argumentation in Artificial Intelligence Co-Located with the 21st Int. Conf. of the Italian Association for Artificial Intelligence (AIxIA 2022), in: CEUR Workshop Proceedings, vol. 3354, CEUR-WS.org, 2022.
[3] P. Baroni, D. Gabbay, M. Giacomin, L. van der Torre (Eds.), Handbook of Formal Argumentation, College Publications, 2018.
[4] T.J.M. Bench-Capon, P.E. Dunne, Argumentation in artificial intelligence, Artif. Intell. 171 (10–15) (2007) 619–641.
[5] I. Rahwan, G.R. Simari (Eds.), Argumentation in Artificial Intelligence, Springer, Berlin, 2009.
[6] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G.R. Simari, M. Thimm, S. Villata, Towards artificial argumentation, AI Mag. 38 (3) (2017) 25–36.
[7] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games, Artif. Intell. 77 (2) (1995) 321–357.
[8] P. Baroni, M. Caminada, M. Giacomin, An introduction to argumentation semantics, Knowl. Eng. Rev. 26 (4) (2011) 365–410.
[9] M. Caminada, On the issue of reinstatement in argumentation, in: Proc. of the 10th European Conf. on Logics in Artificial Intelligence JELIA, in: Lecture Notes in Computer Science, vol. 4160, Springer, 2006, pp. 111–123.
[10] P. Baroni, M. Giacomin, On principle-based evaluation of extension-based argumentation semantics, Artif. Intell. 171 (10/15) (2007) 675–700.
[11] L. van der Torre, S. Vesic, The principle-based approach to abstract argumentation semantics, in: P. Baroni, D. Gabbay, M. Giacomin, L. van der Torre (Eds.), Handbook of Formal Argumentation, College Publications, 2018, pp. 797–837.
[12] A.J. García, G.R. Simari, Defeasible logic programming: an argumentative approach, Theory Pract. Log. Program. 4 (1–2) (2004) 95–138.
[13] H. Jakobovits, D. Vermeir, Robust semantics for argumentation frameworks, J. Log. Comput. 9 (2) (1999) 215–261.
[14] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artif. Intell. 195 (2013) 361–397.
[15] P. Baroni, M. Giacomin, B. Liao, Dealing with generic contrariness in structured argumentation, in: Proc. of the 24th Int. Joint Conf. on Artificial Intelligence, IJCAI 2015, 2015, pp. 2727–2733.
[16] L. Amgoud, J. Ben-Naim, D. Doder, S. Vesic, Acceptability semantics for weighted argumentation frameworks, in: Proc. of the 26th Int. Joint Conf. on Artificial Intelligence (IJCAI), 2017, pp. 56–62.
[17] P. Baroni, A. Rago, F. Toni, From fine-grained properties to broad principles for gradual argumentation: a principled spectrum, Int. J. Approx. Reason. 105 (2019) 252–286.
[18] P. Baroni, R. Riveret, Enhancing statement evaluation in argumentation via multi-labelling systems, J. Artif. Intell. Res. 66 (2019) 793–860.
[19] O. Arieli, Conflict-free and conflict-tolerant semantics for constrained argumentation frameworks, J. Appl. Log. 13 (4) (2015) 582–604.
[20] C. Cayrol, M.-C. Lagasquie-Schiex, Gradual acceptability in argumentation systems, in: Proc. of the 3rd Int. Workshop on Computational Models of Natural Argument (CMNA 2003), 2003, pp. 55–58.
[21] P. Baroni, M. Giacomin, G. Guida, Towards a formalization of skepticism in extension-based argumentation semantics, in: Proc. of the 4th Workshop on Computational Models of Natural Argument, CMNA 2004, 2004, pp. 47–52.
[22] Y. Wu, M. Caminada, A labelling-based justification status of arguments, Stud. Log. 3 (4) (2010) 12–29.
[23] A. Hunter, M. Thimm, Probabilistic reasoning with abstract argumentation frameworks, J. Artif. Intell. Res. 59 (2017) 565–611.
[24] D.M. Gabbay, Equational approach to argumentation networks, Argument Comput. 3 (2–3) (2012) 87–142.
[25] M. Caminada, L. Amgoud, On the evaluation of argumentation formalisms, Artif. Intell. 171 (2007) 286–310.
[26] M. Ulbricht, Understanding inconsistency - a contribution to the field of non-monotonic reasoning, Ph.D. thesis, University of Leipzig, 2019.
[27] J.L. Pollock, Defeasible reasoning, Cogn. Sci. 11 (4) (1987) 481–518.
[28] J.F. Horty, Argument construction and reinstatement in logics for defeasible reasoning, Artif. Intell. Law 9 (1) (2001) 1–28.
[29] H. Prakken, Intuitions and the modelling of defeasible reasoning: some case studies, in: S. Benferhat, E. Giunchiglia (Eds.), 9th International Workshop on Non-Monotonic Reasoning (NMR 2002), Proceedings, April 19-21, Toulouse, France, 2002, pp. 91–102.

[30] A. Urquhart, Many-valued logic, in: D. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic: Volume III: Alternatives in Classical Logic, Springer, Netherlands, Dordrecht, 1986, pp. 71–116.
[31] P. Baroni, M. Giacomin, G. Guida, SCC-recursiveness: a general schema for argumentation semantics, Artif. Intell. 168 (1–2) (2005) 162–210.
[32] W. Dvorák, S.A. Gaggl, Stage semantics and the SCC-recursive schema for argumentation semantics, J. Log. Comput. 26 (4) (2016) 1149–1202.
[33] R. Baumann, G. Brewka, M. Ulbricht, Shedding new light on the foundations of abstract argumentation: modularization and weak admissibility, Artif. Intell. 310 (2022) 103742.
[34] L. Amgoud, C. Cayrol, M. Lagasquie-Schiex, P. Livet, On bipolarity in argumentation frameworks, Int. J. Intell. Syst. 23 (10) (2008) 1062–1093.
[35] G. Brewka, S. Woltran, Abstract dialectical frameworks, in: Proc. of the 12th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR), 2010, pp. 102–111.
[36] P.E. Dunne, A. Hunter, P. McBurney, S. Parsons, M.J. Wooldridge, Weighted argument systems: basic definitions, algorithms, and complexity results, Artif. Intell. 175 (2) (2011) 457–486.
[37] J. Janssen, M.D. Cock, D. Vermeir, Fuzzy argumentation frameworks, in: Proc. of the 12th Int. Conf. on Information Processing and Management in Knowledge-Based Systems (IPMU 08), 2008, pp. 513–520.
[38] C. da Costa Pereira, A. Tettamanzi, S. Villata, Changing one's mind: erase or rewind?, in: Proc. of the 22nd International Joint Conf. on Artificial Intelligence (IJCAI), 2011, pp. 164–171.
[39] L. Amgoud, J. Ben-Naim, Axiomatic foundations of acceptability semantics, in: Proc.of the 15th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR), AAAI Press, 2016, pp. 2–11.
[40] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, A comparative study of ranking-based semantics for abstract argumentation, in: Proc. of the 30th AAAI Conference on Artificial Intelligence, 2016, pp. 914–920.
[41] W. Dvořák, A. Greßler, A. Rapberger, S. Woltran, The complexity landscape of claim-augmented argumentation frameworks, Artif. Intell. 317 (2023) 103873.
[42] W. Dvorák, A. Rapberger, S. Woltran, Argumentation semantics under a claim-centric view: properties, expressiveness and relation to setafs, in: Proc. of the 17th Int. Conf. on Principles of Knowledge Representation and Reasoning, KR, 2020, pp. 341–350.
[43] W. Dvorák, S. Woltran, Complexity of abstract argumentation under a claim-centric view, Artif. Intell. 285 (2020) 103290.
[44] K. Schlechta, Directly sceptical inheritance cannot capture the intersection of extensions, J. Log. Comput. 3 (5) (1993) 455–467.
[45] D. Makinson, K. Schlechta, Floating conclusions and zombie paths: two deep difficulties in the "directly skeptical" approach to defeasible inheritance nets, Artif. Intell. 48 (2) (1991) 199–209.
[46] J.F. Horty, Skepticism and floating conclusions, Artif. Intell. 135 (1–2) (2002) 55–72.
[47] H. Prakken, G.A.W. Vreeswijk, Logics for defeasible argumentation, in: D.M. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic, vol. 4, Springer, 2001, pp. 219–318.
[48] P. Baroni, F. Cerutti, M. Giacomin, G. Guida, AFRA: argumentation framework with recursive attacks, Int. J. Approx. Reason. 52 (1) (2011) 19–37.
[49] G. Boella, D.M. Gabbay, L.W.N. van der Torre, S. Villata, Meta-argumentation modelling I: methodology and techniques, Stud. Log. 93 (2–3) (2009) 297–355.
[50] A. Bikakis, A. Cohen, W. Dvorák, G. Flouris, S. Parsons, Joint attacks and accrual in argumentation frameworks, J. Appl. Log. 8 (6) (2021) 1437–1501, www.scopus.com.
[51] K. Cyras, A. Rago, E. Albini, P. Baroni, F. Toni, Argumentative XAI: a survey, in: Proc. of the 30th Int. Joint Conf. on Artificial Intelligence (IJCAI), 2021, pp. 4392–4399.
[52] X. Fan, F. Toni, On computing explanations in argumentation, in: Proc. of the 29th AAAI Conference on Artificial Intelligence, AAAI Press, 2015, pp. 1496–1502.
[53] Z.G. Saribatur, J.P. Wallner, S. Woltran, Explaining non-acceptability in abstract argumentation, in: Proc. of the 24th European Conf. on Artificial Intelligence (ECAI), 2020, pp. 881–888.
[54] E.I. Sklar, M.Q. Azhar, Explanation through argumentation, in: Proc. of the 6th Int. Conf. on Human-Agent Interaction (HAI), 2018, pp. 277–285.
[55] R. Booth, D.M. Gabbay, S. Kaci, T. Rienstra, L.W.N. van der Torre, Abduction and dialogical proof in argumentation and logic programming, in: Proc. of the 21st European Conference on Artificial Intelligence (ECAI), 2014, pp. 117–122.
[56] K.P. Yoon, C.-L. Hwang, Multiple Attribute Decision Making: an Introduction, Sage Publications, 1995.
[57] D. Dubois, H. Fargier, J. Bonnefon, On the qualitative comparison of decisions having positive and negative features, J. Artif. Intell. Res. 32 (2008) 385–417.
[58] S. Béal, M. Deschamps, M. Diss, R.T. Takeng, Multiwinner elections with diversity constraints on individual preferences, Working Papers 2024-07, CRESE, 2024, https://ideas.repec.org/p/crb/wpaper/2024-07.html.
[59] E. Elkind, P. Faliszewski, P. Skowron, A. Slinko, Properties of multiwinner voting rules, in: Proc. of the 13th Int. Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS), 2014, pp. 53–60.
[60] T.N. Tideman, F. Plassmann, Which voting rule is most likely to choose the "best" candidate?, Public Choice 158 (3) (2014) 331–357.
[61] T.F. Gordon, D. Walton, Proof burdens and standards, in: I. Rahwan, G.R. Simari (Eds.), Argumentation in Artificial Intelligence, Springer, 2009, pp. 239–258.
[62] H.C. Black, Black's Law Dictionary, West Publishing Co., 1979.
[63] C.E. Alchourrón, P. Gärdenfors, D. Makinson, On the logic of theory change: partial meet contraction and revision functions, J. Symb. Log. 50 (2) (1985) 510–530.
[64] P. Gärdenfors, Knowledge in Flux: Modeling the Dynamics of Epistemic States, MIT Press, 1988.
[65] M. Garapa, E. Fermé, M.D.L. Reis, System of spheres-based two level credibility-limited revisions, in: Proc. of the 19th Conf. on Theoretical Aspects of Rationality and Knowledge, TARK 2023, 2023, pp. 287–298.
[66] P. Baroni, M. Giacomin, B. Liao, A general semi-structured formalism for computational argumentation: definition, properties, and examples of application, Artif. Intell. 257 (2018) 158–207.
[67] S.C. Hora, Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management, Reliab. Eng. Syst. Saf. 54 (2) (1996) 217–223.
[68] E. Hüllermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods, Mach. Learn. 110 (3) (2021) 457–506, https://doi.org/10.1007/s10994-021-05946-3, arXiv:1910.09457.