Generalizing Consistency and Reinstatement in Abstract Argumentation

Pietro Baroni¹, Federico Cerutti^{1,2} and Massimiliano Giacomin^{1,*}

¹Department of Information Engineering (University of Brescia), Italy ²Cardiff University, UK

Abstract

We introduce two generic notions related to consistency and reinstatement in an abstract labelling setting, based on a relation of intolerance between the labelled elements and two specific relations, called incompatibility and reinstatement violation, between the labels assigned to them. This way, the approach allows a spectrum of consistency and reinstatement requirements depending on the actual choice of these relations. As a first application to formal argumentation, we show that traditional Dung's semantics can be expressed as combinations of different consistency and reinstatement requirements in this context.

Keywords

Consistency, Reinstatement, Argumentation semantics, Argument justification

1. Introduction

In formal argumentation, the presence of conflicts between arguments is a key aspect that calls for mechanisms able to produce sensible reasoning outcomes. In particular, these outcomes are typically required to satisfy two somewhat dual properties, i.e. those that have intuitively to do with the notion of consistency and those reflecting a requirement of completeness. For instance, in abstract argumentation semantics [1, 2] either extensions or labellings are typically required to satisfy the property of conflict-freeness, and it is also desired that arguments are accepted when all of their attackers are definitely outright, as indicated by the definition of complete extensions in [1] and the various notions of *reinstatement* introduced in [3].

In order to provide a common reference framework to bridge together the consistency notions considered in different formalisms and possibly investigate variations and developments thereof, a general formal treatment of consistency has been investigated in [4]. In this paper we complement this investigation by integrating a generalized notion of reinstatement.

The paper is organized as follows. Section 2 introduces the generalized notions of consistency and reinstatement, applicable in any context where a labelling approach is adopted. In particular, an intolerance relation indicates pairs of labelled elements that cannot stand each other, while

 A^{β} 2022: 6th Workshop on Advances in Argumentation in Artificial Intelligence

^{*}Corresponding author.

 [☆] pietro.baroni@unibs.it (P. Baroni); federico.cerutti@unibs.it (F. Cerutti); massimiliano.giacomin@unibs.it (M. Giacomin)

D 0000-0001-5439-9561 (P. Baroni); 0000-0003-0755-0358 (F. Cerutti); 0000-0003-4771-4265 (M. Giacomin)

^{© 0 2022} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CEUR Workshop Proceedings (CEUR-WS.org)

an incompatibility relation and a reinstatement violation relation between the labels are at the basis of the proposed notions of consistency and reinstatement. To show the application of the proposed concepts in the context of abstract argumentation, we prove in Section 3 that Dung's traditional semantics can be expressed as combinations of different consistency and reinstatement requirements, particularly with different incompatibility and reinstatement violation relations. The relationships of this work with previous literature and various perspectives of future development are finally discussed in Section 4.

2. Generalizing Consistency and Reinstatement for Labelling-based Systems

In a variety of contexts, the assessments of entities of various kind are expressed by assigning them a label. In order to provide a common ground to characterize such different contexts, in [4] we have introduced a three-layer model including 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 a positive assessment means 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 is considered that can be assessed by assigning each entity a label.

In the following, we introduce the notions of the model from the top level to the bottom level.

Definition 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 $c_1 \le c_2$ meaning that c_2 corresponds to an at least as positive assessment as c_1 . In the following we will mostly use a tripolar sac $C^3 = \{\text{pos}, \text{mid}, \text{neg}\}$ with $\text{neg} \le \text{mid} \le \text{pos}$ and the intuitive meaning that pos corresponds to a definitely positive assessment, neg to a definitely negative assessment, and mid to an intermediate situation. 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. 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 \leq where $\lambda_1 \leq \lambda_2$ iff $C_{\Lambda}(\lambda_1) \leq C_{\Lambda}(\lambda_2)$. As usual, $\lambda_1 < \lambda_2$ will denote $\lambda_1 \leq \lambda_2$ and $\lambda_2 \not\leq \lambda_1$

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

Proposition 1. Given a set of assessment classes C and a C-classified set of assessment labels Λ , the relation \leq as introduced in Definition 2 is reflexive and transitive, and for any $\lambda_1, \lambda_2 \in \Lambda$, $\lambda_1 \leq \lambda_2$ or $\lambda_2 \leq \lambda_1$.

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. $\lambda_1 \leq \lambda_1$. Similarly, if $\lambda_1 \leq \lambda_2$ and $\lambda_2 \leq \lambda_3$, by transitivity of \leq it holds that $C_{\Lambda}(\lambda_1) \leq C_{\Lambda}(\lambda_3)$, i.e. $\lambda_1 \leq \lambda_3$. Finally, since \leq is total it holds that $C_{\Lambda}(\lambda_1) \leq C_{\Lambda}(\lambda_2)$ or $C_{\Lambda}(\lambda_2) \leq C_{\Lambda}(\lambda_1)$, i.e. $\lambda_1 \leq \lambda_2$ or $\lambda_2 \leq \lambda_1$.

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

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 \leq_{Λ} .

The notion of labelling based on a sal is the usual one.

Definition 3. Given a sal Λ and a set S, a Λ -labelling of S is a function $L : S \to \Lambda$.

Different sals can be used to express assessments in distinct, but possibly related, evaluation contexts. For instance, in the context of argument acceptance evaluation based on the labelling-based version of Dung's semantics [1, 2], the sal $\Lambda^{IOU} = \{in, out, und\}$ is used, while in Defeasible Logic Programming (*DeLP*) arguments are marked as D(efeated) or U(ndefeated) corresponding to the use of the sal $\Lambda^{De} = \{D, U\}$, and in [5] an approach using the set of four labels $\Lambda^{JV} = \{+, -, \pm, \emptyset\}$ is proposed. We assume that the sals mentioned above are C^3 -classified as follows: $C^3_{\Lambda IOU} = \{(in, pos), (out, neg), (und, mid)\}; C^3_{\Lambda De} = \{(D, neg), (U, pos)\}; C^3_{\Lambda JV} = \{(-, neg), (+, pos), (\pm, mid), (\emptyset, mid)\}.$

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

- An inconsistency arises when two elements of a set which cannot stand each other are assigned labels which are 'too positive' altogether. 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 which cannot stand 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 stand each other.

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 stand 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 4. Given a set S, an intolerance relation on S is a binary relation int $\subseteq S \times S$, where $(s_1, s_2) \in$ int indicates that s_1 is intolerant of s_2 and will be denoted as $s_1 \odot s_2$, while $(s_1, s_2) \notin$ int will be denoted as $s_1 \ominus s_2$.

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

To exemplify, 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 [6, 7]. At the argument level, the attack relation in Dung's frameworks 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 5. Given a sac C, an incompatibility relation on C is a relation inc $\subseteq C \times C$, where $(c_1, c_2) \in$ inc indicates that c_1 is incompatible with c_2 and will be denoted as $c_1 \square c_2$, while $(c_1, c_2) \notin$ inc will be denoted as $c_1 \square c_2$. Given a C-classified sal Λ , we define the induced incompatibility relation inc' $\subseteq \Lambda \times \Lambda$ as follows: for every $\lambda_1, \lambda_2 \in \Lambda, (\lambda_1, \lambda_2) \in$ inc' iff $(C_{\Lambda}(\lambda_1), C_{\Lambda}(\lambda_2)) \in$ inc. With a little abuse of notation we will also denote $(\lambda_1, \lambda_2) \in$ inc' as $\lambda_1 \square \lambda_2$, and analogously for $\lambda_1 \square \lambda_2$.

Definition 6. Given a sac C, a reinstatement violation relation on C is a relation $rv \subseteq C \times C$, where $(c_1, c_2) \in rv$ indicates that c_1 is not sufficiently positive to justify c_2 and will be denoted as $c_1 \square c_2$, while $(c_1, c_2) \notin rv$ will be denoted as $c_1 \square c_2$. Given a C-classified sal Λ , we define the induced reinstatement violation relation $rv' \subseteq \Lambda \times \Lambda$ as follows: for every $\lambda_1, \lambda_2 \in \Lambda$, $(\lambda_1, \lambda_2) \in rv'$ iff $(C_{\Lambda}(\lambda_1), C_{\Lambda}(\lambda_2)) \in rv$. With a little abuse of notation we will also denote $(\lambda_1, \lambda_2) \in rv'$ as $\lambda_1 \square \lambda_2$, and analogously for $\lambda_1 \square \lambda_2$.

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 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:

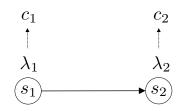


Figure 1: Two elements $s_1, s_2 : s_1 \odot s_2$. The intolerance relation is graphically represented by an arrow.

- inc is monotonic, i.e. given $c_1, c_2 \in C$ such that $c_1 \underline{\Box} c_2$, 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 $c'_1 \underline{\Box} c'_2$
- inc is non empty, i.e. $inc \neq \emptyset$
- $\forall c_1 \in C, \exists c_2 \in C \text{ such that } c_1 \sqsubseteq c_2 \text{ and } \exists c_3 \in C \text{ such that } c_3 \sqsubseteq c_1$

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 Figure 1, 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 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, $c_1 \square c_2$ indicates that the simultaneous positiveness of the 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 $c'_1 \square c'_2$.

The second property requires at least a 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 much weak.

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 $\min(C) \sqsubseteq \min(C)$, $\max(C) \trianglerighteq \min(C)$ and $\min(C) \trianglerighteq \max(C)$.

Proposition 2. *Given a sac C, let inc be a well-founded incompatibility relation on C. It then holds that:*

- $\max(C) \subseteq \max(C)$
- $\exists c \in C \text{ such that } c \subseteq \min(C) \text{ or } \min(C) \subseteq c$

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

As to the second property, assume by contradiction that $\exists c \in C$ such that $c \subseteq \min(C)$. By the monotonicity property and the definition of $\min(C)$, $\forall c' \in C$ it holds that $c \subseteq c'$, violating the third condition of Definition 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 $\underline{inc}_C = \{(\max(C), \max(C))\}$.

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

Definition 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, *i.e.* given $c_1, c_2 \in C$ such that $c_1 \overline{\Box} c_2$, 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 $c'_1 \overline{\Box} c'_2$
- rv is non empty, i.e. $rv \neq \emptyset$
- $\forall c_1 \in C, \exists c_2 \in C \text{ such that } c_1 \overline{\boxminus} c_2 \text{ and } \exists c_2 \in C \text{ such that } c_2 \overline{\boxminus} c_1$

In order to provide an explanation of these requirements, let us refer again to the simple case depicted in Figure 1.

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 Figure 1, $c_1 \square c_2$. 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 $c'_1 \le c_1$ (i.e. positiveness of λ_1 does not increase) and $c'_2 \le c_2$ (i.e. positiveness of λ_2 does not increase), then it must also hold $c'_1 \overline{\Box} c'_2$.

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 same condition appearing in Definition 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 elements¹. Second, the maximally positive label is compatible with any other label, in particular $\max(C) \equiv \max(C)$, $\max(C) \equiv \min(C)$ and $\min(C) \equiv \max(C)$.

Proposition 3. *Given a sac C, let rv be a well-founded reinstatement violation relation on C. It then holds that:*

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

¹This refers to the simple case of Figure 1, while the general case is handled according to Definition 10.

Proof: As to the first property, since $rv \neq \emptyset$ 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 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 nonempty reinstatement violation relation as $\underline{rv}_{C} = \{(\min(C), \min(C))\}$.

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 from our generalized notion of inconsistency of a labelling.

Definition 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

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

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

$$\forall s_1, s_2 \in S \text{ such that } s_1 \odot s_2, \text{ it holds that } L(s_1) \underline{\square} L(s_2)$$
(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 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'-inconsistent, and an int-inc'-consistent Λ -labelling L of S is also int-inc'-inconsistent.

Turning to reinstatement violation, duality w.r.t. inconsistency is reflected also in the counterpart of Definition 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

$$\exists s_2 : \forall s_1 \in S \text{ such that } s_1 \odot s_2 \text{ it holds that } L(s_1) \overline{\boxdot} L(s_2)$$
(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*² 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

 $^{^{2}}$ We borrow the terminology from abstract argumentation, where initial nodes are those without attackers.

(3) and never satisfy condition (4), entailing that no labelling is able to satisfy reinstatement whenever there are initial elements in *S*.

A special 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 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 from the way labels are selected for non initial elements.

We are thus lead 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 initial elements can be considered:

1.
$$\{\lambda \in \Lambda \mid \min(C) \boxtimes C_{\Lambda}(\lambda)\}$$

2.
$$\{\lambda \in \Lambda \mid \forall c \in C, c \exists C_{\Lambda}(\lambda)\}$$

Intuitively, according to the first option initial elements are equated to non initial elements where intolerant elements w.r.t. 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 such that $s_1 \odot s_2$, 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 minimal acceptance degree.

The second option allows for initial nodes all those labels that would be allowed whatever 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 prevent only the labels that would be prevented by the reinstatement violation relation whatever the reason against their acceptance, i.e. when there are intolerant elements whatever their labels.

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 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 $\min(C) \boxtimes C_{\Lambda}(\lambda)$. By the definition of minimum, $\forall c \in C, \min(C) \leq c$. If by contradiction $c \boxtimes C_{\Lambda}(\lambda)$ then by dual monotonicity of rv it would be the case that $\min(C) \boxtimes C_{\Lambda}(\lambda)$, contradicting the hypothesis that $\min(C) \boxtimes C_{\Lambda}(\lambda)$.

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

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

Definition 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

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

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

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

A corresponding result w.r.t. Proposition 4 holds.

Proposition 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'-uncompliant, and an int-rv'-compliant Λ -labelling L of S is also int-rv-uncompliant.

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 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 9 so as to enforce a specific treatment for initial elements, similarly to int-rv-compliant labellings (see Definition 10). Taking into account the intuition behind consistency, the following two options for the possible labels of initial elements can be considered:

1.
$$\{\lambda \in \Lambda \mid \min(C) \sqsubseteq C_{\Lambda}(\lambda)\}$$

2.
$$\{\lambda \in \Lambda \mid \exists c \in C : c \sqsubseteq C_{\Lambda}(\lambda)\}$$

Similarly 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. Again, the absence of intolerant elements (and thus of any

simultaneous positivity) is considered equivalent to intolerant elements 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 is allowed in case there is an intolerant element, whatever its label, must also be allowed for initial elements. It is easy to see that, adopting 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 there is no $c \in C$ such that min $(C) \square c$, entailing that $\forall \lambda \in \Lambda$ min $(C) \square C_{\Lambda}(\lambda)$. Of course, the second option enforces a weaker constraints w.r.t. the first one, as it is evident by considering $c = \min(C)$, thus it must allow as the first option all of the labels in Λ . Summing up, explicitly considering initial elements would not bring any modification to Definition 9, which thus conceptually corresponds to the dual counterpart of Definition 10. From a theoretical perspective, these considerations support the well-foundedness and generality of both definitions.

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. In particular, different combinations of (in)consistency and reinstatement (violation) notions are able to capture different argumentation semantics in Dung's framework, as discussed next.

3. Consistency and reinstatement properties in argumentation semantics

As well-known, in abstract argumentation an argumentation semantics 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), between a set \mathscr{A} of arguments. A set of arguments and the relevant attack relation are modelled by the traditional notion of argumentation framework [1].

Definition 11. 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 criterion*, as well as some of its variants, has been made explicit in [3]: 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 arguments labellings instead of extensions, i.e. as

assignments of labels, taken from a given set, to the set of arguments \mathscr{A} . Using the set of three labels Λ^{IOU} a correspondence can be drawn between extensions and labellings, while in general the labelling-based approach is more expressive than the extension-based approach.

Combining the generalized notions of consistency and reinstatement with three-valued labellings enables 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_{AIOU} introduced above. 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 [2] showing that they can be expressed as combinations of specific instances of generalized consistency and reinstatement properties.

The simplest semantics notion is the one of conflict-freeness, recalled in Definition 12.

Definition 12. 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* $L(\alpha) = \text{in then } \nexists \beta \in \alpha^- : L(\beta) = \text{in}$
- 2. *if* $L(\alpha) = \text{out then } \exists \beta \in \alpha^- : L(\beta) = \text{in}$

Item 1 in Definition 12 corresponds exactly 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: $rv_{C^3}^{cf} = \{(\text{neg}, \text{neg}), (\text{mid}, \text{neg})\}$.

Proposition 7. Let *L* be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. Then, *L* is $\rightarrow -rv_{C^3}^{cf}$ -compliant iff for each $\alpha \in \mathcal{A}$ it holds that if $L(\alpha) = \text{out then } \exists \beta \in \alpha^- : L(\beta) = \text{in.}$

Proof: Let L be $\rightarrow -\operatorname{rv}_{C^3}^{cf}$ -compliant and assume by contradiction that there is an argument $\alpha \in \mathscr{A}$ such that $L(\alpha) = \operatorname{out}$ and $\forall \beta \in \alpha^- L(\beta) \neq \operatorname{in}$. If α is initial, according to Definition 10 it must be the case that $\min(C) \boxplus C_{\Lambda}(L(\alpha))$, i.e. taking into account the definition of $\operatorname{rv}_{C^3}^{cf}$ it holds that $C_{\Lambda}(L(\alpha)) \in \{\operatorname{mid}, \operatorname{pos}\}$, which contradicts $L(\alpha) = \operatorname{out}$. If α is not initial, according to Definition 10 it holds that $\exists \beta \in S : \beta \odot \alpha$ (i.e. $\beta \in \alpha^-$) and $L(\beta) \boxplus L(\alpha)$, i.e. taking again into account the definition of $\operatorname{rv}_{C^3}^{cf}$ and the fact that $L(\alpha) = \operatorname{out}$ we have that $\exists \beta \in S : \beta \in \alpha^-$ and $L(\beta) \in \{\operatorname{in}\}$, contradicting the initial assumption that $\forall \beta \in \alpha^- L(\beta) \neq \operatorname{in}$.

As to the reverse direction, assume that for each $\alpha \in \mathscr{A}$ if $L(\alpha) = \text{out then } \exists \beta \in \alpha^- : L(\beta) = \text{in}$, and assume by contradiction that L is $\rightarrow -\text{rv}_{C^3}^{cf}$ -uncompliant. According to Definition 10, at least one of the following two cases holds.

- 1. There is $\alpha \in \mathscr{A}$ such that α is initial and $\min(C) \overline{\boxdot} C_{\Lambda}(L(\alpha))$. Taking into account the definition of $\operatorname{rv}_{C^3}^{cf}$, this entails $L(\alpha) = \operatorname{out}$. By the initial assumption $\exists \beta \in \alpha^- : L(\beta) = \operatorname{in}$, 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 $s_1 \in \alpha^-$ it holds that $L(\beta)\overline{\boxdot}L(\alpha)$. Taking into account the definition of $\operatorname{rv}_{C^3}^{cf}$, it must be the case that $L(\alpha) = \operatorname{out}$ and $\forall \beta \in S$ such that $\beta \in \alpha^-$, $L(\beta) \in \{\operatorname{out}, \operatorname{und}\}$. This contradicts the assumption that $\exists \beta \in \alpha^- : L(\beta) = \operatorname{in}$.

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

Proposition 8. The set of conflict-free labellings coincides with the set of labellings which are $\rightarrow -\underline{inc}_{C^3}$ -consistent and $\rightarrow -rv_{C^3}^{cf}$ -compliant.

Proof: The proof is immediate taking into account the correspondence between Item 1 in Definition 12 and consistency under \underline{inc}_{C^3} , as well as correspondence between Item 2 and reinstatement compliance under $rv_{C^3}^{cf}$ ensured by Proposition 7.

Admissibility of a set of arguments was introduced in [1] 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 13.

Definition 13. 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* $L(\alpha) = \text{ in then } \forall \beta \in \alpha^- : L(\beta) = \text{ out }$
- 2. *if* $L(\alpha) = \text{out } then \exists \beta \in \alpha^- : L(\beta) = \text{in}$

Item 1 in Definition 13 is a strengthening of item 1 of Definition 12, while item 2 is the same in both Definition 12 and 13. Interestingly, this strengthening 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 $\operatorname{inc}_{C^3}^a = \{(\operatorname{pos}, \operatorname{pos}), (\operatorname{mid}, \operatorname{pos})\}.$

Proposition 9. The set of admissible labellings coincides with the set of labellings which are $\rightarrow -inc_{C3}^{a}$ -consistent and $\rightarrow -rv_{C3}^{cf}$ -compliant.

Proof: It has been shown in [4] that admissible labellings correspond to the set of conflict-free labellings that are \rightarrow -inc^{*a*}_{*C*³}-consistent. Then the conclusion easily follows from Proposition 8. \Box

Completeness of a set of arguments was introduced in [1] 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 14.

Definition 14. 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 $L(\alpha) = \text{und then } \nexists \beta \in \alpha^- : L(\beta) = \text{in and } \exists \beta \in \alpha^- : L(\beta) = \text{und}$

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

Definition 15. A labelling *L* satisfies the reinstatement property if $\forall \alpha \in \mathcal{A}$ it holds that if $\forall \beta \in \alpha^ L(\beta) = \text{out then } L(\alpha) = \text{in.}$

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

Proposition 10. Let *L* be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. Then, *L* is \rightarrow - $rv_{C^3}^{\epsilon}$ -compliant iff it satisfies the reinstatement property and for each $\alpha \in \mathcal{A}$ it holds that if $L(\alpha) = \text{out then } \exists \beta \in \alpha^- : L(\beta) = \text{in.}$

Proof: Assume that L is $\rightarrow -\operatorname{rv}_{C^3}^c$ -compliant. By Proposition 6, L is $\rightarrow -\operatorname{rv}_{C^3}^{cf}$ -compliant, and by Proposition 7 the second condition of the thesis holds. To prove the reinstatement property, let us consider an argument $\alpha \in \mathcal{A}$ such that $\forall \beta \in \alpha^-$, $L(\beta) = \operatorname{out}$, and let us show that $L(\alpha) = \operatorname{in}$. If α is initial, by Definition 10 it must be the case that $\min(C) \overline{\boxminus} C_{\Lambda}(L(\alpha))$, i.e. taking into account the definition of $\operatorname{rv}_{C^3}^c$ we have that $C_{\Lambda}(L(\alpha)) = \operatorname{pos}$, which holds iff $L(\alpha) = \operatorname{in}$. If α is non initial, by Definition 10 there is an argument $\beta \in \alpha^-$ such that $L(\beta) \overline{\boxminus} L(\alpha)$. Since by the hypothesis $L(\beta) = \operatorname{out}$, according to the definition of $\operatorname{rv}_{C^3}^c$ it must be the case that $L(\alpha) = \operatorname{in}$.

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 $\rightarrow \operatorname{rv}_{C^3}^c$ -compliant. According to Definition 10, we can consider an argument α and distinguish two cases for it. If α is initial, by the reinstatement property $L(\alpha) = \operatorname{in}$, thus $C_{\Lambda}(L(\alpha)) = \operatorname{pos}$ which satisfies the required condition $\min(C) \boxtimes C_{\Lambda}(L(\alpha))$. If α is not initial, referring to Definition 10 assume by contradiction that there is no $\beta \in \alpha^-$ such that $L(\beta) \boxtimes L(\alpha)$. This means that $\forall \beta \in \alpha^-$, $L(\beta) \boxtimes L(\alpha)$, i.e. $(L(\beta), L(\alpha)) \in \{(\operatorname{out}, \operatorname{out}), (\operatorname{out}, \operatorname{und}), (\operatorname{und}, \operatorname{out})\}$. According to the second condition of the hypothesis if $L(\alpha) = \operatorname{out}$ then $\exists \beta \in \alpha^- : L(\beta) = \operatorname{in}$, which entails that $L(\alpha) \neq \operatorname{out}$. Then $L(\alpha) = \operatorname{und} \operatorname{and} \forall \beta \in \alpha^- L(\beta) = \operatorname{out}$, contradicting the reinstatement property.

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

Proposition 11. The set of complete labellings coincides with the set of labellings which are \rightarrow -inc^c_{C3}-consistent and \rightarrow -rv^c_{C3}-compliant.

Proof: It is shown in [4] that complete labellings coincide with admissible labellings that are \rightarrow -inc^c_{C³}-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 $L(\alpha) =$ out then $\exists \beta \in \alpha^- : L(\beta) =$ in, entailing by Proposition 10 that it is \rightarrow -rv^c_{C³}-compliant. Conversely, if a labelling is \rightarrow -inc^c_{C³}-consistent and \rightarrow -rv^c_{C³}-compliant then by Proposition 4 and Proposition 6 it is also \rightarrow -inc^a_{C³}-consistent and \rightarrow -rv^c_{C³}-compliant, and thus admissible by Proposition 9. Moreover, by Proposition 10 it satisfies the reinstatement property. As a consequence, taking into account the aforementioned result from [4], the labelling is complete. \Box

Stability of a 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 16.

Definition 16. Let *L* be a labelling of an argumentation framework $AF = (\mathcal{A}, \rightarrow)$. *L* is stable if it is complete and $\nexists \alpha \in \mathcal{A} : L(\alpha) =$ 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 12. The set of stable labellings coincides with the set of labellings which are $\rightarrow -\overline{inc}_{C^3}$ consistent and $\rightarrow -rv_{C^3}^c$ -compliant.

Proof: It is shown in [4] that stable labellings coincide with complete labellings which are $\rightarrow -\overline{\text{inc}}_{C^3}$ -consistent. Then the conclusion follows from Proposition 11.

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 further strengthenings of generalized consistency and reinstatement characterize complete and stable labellings.

4. Discussion and conclusions

In this paper, we have introduced the generalized notions of consistency and reinstatement encompassing Dung's traditional semantics. To our knowledge, providing a generalized form of the notion of consistency has only been considered in [4], whereof this paper represents an extension which integrates the notion of reinstatement. Some complementary research direction has been previously pursued concerning the notion of consistency, e.g. encompassing some inconsistency tolerance in the semantics of weighted argumentation systems [8], introducing the notion of conflict-tolerant semantics [9], measuring inconsistency in (abstract and structured) argumentation formalisms [10]. On the one hand, drawing correspondences between our approach and these proposals is an interesting future work. On the other hand, in the light of our general model it seems reasonable to apply these research directions also to reinstatement.

Several other research work can be envisaged. First of all, in [4] we have considered the issue of consistency preservation when a labelling is obtained as a synthesis of a set of labellings. Extending the relevant results to reinstatement preservation is a first step of future research.

Extending the analysis beyond tripolar classifications is another important future development. For example, more articulated notions of argument justification have been considered in the literature [11, 12, 13]. Dealing with consistency and reinstatement in such a context might require considering different sets of assessment classes and defining a notion of refinement between them. Addressing the evaluation of argument conclusions and their consistency is a further key step. In particular, it would be interesting to extend the notions presented in this paper to the formalism of multi-labelling systems [14], which can capture a variety of approaches to derive the assessment of conclusions from the assessment of arguments.

Acknowledgments

This work has been partially supported by the research project GNCS-INdAM CUP E55F22000270001, "Verifica Formale di Dibattiti nella Teoria dell'Argomentazione".

References

- [1] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games, Artif. Intell. 77 (1995) 321–357.
- [2] P. Baroni, M. Caminada, M. Giacomin, An introduction to argumentation semantics, Knowledge Engineering Review 26 (2011) 365–410.
- [3] P. Baroni, M. Giacomin, On principle-based evaluation of extension-based argumentation semantics, Artif. Intell. 171 (2007) 675–700.
- [4] 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.
- [5] H. Jakobovits, D. Vermeir, Robust semantics for argumentation frameworks, J. of Logic and Computation 9 (1999) 215–261.
- [6] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artif. Intell. 195 (2013) 361 – 397.
- [7] 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.
- [8] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, M. J. Wooldridge, Weighted argument systems: Basic definitions, algorithms, and complexity results, Artif. Intell. 175 (2011) 457–486.
- [9] O. Arieli, Conflict-free and conflict-tolerant semantics for constrained argumentation frameworks, J. Appl. Log. 13 (2015) 582–604.
- [10] A. Hunter, Measuring inconsistency in argument graphs, CoRR abs/1708.02851 (2017). URL: http://arxiv.org/abs/1708.02851. arXiv:1708.02851.
- [11] Y. Wu, M. W. A. Caminada, A labelling-based justification status of arguments, Studies in Logic 3 (2010) 12–29.
- [12] W. Dvořák, On the complexity of computing the justification status of an argument, in: S. Modgil, N. Oren, F. Toni (Eds.), Proc. of the 1st Int. Workshop on Theory and Applications of Formal Argumentation (TAFA 2011), number 7132 in Lecture Notes in Computer Science, Springer, 2012, pp. 32–49.
- [13] P. Baroni, M. Giacomin, G. Guida, Towards a formalization of skepticism in extensionbased argumentation semantics, in: Proc. of the 4th Workshop on Computational Models of Natural Argument (CMNA 2004), 2004, pp. 47–52.
- [14] P. Baroni, R. Riveret, Enhancing statement evaluation in argumentation via multi-labelling systems, J. Artif. Intell. Res. 66 (2019) 793–860.