TPLP: Page 1–24.
doi:10.1017/S1471068422000114

*Abstract Argumentation and Answer Set Programming: Two Faces of Nelson's Logic**

JORGE FANDINNO
University of Nebraska Omaha, Omaha, NE, USA
(*e-mail:* jfandinno@unomaha.edu)

LUIS FARIÑAS DEL CERRO
IRIT, Université de Toulouse, CNRS, Toulouse, France
(*e-mail:* luis@irit.fr)

submitted 18 December 2020; revised 21 December 2022; accepted 7 April 2022

Abstract

In this work, we show that both logic programming and abstract argumentation frameworks can be interpreted in terms of Nelson's constructive logic N4. We do so by formalising, in this logic, two principles that we call noncontradictory inference and strengthened closed world assumption: the first states that no belief can be held based on contradictory evidence while the latter forces both unknown and contradictory evidence to be regarded as false. Using these principles, both logic programming and abstract argumentation frameworks are translated into constructive logic in a modular way and using the object language. Logic programming implication and abstract argumentation supports become, in the translation, a new implication connective following the noncontradictory inference principle. Attacks are then represented by combining this new implication with strong negation. *Under consideration in Theory and Practice of Logic Programming (TPLP).*

KEYWORDS: logic programming, abstract argumentation, equilibrium logic, constructive logic

1 Introduction

Logic programming (LP) and Abstract Argumentation Frameworks (AFs) are two well-established formalisms for Knowledge Representation and Reasoning (KR) whose close relation is well-known since the introduction of the latter: besides introducing AFs, Dung (1995) studied how logic programs under the *stable models* (Gelfond and Lifschitz 1988) and the *well-founded semantics* (Van Gelder *et al.* 1991) can be translated into abstract argumentation frameworks. Since then, this initial connection has been further studied and extended, providing relations between other semantics and ways to translate argumentation frameworks into logic programs (Nieves *et al.* 2008; Caminada

* This work is an extended version of a paper presented at the Sixteenth International Conference on Principles of Knowledge Representation and Reasoning and entitled *Constructive Logic Covers Argumentation and Logic Programming* (Fandinno and Fariñas del Cerro 2018). We are thankful to Seiki Akama, Pedro Cabalar, Marcelo Coniglio, David Pearce, Newton Peron, and Agustín Valverde for their suggestions and comments on earlier versions of this work. We also thank the anonymous reviewers of the Sixteenth International Conference on Principles of Knowledge Representation and Reasoning for their comments on a preliminary version of this work.

and Gabbay 2009; Wu and Caminada 2010; Toni and Sergot 2011; Dvorák *et al.* 2011; Caminada *et al.* 2015).

On the other hand, Nelson's *constructive logic* (Nelson 1949) is a conservative extension of *intuitionistic logic*, which introduces the notion of *strong negation* as a means to deal with constructive falsity, in an analogous way as intuitionism deals with constructive truth. Pearce (1996; 2006) showed that a particular selection of models of constructive logic, called *equilibrium logic*, precisely characterize the stable models of a logic program. This characterization was later extended to the *three-valued stable model* (Przymusinski 1991) and the well-founded semantics by Cabalar *et al.* (2007). Versions of constructive logic without the "explosive" axiom $\varphi \rightarrow (\sim\varphi \rightarrow \psi)$ have been extensively studied in the literature (Nelson 1959; López-Escobar 1972; Thomason 1969; Almukdad and Nelson 1984; Odintsov 2005; Odintsov and Rybakov 2015; Kamide and Wansing 2015) and can be considered a kind of *paraconsistent* logics, in the sense, that some formulas may be constructively true and false at the same time. The notion of equilibrium has been extended to one of these logics by Odintsov and Pearce (2005), who also showed that this precise characterize the *paraconsistent stable semantics* (Sakama and Inoue 1995).

In this paper, we formalize in Nelson's constructive logic a reasoning principle, to be called *non-contradictory inference* (denoted **NC**), which states that

NC "*no belief can be held based on contradictory evidence.*"

Interestingly, though different from the logic studied by Odintsov and Pearce, the logic presented here is also a conservative extension of equilibrium logic (and, thus, also of LP under the stable models semantics) that allows us to deal with inconsistent information in LP. The interesting feature of this new logic is that, besides LP, it also captures several classes of AFs, under the stable semantics. It is worth to mention that the representation of AFs in this new logic is modular and it is done using an *object language level*. Recall that by object language level, we mean that AFs and its logical translation *share the same language* (each argument in the AF becomes an atom in its corresponding logical theory) and the relation between arguments in the AF (attacks or supports) are expressed by means of logical connectives. This contrast with *meta level approaches*, which talk about the AFs from "above," using another language and relegating logic to talk about this new language. It is important to note that, as highlighted by Gabbay and Gabbay (2015), the object language oriented approaches have the remarkable property of providing alternative intuitive meaning to the translated concepts through their interpretation in logic. In this sense, from the viewpoint of constructive logic, AFs can be understood as a *strengthened closed world assumption* (Reiter 1980) that we denote as **CW**:

CW "*everything for which we do not have evidence of being true or for which we have contradictory evidence, should be regarded as false*"

The relation between AFs and logic has been extensively studied in the literature and, as mentioned above, can be divided in two categories: those that follow an object language approach (Caminada and Gabbay 2009; Gabbay and Gabbay 2015; 2016) and those that follow a meta-level approach (Besnard and Doutre 2004; Caminada and Gabbay 2009; Grossi 2011; Dvorák *et al.* 2012; Arieli and Caminada 2013; Doutre *et al.* 2014; Besnard *et al.* 2014; Dvorák *et al.* 2015). In particular, the approach we take here shares with the work by Gabbay and Gabbay (2015) the use of strong negation to capture attacks, but differs in the underlying logic: constructive logic in our case and classical logic in the case of Gabbay and Gabbay's work. On the intuitive level, under the constructive logic point of view, *attacks* can be understood as

AT "*means to construct a proof of the falsity of the attacked argument based on the acceptability of the attacker*"

On the practical level, the use of constructive logic allows for a more *compact* and *modular translation*: each attack becomes a (rule-like) formula with the attacker – or a conjunction of attackers in the case of set attacking arguments (Nielsen and Parsons 2007) – as the antecedent and the attacked argument as the consequent. Moreover, when attacks are combined with LP implication, we show that the latter captures the notion of *support* in Evidential-Based Argumentation Frameworks (EBAFs; Oren and Norman 2008): for accepting an argument, these frameworks require, not only its *acceptability* as in Dung's sense, but also that it is supported by some chain of supports rooted in a kind of special arguments called *prima-facie*.

2 Background

In this section, we recall the needed background regarding Nelson's constructive logic, logic programming, and argumentation frameworks.

2.1 Nelson's constructive logic

The concept of constructive falsity was introduced into logic by Nelson (1949) and it is often denoted as **N3**. It was first axiomatized by Vorob'ev (1952), and later studied by Markov (1953), who related intuitionistic and strong negation, and by Rasiowa (1969), who provided an algebraic characterization. Versions of constructive logic without the "explosive" axiom $\varphi \rightarrow ({\sim}\varphi \rightarrow \psi)$ are usually denoted as **N4** and they are based on a four valued assignment for each world corresponding to the values *unknown*, *(constructively) true*, *(constructively) false* and *inconsistent* (or *overdetermined*). The logic **N3** can be obtained by adding back the "explosive" axiom. We describe next a Kripke semantics for a version of **N4** (Thomason 1969; Gurevich 1977) with the falsity constant $\bot$, which is denoted as $\mathbf{N4}^{\bot}$ by Odintsov and Rybakov (2015). We follow here an approach with two forcing relations in the style of the work by Akama (1987). An alternative characterization using 2-valued assignments plus an involution has been described by Routley (1974).

Syntactically, we assume a logical language with a *strong negation* connective "${\sim}$". That is, given some (possibly infinite) set of atoms At, a *formula* φ is defined using the grammar:

$$\varphi \quad ::= \quad \bot \ \mid \ a \ \mid \ {\sim}\varphi \ \mid \ \varphi \wedge \varphi \ \mid \ \varphi \vee \varphi \ \mid \ \varphi \rightarrow \varphi,$$

with $a \in At$. We use Greek letters φ and ψ and their variants to stand for propositional formulas. *Intuitionistic negation* is defined as $\neg\varphi \stackrel{\text{def}}{=} (\varphi \rightarrow \bot)$. We also define the derived operators $\varphi \leftrightarrow \psi \stackrel{\text{def}}{=} (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$ and $\top \stackrel{\text{def}}{=} {\sim}\bot$.

A Kripke frame $\mathcal{F} = \langle W, \leq \rangle$ is a pair where W is a nonempty set of worlds and $\leq$ is a partial order on W. A valuation $V : W \longrightarrow 2^{At}$ is a function mapping each world to a subset of atoms. A Nelson's interpretation (N-interpretation) is a 3-tuple $\mathcal{I} = \langle \mathcal{F}, V^+, V^- \rangle$ where $\mathcal{F} = \langle W, \leq \rangle$ is a Kripke frame and where both V^+ and V^- are valuations satisfying, for every pair of worlds $w, w' \in W$ with $w \leq w'$ and every atom $a \in At$, the following preservation properties:

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

(i) $V^+(w) \subseteq V^+(w')$, and
(ii) $V^-(w) \subseteq V^-(w')$.

Intuitively, V^+ represents our knowledge about constructive truth while V^- represents our knowledge about constructive falsity. We say that $\mathcal{I}$ is *consistent* if, in addition, it satisfies:

(iii) $V^+(w) \cap V^-(w) = \varnothing$ for every world $w \in W$.

Two forcing relations $\models^+$ and $\models^-$ are defined with respect to any N-interpretation $\mathcal{I} = \langle \mathcal{F}, V^+, V^- \rangle$, world $w \in W$ and atom $a \in At$ as follows:

$$\begin{aligned} \mathcal{I}, w \models^+ a \quad &\text{iff} \quad a \in V^+(w), \\ \mathcal{I}, w \models^- a \quad &\text{iff} \quad a \in V^-(w). \end{aligned}$$

These two relations are extended to compounded formulas as follows:

$$\begin{array}{ll} \mathcal{I}, w \not\models^+ \bot & \\ \mathcal{I}, w \models^+ \varphi_1 \wedge \varphi_2 & \text{iff } \mathcal{I}, w \models^+ \varphi_1 \text{ and } \mathcal{I}, w \models^+ \varphi_2 \\ \mathcal{I}, w \models^+ \varphi_1 \vee \varphi_2 & \text{iff } \mathcal{I}, w \models^+ \varphi_1 \text{ or } \mathcal{I}, w \models^+ \varphi_2 \\ \mathcal{I}, w \models^+ \varphi_1 \rightarrow \varphi_2 & \text{iff } \forall w' \geq w \; \mathcal{I}, w' \not\models^+ \varphi_1 \text{ or } \mathcal{I}, w' \models^+ \varphi_2 \\ \mathcal{I}, w \models^+ {\sim}\varphi & \text{iff } \mathcal{I}, w \models^- \varphi \\ \mathcal{I}, w \models^- \bot & \\ \mathcal{I}, w \models^- \varphi_1 \wedge \varphi_2 & \text{iff } \mathcal{I}, w \models^- \varphi_1 \text{ or } \mathcal{I}, w \models^- \varphi_2 \\ \mathcal{I}, w \models^- \varphi_1 \vee \varphi_2 & \text{iff } \mathcal{I}, w \models^- \varphi_1 \text{ and } I, w \models^- \varphi_2 \\ \mathcal{I}, w \models^- \varphi_1 \rightarrow \varphi_2 & \text{iff } \mathcal{I}, w \models^+ \varphi_1 \text{ and } \mathcal{I}, w \models^- \varphi_2 \\ \mathcal{I}, w \models^- {\sim}\varphi & \text{iff } \mathcal{I}, w \models^+ \varphi. \end{array}$$

An N-interpretation is said to be an *N-model* of a formula φ, in symbols $\mathcal{I} \models^+ \varphi$, iff $\mathcal{I}, w \models^+ \varphi$ for every $w \in W$. It is said to be *N-model* of a theory Γ, in symbols also $\mathcal{I} \models^+ \Gamma$, iff it is an N-model of all its formulas $\mathcal{I} \models^+ \varphi$. A formula φ is said to be a *consequence* of a theory Γ iff every model of Γ is also a model of φ, that is $\mathcal{I} \models^+ \varphi$ for every $\mathcal{I} \models^+ \Gamma$. This formalization characterizes **N4** while a restriction to consistent N-interpretations would characterize **N3**. As mentioned above, **N4** is "somehow" paraconsistent in the sense that a formula φ and its strongly negated counterpart ${\sim}\varphi$ may simultaneously be consequences of some theory: for instance, we have that $\{a, {\sim}a\} \models^+ a$ and $\{a, {\sim}a\} \models^+ {\sim}a$. Intuitively, these two forcing relations determine the four values above mentioned: a formula φ satisfying $\mathcal{I} \not\models^+ \varphi$ and $\mathcal{I} \not\models^- \varphi$ is understood as *unknown*. If it satisfies $\mathcal{I} \models^+ \varphi$ and $\mathcal{I} \not\models^- \varphi$, is understood as *true*. *False* if $\mathcal{I} \not\models^+ \varphi$ and $\mathcal{I} \models^- \varphi$, and *inconsistent* if $\mathcal{I} \models^+ \varphi$ and $\mathcal{I} \models^- \varphi$.

2.2 Logic programming, equilibrium logic, and here-and-there Nelson's models

In order to accommodate logic programming conventions, we will indistinctly write $\varphi \leftarrow \psi$ instead of $\psi \rightarrow \varphi$ when describing logic programs. An *explicit literal* is either an atom $a \in At$ or an atom preceded by strong negation ${\sim}a$. A *literal* is either an explicit literal l or an explicit literal preceded by intuitionistic negation $\neg l$. A literal that

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

contains intuitionistic negation is called *negative*. Otherwise, it is called *positive*. A *rule* is a formula of the form $H \leftarrow B$ where H is a disjunction of atoms and B is a conjunction of literals. A logic program Π is a set of rules.

Given some set of explicit literals $\mathbf{T}$ and some formula φ, we write $\mathbf{T} \models^+ \varphi$ when $\langle \mathcal{F}, V^+, V^- \rangle \models^+ \varphi$ holds for the Kripke frame $\mathcal{F}$ with a unique world w and valuations: $V^+(w) = \mathbf{T} \cap At$ and $V^-(w) = \{ \, a \mid {\sim} a \in \mathbf{T} \, \}$. A set of explicit literals $\mathbf{T}$ is said to be *closed* under Π if $\mathbf{T} \models^+ H \leftarrow B$ for every rule $H \leftarrow B$ in Π.

Next, we recall the notions of reduct and answer set (Gelfond and Lifschitz 1991):

Definition 1 (*Reduct and Answer Set*)
The *reduct* of program Π w.r.t. some set of explicit literals $\mathbf{T}$ is defined as follows

(i) Remove all rules with $\neg l$ in the body s.t. $l \in \mathbf{T}$,
(ii) Remove all negative literals for the remaining rules.

Set $\mathbf{T}$ is a *stable model* of Π if $\mathbf{T}$ is a $\subseteq$-minimal closed set under Π.

For characterizing logic programs in constructive logic, we are only interested in a particular kind of N-interpretations over *Here-and-There* (HT) frames. These frames are of the form $\mathcal{F}_{HT} = \langle \{h, t\}, \leq \rangle$ where $\leq$ is a partial order satisfying $h \leq t$. We refer to N-interpretations with an HT-frame as *HT-interpretations*. A *HT-model* is an N-model which is also a HT-interpretation. We use the generic terms *interpretation* (resp. *model*) for both HT and N-interpretations (resp. models) when it is clear by the context. At first sight, it may look that restricting ourselves to HT frames is an oversimplification. However, once the closed world assumption is added to intuitionistic logic, this logic can be replaced without loss of generality by any proper intermediate logic (Osorio *et al.* 2005; Cabalar *et al.* 2017).

Given any HT-interpretation, $\mathcal{I} = \langle \mathcal{F}_{HT}, V^+, V^- \rangle$, we define four sets of atoms as follows:

$$H^+_{\mathcal{I}} \stackrel{\text{def}}{=} V^+(h) \qquad T^+_{\mathcal{I}} \stackrel{\text{def}}{=} V^+(t)$$
$$H^-_{\mathcal{I}} \stackrel{\text{def}}{=} V^-(h) \qquad T^-_{\mathcal{I}} \stackrel{\text{def}}{=} V^-(t).$$

These sets of atoms correspond to the atoms verified at each corresponding world and valuation. Every HT-interpretation $\mathcal{I}$ is fully determined by these four sets. We will omit the subscript and write, for instance, H^+ instead of $H^+_{\mathcal{I}}$ when $\mathcal{I}$ is clear from the context. Furthermore, any HT-interpretations can be succinctly rewritten as a pair $\mathcal{I} = \langle \mathbf{H}, \mathbf{T} \rangle$ where $\mathbf{H} = H^+ \cup {\sim} H^-$ and $\mathbf{T} = T^+ \cup {\sim} T^-$ are sets of literals.[1] Note that, by the preservation properties of N-interpretations, we have that $\mathbf{H} \subseteq \mathbf{T}$. We say that an HT-interpretation $\mathcal{I} = \langle \mathbf{H}, \mathbf{T} \rangle$ is *total* iff $\mathbf{H} = \mathbf{T}$. Given HT-interpretations $\mathcal{I} = \langle \mathbf{H}, \mathbf{T} \rangle$ and $\mathcal{I}' = \langle \mathbf{H}', \mathbf{T}' \rangle$, we write $\mathcal{I} \leq \mathcal{I}'$ iff $\mathbf{H} \subseteq \mathbf{H}'$ and $\mathbf{T} = \mathbf{T}'$. As usual, we write $\mathcal{I} < \mathcal{I}'$ iff $\mathcal{I} \leq \mathcal{I}'$ and $\mathcal{I} \neq \mathcal{I}'$.

Next, we introduce the definition of equilibrium model (Pearce 1996).

Definition 2 (*Equilibrium model*)
A HT-model $\mathcal{I}$ of a theory Γ is said to be an *equilibrium model* iff it is total and there is no other HT-model $\mathcal{I}'$ of Γ s.t. $\mathcal{I}' < \mathcal{I}$.

[1] We denote by ${\sim} S \stackrel{\text{def}}{=} \{ \, {\sim}\varphi \mid \varphi \in S \, \}$ the of set strongly negated formulas of a given set S. Similarly, we also define $\neg S \stackrel{\text{def}}{=} \{ \, \neg\varphi \mid \varphi \in S \, \}$.

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

Interestingly, consistent equilibrium models precisely capture the answer set of a logic program. The following is a rephrase of Proposition 2 by Pearce (1996) using our notation.

Proposition 1
Let Π be a logic program. A consistent set $\mathbf{T}$ of explicit literals is a stable model of Π if and only if $\mathbf{T}$ is the set of explicit literals true in some consistent equilibrium model of Π.

More in general, it has been shown by Odintsov and Pearce (2005) that the (possible nonconsistent) equilibrium models of a logic program capture its paraconsistent answer sets (Sakama and Inoue 1995).

The following propositions characterizes some interesting properties of HT and strong negation that will be useful through the paper[2] :

Proposition 2 (*Persistence*)
Any HT-interpretation $\mathcal{I}$, formula φ and world $w \in \{h, t\}$ satisfy:

1. $I, w \models^+ \varphi$ implies $I, t \models^+ \varphi$, and
2. $I, w \models^- \varphi$ implies $I, t \models^- \varphi$.

Proposition 3 (*HT-negation*)
Any HT-interpretation $\mathcal{I}$, formula φ and world $w \in \{h, t\}$ satisfy:

i) $\mathcal{I}, w \models^+ \neg\varphi$ iff $\mathcal{I}, t \not\models^+ \varphi$, and
ii) $\mathcal{I}, w \models^+ \neg\neg\varphi$ iff $\mathcal{I}, t \models^+ \varphi$, and
iii) $\mathcal{I}, w \models^+ \neg\neg\neg\varphi$ iff $\mathcal{I}, w \models^+ \neg\varphi$, and
iv) $\mathcal{I}, w \models^- \neg\varphi$ iff $\mathcal{I}, w \models^- {\sim}\varphi$.

2.3 Abstract argumentation frameworks

Since their introduction, the syntax of AFs have been extended in different ways. One of these extensions, usually called SETAFs, consists in generalizing the notion of binary attacks to collective attacks such that a set of arguments B attacks some argument a (Nielsen and Parsons 2007). Another such extension, usually called Bipolar AFs (BAFs), consists in frameworks with a second positive relation called *support* (Karacapilidis and Papadias 2001; Verheij 2003a; Amgoud *et al.* 2004). In particular, Verheij (2003b) introduced the idea that, in AFs, arguments are considered as *prima-facie* justified statements, which can be considered true until proved otherwise, that is, until they are defeated. This allows introducing a second class of *ordinary arguments*, which cannot be considered true unless get supported by the prima-facie ones. Later, Polberg and Oren (2014) developed this idea by introducing Evidence-Based AFs (EBAFs), an extension of SETAFs (and, this, of AFs) which incorporates the notions of support and prima-facie arguments. Next we introduce an equivalent definition by Cayrol *et al.* (2018), which is closer to the logic formulation we pursue here.

Definition 3 (*Evidence-Based Argumentation framework*)
An Evidence-Based Argumentation framework $\mathbf{EF} = \langle \mathbf{A}, \mathbf{R}_a, \mathbf{R}_s, \mathbf{P} \rangle$ is a 4-tuple where $\mathbf{A}$ represents a (possibly infinite) set of arguments, $\mathbf{R}_a \subseteq 2^{\mathbf{A}} \times \mathbf{A}$ is an attack relation,

[2] For the sake of clarity, proofs of formal results are moving to an appendix.

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

$\mathbf{R}_s \subseteq 2^{\mathbf{A}} \times \mathbf{A}$ is a support relation and $\mathbf{P} \subseteq \mathbf{A}$ is a set of distinguished *prima-facie* arguments. We say that an **EF** is *finitary* iff B is finite for every attack or support $(B, a) \in \mathbf{R}_a \cup \mathbf{R}_s$.

The notion of acceptability is extended by requiring not only defense against all attacking arguments, but also support from some prima-facie arguments. Furthermore, the defense can be provided not only by defeating all attacking sets of arguments, but also by denying the necessary support for some of the non-prima-facie arguments of these attacks.

Definition 4 (*Defeat/Acceptability*)
Given some argument $a \in \mathbf{A}$ and set of arguments $E \subseteq \mathbf{A}$, we say

1. a is *defeated* w.r.t. E iff there is some $B \subseteq E$ s.t. $(B, a) \in \mathbf{R}_a$,

Def(E) will denote the set of arguments that are defeated w.r.t. E.

2. a is *supported* w.r.t. E iff either $a \in \mathbf{P}$ or there is some $B \subseteq E \setminus \{a\}$ whose elements are supported w.r.t. $E \setminus \{a\}$ and such that $(B, a) \in \mathbf{R}_s$,
3. a is *supportable* w.r.t. E iff it is supported w.r.t. $\mathbf{A} \setminus \mathit{Def}(E)$,
4. a is *unacceptable* w.r.t. E iff it is either defeated or not supportable,
5. a is *acceptable* w.r.t. E iff it is supported and, for every $(B, a) \in \mathbf{R}_a$, there is $b \in B$ such that b is unacceptable w.r.t. E

Sup(E) (resp. *UnAcc*(E) and *Acc*(E)) will denote the set of arguments that are supported (resp. unacceptable and acceptable) w.r.t. E.

Then, semantics are defined as follows:

Definition 5
A set of arguments $E \subseteq \mathbf{A}$ is said to be:

1. *self-supporting* iff $E \subseteq \mathit{Sup}(E)$,
2. *conflict-free* iff $E \cap \mathit{Def}(E) = \varnothing$,
3. *admissible* iff it is conflict-free and $E \subseteq \mathit{Acc}(E)$,
4. *complete* iff it is conflict-free and $E = \mathit{Acc}(E)$,
5. *preferred* iff it is a $\subseteq$-maximal admissible set,
6. *stable* iff $E = \mathbf{A} \setminus \mathit{UnAcc}(E)$.

SETAFs can be seen as special cases where the set of supports is empty and all arguments are prima-facie. In this sense, we write $\mathbf{SF} = \langle \mathbf{A}, \mathbf{R}_a \rangle$ instead $\mathbf{EF} = \langle \mathbf{A}, \mathbf{R}_a, \varnothing, \mathbf{A} \rangle$. Furthermore, in their turn, AFs can be seen as a special case of SETAFs where all attacks have singleton sources. In such case, we just write $\mathbf{AF} = \langle \mathbf{A}, \mathbf{R} \rangle$ instead $\mathbf{SF} = \langle \mathbf{A}, \mathbf{R}_a \rangle$, where $\mathbf{R} = \{\ (b, a) \mid (\{b\}, a) \in \mathbf{R}_a\ \}$ For this kind of frameworks, the respective notions of conflict-free (resp. admissible, complete, preferred or stable) coincide with those being defined by Nielsen and Parsons (2007) and Dung (1995), respectively.

To illustrate the notions of support and prima-facie arguments, consider the well-known Tweety example:

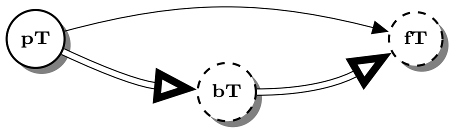
Example 1
Suppose we have the knowledge base that includes the following statements:

1. birds (normally) can fly,
2. penguins are birds,

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

3. penguins cannot fly and
4. Tweety is a penguin.

We can formalize this by the following graph:



where pT, bT and fT respectively stand for "Tweety is a penguin", "Tweety is a bird" and "Tweety can fly." Double arrows represent support while simple ones represent attacks. Furthermore, circles with solid border represent prima-facie arguments while dashed border ones represent ordinary ones. That is, "Tweety is a penguin" is considered a prima-facie argument that supports that "Tweety is a bird" which, in its turn, supports that "Tweety can fly." The latter is then considered also prima-facie, that is, true unless proven otherwise. Note that "Tweety is a penguin" also attacks that "Tweety can fly", so the latter cannot be accepted as true. Formally, this corresponds to the framework $\mathbf{EF}_1 = \langle \mathbf{A}, \mathbf{R}_a, \mathbf{R}_s, \mathbf{P} \rangle$ with $\mathbf{R}_a = \{(\{pT\}, fT)\}$ and $\mathbf{R}_s = \{(\{pT\}, bT), (\{bT\}, fT)\}$ and $\mathbf{P} = \{pT\}$ whose unique admissible, complete, preferred and stable extension is $\{pT, bT\}$. In other words, we conclude that "Tweety cannot fly." Note that "Tweety is a penguin" provides conflicting evidence for whether it can fly or not. In EBAFs, this is solved by giving priority to the attack relation, so "Tweety cannot fly" is inferred.

3 Reasoning with contradictory evidence in equilibrium logic

In this section, we formalize principles **NC** and **CW** in constructive logic, obtaining as a result a formalism which is a conservative extension of logic programming under the answer set semantics (see Theorem 1 and Corollary 1 below) and which is capable of reasoning with contradictory evidence. We start by defining a new implication connective that captures **NC** in terms of intuitionistic implication and strong negation:

$$\varphi_1 \Rightarrow \varphi_2 \quad \stackrel{\text{def}}{=} \quad (\neg{\sim}\varphi_1 \wedge \varphi_1) \rightarrow \varphi_2. \tag{1}$$

Recall that intuitionistic implication $\varphi_1 \rightarrow \varphi_2$ can be informally understood as a means to construct a proof of the truth of the consequent φ_2 in terms of a proof of truth of the antecedent φ_1. In this sense, (1) can be understood as a means to construct a proof of the truth of the consequent φ_2 in terms of proof of the truth of the antecedent φ_1 and the absence of a proof of its falsity, or in other words, in terms of a *consistent proof* of the antecedent φ_1. It is easy to see that (1) is weaker than intuitionistic implication, that is, that

$$\varphi_1 \rightarrow \varphi_2 \;\models^+\; \varphi_1 \Rightarrow \varphi_2,$$

holds for every pair of formulas φ_1 and φ_2. We can use the following simple example to illustrate the difference between intuitionistic implication and (1).

Example 2
Let Γ_2 be the following set of formulas:

$$a \qquad\qquad b \qquad\qquad {\sim}b \qquad\qquad a \Rightarrow c \qquad\qquad b \Rightarrow d,$$

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

and let Γ_2' be the theory obtained by replacing each occurrence of implication $\Rightarrow$ by intuitionistic implication $\rightarrow$. On the one hand, we have that both, Γ_2 and Γ_2', entail atoms a and c. On the other hand, we have: $\Gamma_2' \models^+ d$ but $\Gamma_2 \not\models^+ d$. This is in accordance with **NC**, since the only way to obtain a proof of d is in terms of b, for which we have contradictory evidence. Note also that an alternative proof of d could be obtained if new consistent evidence becomes available: for the theory $\Gamma_3 = \Gamma_2 \cup \{a \Rightarrow d\}$ we obtain $\Gamma_3 \models^+ d$. It is also worth highlighting that, in contrast with intuitionistic implication, this new connective (1) is not monotonic: for $\Gamma_4 = \{b,\ b \Rightarrow d\}$ we have $\Gamma_4 \models^+ d$ and $\Gamma_4 \cup \{\sim b\} \not\models^+ d$. Obviously, it is not antimonotonic either: $\Gamma_4 \setminus \{b\} \not\models^+ d$.

The following result shows that, when dealing with consistent evidence, these differences disappear and (1) collapses into intuitionistic implication:

Proposition 4
Let $\mathcal{I}$ be a consistent N-interpretation and let φ_1 and φ_2 be any pair of formulas. Then, $\mathcal{I} \models^+ \varphi_1 \Rightarrow \varphi_2$ iff $\mathcal{I} \models^+ \varphi_1 \rightarrow \varphi_2$.

Let us now formalize the **CW** assumption. As usual nonmonotonicity is obtained by considering equilibrium models (Definition 2). However, to capture **CW**, we need to restrict the consequences of these models to those that are consistent. We do so by introducing a new *cw-inference* relation which, precisely, restricts the consequences of $\models^+$ to those which are consistent:

$$\mathcal{I}, w \models \varphi \quad \text{iff} \quad \mathcal{I}, w \models^+ \neg{\sim}\varphi \wedge \varphi. \tag{2}$$

Furthermore, as usual, we write $\mathcal{I} \models \varphi$ iff $\mathcal{I}, w \models \varphi$ for all $w \in W$. We also write $\Gamma \models \varphi$ iff $\mathcal{I} \models \varphi$ holds for every equilibrium model $\mathcal{I}$ of Γ. For instance, in Example 2, it is easy to see that $\Gamma_2 \models^+ b$ and $\Gamma_2 \models^+ \sim b$, but $\Gamma_2 \not\models b$ and $\Gamma_2 \not\models \sim b$ because the unique equilibrium model of Γ_2 contains contradictory evidence for b. On the other hand, as may be expected, when we deal with noncontradictory evidence cw-inference $\models$ just collapses to the regular inference relation $\models^+$ (see Proposition 5 below).

To finalize the formalization of **CW**, we also need to define *default negation*. This is accomplished by introducing a new connective *not* and adding the following two items to the Nelson's forcing relations:

$$\mathcal{I}, w \models^+ \mathit{not}\ \varphi \text{ iff } \mathcal{I}, w \models^+ \neg\varphi \vee (\varphi \wedge {\sim}\varphi),$$
$$\mathcal{I}, w \models^- \mathit{not}\ \varphi \text{ iff } \mathcal{I}, w \models^+ \varphi \text{ and } \mathcal{I}, w \not\models^- \varphi.$$

Then, an *extended formula* φ is defined using the following grammar:

$$\varphi \quad ::= \quad \bot \mid a \mid {\sim}\varphi \mid \mathit{not}\ \varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \rightarrow \varphi,$$

with $a \in At$ an atom. The following result shows that cw-inference and default negation are conservative extensions of the satisfaction relation $\models^+$ and HT-negation $\neg$ when restricted to consistent knowledge.

Proposition 5
Let $\mathcal{I}$ be a consistent N-interpretation and φ be any extended formula. Then, the following conditions hold:

(i) $\mathcal{I} \models \varphi$ iff $\mathcal{I} \models^+ \varphi$
(ii) $\mathcal{I} \models \mathit{not}\ \varphi$ iff $\mathcal{I} \models \neg\varphi$.

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

Despite the relation between default negation *not* and HT-negation $\neg$ on consistent interpretations, in general, they no not coincide. The following example illustrates the difference between these two kinds of negations:

Example 3
Let Γ_5 be the following theory:

$$a \qquad\qquad \sim a \qquad\qquad \mathit{not} \sim a \Rightarrow b.$$

This theory has a unique equilibrium model $\mathcal{I} = \langle \mathbf{T}, \mathbf{T} \rangle$ with $\mathbf{T} = \{a, \sim a, b\}$. Note that, every model $\mathcal{J}$ of Γ_5 must satisfy $\mathcal{J} \models^+ a \wedge \sim a$ and, thus, it must also satisfy $\mathcal{J} \models \mathit{not} \sim a$ and $\mathcal{J} \models^+ b$ follows (Proposition 6). Hence, $\mathcal{I}$ is a $\leq$-minimal model and, thus, an equilibrium model. On the other hand, let Γ_6 be the theory:

$$a \qquad\qquad \sim a \qquad\qquad \neg\sim a \Rightarrow b.$$

In this case, we can check that $\mathcal{J} = \langle \mathbf{H}, \mathbf{T} \rangle$ with $\mathbf{H} = \{a, \sim a\}$ is a model of Γ_6 because $\mathcal{J} \not\models \neg\sim a$ and, thus, now $\mathcal{I}$ is not an equilibrium model. In fact, $\langle \mathbf{H}, \mathbf{H} \rangle$ is the unique equilibrium model of Γ_6.

The following result shows the relation between default negation, implication, and cw-inference.

Proposition 6
Let $\mathcal{I}$ be any N-interpretation and φ be any formula. Then,

(i) $\mathcal{I} \models \varphi$ and $\mathcal{I} \models^+ \varphi \Rightarrow \psi$ implies $\mathcal{I} \models^+ \psi$,
(ii) $\mathcal{I} \models \mathit{not}\ \varphi$ implies $\mathcal{I} \not\models \varphi$.

Furthermore, if $\mathcal{I}$ is a total HT-interpretation, then

(iii) $\mathcal{I} \models \mathit{not}\ \varphi$ iff $\mathcal{I} \not\models \varphi$.

Condition (i) formalizes a kind of *modus ponens* for $\Rightarrow$ in the sense that, if the we have a consistent proof of the antecedent, then we have a (possibly inconsistent) proof of the consequent. It is clear that this statement cannot be strengthened to provide a consistent proof of the consequent because any other formula could provide the contradictory evidence to make it inconsistent. Note also that this relation is nonmonotonic as adding new information may result in a contradictory antecedent. Condition (iii) formalizes the **CW** assumption, that is, *not* φ holds whenever φ is not known to be true or we have contradictory evidence for it. Note that, according to this, the default negation of an inconsistent formula is true and, therefore, the evaluation of default negation itself is always consistent (even if the formula is inconsistent): that is, $\mathcal{I}, w \not\models^+ \mathit{not}\ \varphi$ or $\mathcal{I}, w \not\models^- \mathit{not}\ \varphi$ holds for any extended formula.

On the contrary that implication $\Rightarrow$, default negation *not* cannot be straightforwardly defined[3] in terms of Nelson's connectives.

Another alternative, we have investigated was defining *not* φ as and $\neg\varphi \vee (\varphi \wedge \sim\varphi)$. in terms of cw-inference. The following result shades light on this attempt.

Proposition 7
Let $\mathcal{I}$ be any N-interpretation and φ be any formula. Then, $\mathcal{I} \models \neg\varphi \vee (\varphi \wedge \sim\varphi)$ iff $\mathcal{I} \models \neg\varphi$.

[3] It is still an open question whether it is definable in terms of Nelson's connectives or not.

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

That is, in terms of cw-inference, $\neg\varphi \vee (\varphi \wedge {\sim}\varphi)$ is equivalent to HT-negation. As illustrated by Example 3, default negation and HT-negation do not behave in the same way.

The following example illustrates that, though default negation allows to derive new knowledge from contradictory information, it does not allow to self justify a contradiction.

Example 4
Let Γ_7 be a logic program containing the following single rule:

$$\mathit{not}\ {\sim}a \Rightarrow a, \tag{3}$$

stating, as usual, that a holds by default. As expected this theory has a unique equilibrium model $\mathcal{I}$ which satisfies $\mathcal{I} \models a$ and $\mathcal{I} \not\models {\sim}a$. Let now $\Gamma_8 = \Gamma_7 \cup \{{\sim}a\}$. This second theory also has a unique equilibrium model $\mathcal{I}$ which now satisfies $\mathcal{I} \models {\sim}a$ and $\mathcal{I} \not\models a$. To see that $\mathcal{J} = \langle \mathbf{T}, \mathbf{T} \rangle$ with $\mathbf{T} = \{a, {\sim}a\}$ is not an equilibrium model of Γ_8, let $\mathcal{J}' = \langle \mathbf{H}, \mathbf{T} \rangle$ with $\mathbf{H} = \{{\sim}a\}$ be an interpretation. Since $\mathcal{J}'$ satisfies $\mathcal{J}' < \mathcal{J}$ and it is a model of ${\sim}a$, it only remains to be shown that $\mathcal{J}'$ is a model of (3). For that, just note $\mathcal{J} \models {\sim}a$ and, thus, $\mathcal{J} \not\models \mathit{not}\ {\sim}a$ follows by Proposition 6. This implies that $\mathcal{J}'$ satisfies (3) and, consequently, that $\mathcal{J}$ is not an equilibrium model. In fact, $\langle \mathbf{H}, \mathbf{H} \rangle$ is the unique equilibrium model of Γ_8.

3.1 A conservative extension of logic programming

Let us now consider the language formed with the set of logical connectives

$$\mathcal{C}_{LP} \stackrel{\text{def}}{=} \{\bot, {\sim}, \wedge, \vee, \Rightarrow, \mathit{not}\,\}.$$

In other words, a $\mathcal{C}_{LP}$*-formula* φ is defined using the following grammar:

$$\varphi \quad ::= \quad \bot \mid a \mid {\sim}\varphi \mid \mathit{not}\ \varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \Rightarrow \varphi,$$

with $a \in \mathit{At}$ being an atom. A $\mathcal{C}_{LP}$-literal is either an explicit literal l or is default negation $\mathit{not}\ l$. A $\mathcal{C}_{LP}$-rule is a formula of the form $H \Leftarrow B$ where H is a disjunction of atoms and B is a conjunction of $\mathcal{C}_{LP}$-literals. $\mathcal{C}_{LP}$*-theories* and $\mathcal{C}_{LP}$*-programs* are respectively defined as sets of $\mathcal{C}_{LP}$-formulas and $\mathcal{C}_{LP}$-rules. The definition of an answer set is applied straightforwardly as in Definition 1. Given any theory $\mathcal{C}_{LP}$-theory Γ, by $\mathcal{C}_N(\Gamma)$ we denote the result of

1. replacing every occurrence of $\Rightarrow$ by $\rightarrow$ and
2. and every occurrence of not by $\neg$.

Then, the following results follow directly from Propositions 4 and 5:

Theorem 1
Let Γ be any $\mathcal{C}_{LP}$*-theory* and $\mathcal{I}$ be any consistent interpretation. Then, $\mathcal{I}$ is an equilibrium model of Γ iff $\mathcal{I}$ is an equilibrium model of $\mathcal{C}_N(\Gamma)$.

Corollary 1
Let P be a $\mathcal{C}_{LP}$-program and $\mathbf{T}$ be any consistent set of explicit literals. Then, $\mathcal{I} = \langle \mathbf{T}, \mathbf{T} \rangle$ is an equilibrium model of P iff $\mathbf{T}$ is an answer set of P.

In other words, the equilibrium models semantics are a conservative extension of the answer set semantics. The following example shows the usual representation of the Tweety

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

scenario in this logic (an alternative representation using contradictory evidence will be discussed in the Discussion section).

Example 5 (*Ex. 1 continued*)
Consider again the Tweety scenario. The following logic program P_9 is a usual way of representing this scenario in LP:

$$\begin{aligned} \mathit{flyTweety} &\Leftarrow \mathit{birdTweety} \wedge \mathit{not} \sim\!\mathit{flyTweety}, & (4)\\ \mathit{birdTweety} &\Leftarrow \mathit{penguinTweety}, & (5)\\ \sim\!\mathit{flyTweety} &\Leftarrow \mathit{penguinTweety} & (6)\\ \mathit{penguinTweety}&, \end{aligned}$$

where rule (4) formalizes the statement "birds normally can fly." This is achieved by considering $\sim\!\mathit{flyTweety}$ as an exception to this rule. It can be checked that P_9 has a unique equilibrium model $\mathcal{I}_9$, which is consistent, and which satisfies $\mathcal{I}_9 \not\models \mathit{flyTweety}$ and $\mathcal{I}_9 \models \mathit{not}\ \mathit{flyTweety}$. In other words, Tweety cannot fly.

Example 6 (*Ex. 2 continued*)
Consider now the theory obtained by replacing formulas $a \Rightarrow c$ and $b \Rightarrow d$ in Γ_2 by the following two formulas:

$$\mathit{not}\ e \wedge a \Rightarrow c \qquad \mathit{not}\ e \wedge b \Rightarrow d.$$

Let Γ_{10} be such theory. It is easy to see that neither Γ_{10} nor $\mathcal{C}_N(\Gamma_{10})$ monotonically entail c nor d. This is due to the fact that the negation of e is not monotonically entailed: $\Gamma_{10} \not\models^+ \mathit{not}\ e$ and $\mathcal{C}_N(\Gamma_{10}) \not\models^+ \neg e$. On the other hand, the negation of e is non-monotonically entailed in both cases: $\Gamma_{10} \models \mathit{not}\ e$ and $\mathcal{C}_N(\Gamma_{10}) \models \neg e$. Note that both Γ_{10} and $\mathcal{C}_N(\Gamma_{10})$ have a unique equilibrium model, $\mathcal{I}_{10} = \langle \mathbf{T}, \mathbf{T} \rangle$ and $\mathcal{I}'_{10} = \langle \mathbf{T}', \mathbf{T}' \rangle$ with $\mathbf{T} = \{a, b, \sim\!b, c\}$ and $\mathbf{T}' = \{a, b, \sim\!b, c, d\}$, respectively, and in both cases we have $\mathcal{I}_{10} \models \mathit{not}\ e$ and $\mathcal{I}'_{10} \models \neg e$. As a result, we get that both theories cautiously entail c. However, as happened in Example 2, only $\mathcal{C}_N(\Gamma_{10})$ cautiously entails d, because the unique evidence for d comes from b for which we have inconsistent evidence. This behavior is different from paraconsistent answer sets (Sakama and Inoue 1995; Odintsov and Pearce 2005). As pointed out by Sakama and Inoue (1995), the truth of d is less credible than the truth of c, since d is derived through the contradictory fact b. In order to distinguish such two facts Sakama and Inoue (1995) also define *suspicious answer sets* which do not consider d as true.[4]

This example also helps us to illustrate the strengthened closed world assumption principle **CW**. On the one hand, we have that $\Gamma_{10} \models \mathit{not}\ e$ holds because there is no evidence for e. On the other hand, we have that $\Gamma_{10} \models \mathit{not}\ b$ holds because we have contradictory evidence for b. Moreover, we have that $\Gamma_{10} \models \mathit{not}\ d$ holds because the only evidence we have for d is based on the contradictory evidence for b.

[4] Suspicious answer sets are based on a 6-value lattice which add the values *suspiciously true* and *suspiciously false* to the four values of **N4**. In the unique suspicious answer set of Γ_{10}, atom d gets assigned the suspiciously true value instead the true value. A formal comparison with suspicious answer sets is left for future work.

4 Argumentation frameworks in equilibrium logic

In this section, we show how AFs, SETAFs, and EBAFs can be translated in this logic in a modular way and using only the object language. This translation is a formalization of the intuition of an attack stated in **AT**. Theorems 2, 3, and 4 show that the equilibrium models of this translation precisely characterize the stable extension of the corresponding framework.

4.1 Dung's argumentation frameworks

Now, let us formalize the notion of attack introduced in **AT**, by defining the following connective:

$$\varphi_1 \rightsquigarrow \varphi_2 \quad \stackrel{\text{def}}{=} \quad \varphi_1 \Rightarrow {\sim}\varphi_2. \tag{7}$$

Here, we identify the acceptability of φ_1 with having a consistent proof of it, or in other words, as having a proof of the truth of φ_1 and not having a proof of its falsity. Then, (7) states that the acceptability of φ_1 allows to construct a proof of the falsity of φ_2. In this sense, we identify a proof of the falsity of φ_2 with φ_2 being defeated.

Proposition 8
Given any N-interpretation $\mathcal{I}$ and any pair of formulas φ_1, φ_2, the following conditions hold:

(i) $\mathcal{I} \models \varphi_1$ and $\mathcal{I} \models^+ \varphi_1 \rightsquigarrow \varphi_2$ imply $\mathcal{I} \models^- \varphi_2$

Using the language $\mathcal{C}_{AF} = \{\rightsquigarrow\}$, we can translate any AF as follows:

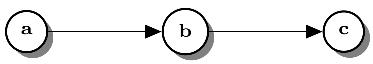
Definition 6
Given some framework $\mathbf{AF} = \langle \mathbf{A}, \mathbf{R} \rangle$, we define the theory:

$$\mathcal{C}_{AF}(\mathbf{AF}) \stackrel{\text{def}}{=} \mathbf{A} \cup \{\, a \rightsquigarrow b \;\big|\; (a,b) \in \mathbf{R} \,\}. \tag{8}$$

In addition, we assign a corresponding set of arguments $E_{\mathcal{I}} \stackrel{\text{def}}{=} \{\, a \in \mathbf{A} \,\big|\, \mathcal{I} \models a \,\}$ to every interpretation $\mathcal{I}$.

Translation $\mathcal{C}_{AF}(\cdot)$ applies the notion of attack introduced in **AT** to translate an AF into a logical theory. The strengthened close world assumption **CW** is used to retrieve the arguments $E_{\mathcal{I}}$ corresponding to each stable model $\mathcal{I}$ of the logical theory obtained from this translation.

Example 7
To illustrate this translation, let $\mathbf{AF}_{11}$ be the framework corresponding to the following graph:



Then, we have that $\mathcal{C}_{AF}(\mathbf{AF}_{11})$ is the theory containing the following two attacks:

$$a \rightsquigarrow b \qquad\qquad b \rightsquigarrow c,$$

plus the facts $\{a, b, c\}$.

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

Proposition 9
Let $\mathbf{AF}$ be some framework and $\mathcal{I}$ be some HT-model of $\mathcal{C}_{AF}(\mathbf{AF})$. Then, the following hold:

(i) if a is defeated w.r.t. $E_{\mathcal{I}}$, then $\mathcal{I} \models^{+} {\sim}a$
(ii) $E_{\mathcal{I}}$ is conflict-free.

If, in addition, $\mathcal{I}$ is an $\leq$-minimal model, then

(iii) a is defeated w.r.t. $E_{\mathcal{I}}$ iff $\mathcal{I} \models^{+} {\sim}a$.

Example 8 (*Ex. 7 continued*)
Continuing with our running example, let $\mathcal{I}_{11} = \langle \mathbf{T}_{11}, \mathbf{T}_{11} \rangle$ and $\mathcal{J}_{11} = \langle \mathbf{T}'_{11}, \mathbf{T}'_{11} \rangle$ be two total models of $\Gamma_{\mathbf{AF}_{11}}$ with $\mathbf{T}_{11} = \{a, b, c, {\sim}b\}$ and $\mathbf{T}'_{11} = \{a, b, c, {\sim}a, {\sim}c\}$. Then, we have that both $S_{\mathcal{I}_{11}} = \{a, c\}$ and $S_{\mathcal{J}_{11}} = \{b\}$ are conflict-free (though only $S_{\mathcal{I}_{11}}$ is stable). Furthermore, we also can see that argument b is the unique defeated argument w.r.t. $S_{\mathcal{I}_{11}}$ and the unique atom for which $\mathcal{I}_{11} \models^{+} {\sim}b$ holds. On the other hand, we get that argument c is the unique defeated argument w.r.t. $S_{\mathcal{J}_{11}}$ and also both $\mathcal{J}_{11} \models^{+} {\sim}a$ and $\mathcal{J}_{11} \models^{+} {\sim}c$ hold. Note that, as stated by (iii) in Proposition 9, this implies that only $S_{\mathcal{I}_{11}}$ can be an equilibrium model. Let us show that it is indeed the case that $\mathcal{J}_{11}$ is not an equilibrium model and let us define, for that purpose, an interpretation $\mathcal{J}'_{11} = \langle \mathbf{H}'_{11}, \mathbf{T}'_{11} \rangle$ with $\mathbf{H}'_{11} = \mathbf{T}'_{11} \setminus \{{\sim}a\} = \{a, b, c, {\sim}c\}$. In other words, interpretation $\mathcal{J}'_{11}$ is as $\mathcal{J}_{11}$, but removing the non-defeated argument a as a negated conclusion ${\sim}a$. It is easy to check that $\mathcal{J}'_{11} \models b \rightsquigarrow c$ because ${\sim}c \in \mathbf{H}'_{11}$ holds. Besides, since ${\sim}a \in \mathbf{T}'_{11}$, we have that $\mathcal{J}'_{11} \not\models a$ and, therefore, that $\mathcal{J}'_{11} \models a \rightsquigarrow b$. This implies that $\mathcal{J}'_{11}$ is a model of $\Gamma_{\mathbf{AF}_{11}}$. Since $\mathcal{J}'_{11} < \mathcal{J}_{11}$, we get that $\mathcal{J}_{11}$ is not an equilibrium model.

In fact, we can generalize this correspondence between the stable extensions and the equilibrium models to any argumentation framework as stated by the following theorem:

Theorem 2
Given some $\mathbf{AF} = \langle \mathbf{A}, \mathbf{R} \rangle$, there is a one-to-one correspondence between its stable extensions and the equilibrium models of $\mathcal{C}_{AF}(\mathbf{AF})$ such that

(i) if $\mathcal{I}$ is an equilibrium model of $\mathcal{C}_{AF}(\mathbf{AF})$, then $E_{\mathcal{I}}$ is a stable extension of $\mathbf{AF}$,
(ii) if E is a stable extension of $\mathbf{AF}$ and $\mathcal{I}$ is a total interpretation such that $T^{+}_{\mathcal{I}} = \mathbf{A}$ and $T^{-}_{\mathcal{I}} = \mathit{Def}(E)$, then $\mathcal{I}$ is an equilibrium model of $\mathcal{C}_{AF}(\mathbf{AF})$.

Proof sketch.[5] First, note that condition (i) follows directly from (iii) in Proposition 9 and the facts that (a) equilibrium models are $\leq$-minimal models and (b) $E_{\mathcal{I}}$ is a stable extension iff $E_{\mathcal{I}}$ are exactly the non-defeated arguments w.r.t. $E_{\mathcal{I}}$. To show (ii), it is easy to see that $E_{\mathcal{I}}$ being a stable extension implies that $\mathcal{I}$ is a model of $\mathcal{C}_{AF}(\mathbf{AF})$. Hence, to show that $\mathcal{I}$ is an equilibrium model what remains is to prove that any $\mathcal{J} < \mathcal{I}$ is not a model of $\mathcal{C}_{AF}(\mathbf{AF})$. Any such $\mathcal{J}$ must satisfy $H^{+}_{\mathcal{J}} = H^{+}_{\mathcal{I}} = \mathbf{A}$ and $H^{-}_{\mathcal{J}} \subset H^{-}_{\mathcal{I}} = T^{-}_{\mathcal{I}} = \mathit{Def}(E)$. Therefore, there is some defeated argument such that $a \notin H^{-}_{\mathcal{J}}$ and some defeating attack $(b, a) \in \mathbf{R}_a$ such that $b \in E = H^{+}_{\mathcal{I}} \setminus T^{-}_{\mathcal{I}} = H^{+}_{\mathcal{J}} \setminus T^{-}_{\mathcal{J}}$. This implies that $b \rightsquigarrow a \in \mathcal{C}_{AF}(\mathbf{AF})$

[5] This theorem is a particualr case of Theorem 3 below. Recall that full proofs are provided in the appendix.

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

and $\mathcal{J} \models b$ which, in its turn, implies that $a \in H^-_{\mathcal{J}}$. This is a contradiction and, consequently, $\mathcal{I}$ is an equilibrium model. □

Theorem 2 captures the relation between the stable extensions of an AF and its translation into a logical theory. As mentioned above, this relation relies on the reasoning principles **AT** and **CW**: An $\mathbf{AF} = \langle \mathbf{A}, \mathbf{R} \rangle$ is translated into a logical theory $\mathcal{C}_{AF}(\mathbf{AF})$ using the notion of attack introduced in **AT**. The stable extension $E_{\mathcal{I}}$ of this AF is then retrieved from the equilibrium model $\mathcal{I}$ of $\mathcal{C}_{AF}(\mathbf{AF})$ using the **CW** principle.

4.2 Set attack argumentation frameworks

We may also extend the results of the previous section to SETAFs using the language $\mathcal{C}_{SF} = \{\rightsquigarrow, \wedge\}$ and a similar translation.

Definition 7
Given some finitary set attack framework $\mathbf{SF} = \langle \mathbf{A}, \mathbf{R}_a \rangle$, we define

$$\Gamma_{\mathbf{R}_a} \stackrel{\text{def}}{=} \left\{ \bigwedge A \rightsquigarrow b \;\middle|\; (A, b) \in \mathbf{R}_a \right\}, \tag{9}$$

and $\mathcal{C}_{SF}(\mathbf{SF}) \stackrel{\text{def}}{=} \mathbf{A} \cup \Gamma_{\mathbf{R}_a}$.

Similar to Definition 6, translation $\mathcal{C}_{SF}(\cdot)$ applies the notion of attack introduced in **AT** to translate an AF into a logical theory. In this case, the set of attacking arguments becomes a conjuntion in the antecedent of the attack connective.

Theorem 3
Given some finitary **SF** there is a one-to-one correspondence between its stable extensions and the equilibrium models of $\mathcal{C}_{SF}(\mathbf{SF})$ such that

(i) if $\mathcal{I}$ is an equilibrium model of $\mathcal{C}_{SF}(\mathbf{SF})$, then $E_{\mathcal{I}}$ is a stable extension of **SF**,
(ii) if E is a stable extension of **SF** and $\mathcal{I}$ is a total interpretation such that $T^+_{\mathcal{I}} = \mathbf{A}$ and $T^-_{\mathcal{I}} = \mathit{Def}(E)$, then $\mathcal{I}$ is an equilibrium model of $\mathcal{C}_{SF}(\mathbf{SF})$.

Proof sketch. The proof follows as in Theorem 2 by noting that any interpretation $\mathcal{I}$ and set of arguments B satisfy: $B \subseteq E_{\mathcal{I}}$ iff $\mathcal{I} \models b$ for all $b \in B$ iff $\mathcal{I} \models \bigwedge B$. □

4.3 Argumentation frameworks with evidence-based support

Let us now extend the language of SETAFs with the LP implication (1), in other words, we consider the language possessing the following set of connectives $\mathcal{C}_{EF} = \{\rightsquigarrow, \wedge, \Rightarrow\}$, so that we can translate any EBAF as follows:

Definition 8
Given any finitary evidence-based framework $\mathbf{EF} = \langle \mathbf{A}, \mathbf{R}_a, \mathbf{R}_s, \mathbf{P} \rangle$, we define its corresponding theory as: $\mathcal{C}_{EF}(\mathbf{EF}) \stackrel{\text{def}}{=} \mathbf{P} \cup \Gamma_{\mathbf{R}_a} \cup \Gamma_{\mathbf{R}_s}$ with

$$\Gamma_{\mathbf{R}_s} \stackrel{\text{def}}{=} \left\{ \bigwedge A \Rightarrow b \;\middle|\; (A, b) \in \mathbf{R}_s \right\}, \tag{10}$$

and $\Gamma_{\mathbf{R}_a}$ as stated in (9).

Note that, in contrast with AFs and SETAFs, the theory corresponding to an EBAFs do not contain all arguments as atoms, but only those that are prima-facie **P**. This

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

reflects the fact that in EBAFs not all arguments can be accepted, but only those that are prima-facie or are supported by those prima-facie. Supports are represented using the LP implication $\Rightarrow$ and supported arguments are captured by the positive evaluation of each interpretation $H^+_{\mathcal{I}}$. The following result extends Proposition 9 to EBAFs including the relation between supported arguments and models.

Proposition 10

Let **EF** be some framework and $\mathcal{I}$ be some HT-model of $\mathcal{C}_{EF}(\mathbf{EF})$. Then, the following hold:

(i) if a is supported w.r.t. $E_{\mathcal{I}}$, then $\mathcal{I} \models^+ a$,
(ii) if a is defeated w.r.t. $E_{\mathcal{I}}$, then $\mathcal{I} \models^+ {\sim}a$,
(iii) $E_{\mathcal{I}}$ is conflict-free.

If, in addition, $\mathcal{I}$ is an $\leq$-minimal HT-model, then

(iii) a is supported w.r.t. $E_{\mathcal{I}}$ iff $\mathcal{I} \models^+ a$,
(iv) a is defeated w.r.t. $E_{\mathcal{I}}$ iff $\mathcal{I} \models^+ {\sim}a$,
(v) $E_{\mathcal{I}}$ is self-supporting.

Example 9 (Ex. 1 continued)

Consider now framework **EF** representing the Tweety scenario.

$$birdTweety \Rightarrow flyTweety, \tag{11}$$

$$penguinTweety \Rightarrow birdTweety, \tag{12}$$

$$penguinTweety \rightsquigarrow flyTweety \tag{13}$$

$$penguinTweety.$$

As mentioned in Example 1, framework $\mathbf{EF}_1$ has a unique stable extension

$$\{penguinTweety, birdTweety\},$$

which does not include the argument *flyTweety*. In other words, Tweety cannot fly. Interestingly, $\mathcal{C}_{SF}(\mathbf{EF}_1)$ has also a unique equilibrium model $\mathcal{I}_{12} = \langle \mathbf{T}_{12}, \mathbf{T}_{12} \rangle$ where $\mathbf{T}_{12}$ stands for the set:

$$\{penguinTweety, birdTweety, flyTweety, {\sim}flyTweety\}.$$

This equilibrium model precisely satisfies the two arguments in that stable extension: $\mathcal{I}_{12} \models penguinTweety$ and $\mathcal{I}_{12} \models birdTweety$. Note that we get $\mathcal{I}_{12} \not\models flyTweety$ from the fact that $\mathcal{I}_{12} \models^+ {\sim}flyTweety$. In fact, this correspondence holds for any EBAF as shown by Theorem 4 below. Though more technically complex, the proof of Theorem 4 is similar that those of Theorems 2 and 3. In particular, it is necessary to prove the following relation between equilibrium models and supportable arguments:

Proposition 11

Let **EF** be some framework and $\mathcal{I}$ be some equilibrium model of $\mathcal{C}_{EF}(\mathbf{EF})$. Then, the following statement holds:

(i) a is supportable w.r.t. $E_{\mathcal{I}}$ iff $\mathcal{I} \models^+ a$.

In contrast with the results for supported arguments stated in Proposition 10, this property does not hold for arbitrary $\leq$-minimal models. This fact can be illustrated by considering a simple $\mathbf{EF}_{13}$ such that $\mathcal{C}_{EF}(\mathbf{EF}_{13}) = \{a,\ a \Rightarrow b\}$. Let $\mathcal{I}_{13} = \langle \mathbf{H}_{13}, \mathbf{T}_{13} \rangle$ be some interpretation with $\mathbf{H}_{13} = \{a\}$ and $\mathbf{T}_{13} = \{a, {\sim}a\}$. It is easy to see that $\mathcal{I}_{13}$ is a $\leq$-minimal model of $\mathcal{C}_{EF}(\mathbf{EF}_{13})$, though it is not an equilibrium model (because it is not a total interpretation). It can also be checked that a is not defeated and, consequently, that b is supportable w.r.t. $E_{\mathcal{I}_{13}} = \varnothing$. On the other hand, the unique equilibrium model of $\mathcal{C}_{EF}(\mathbf{EF}_{13})$ is $\mathcal{J}_{13} = \langle \mathbf{H}'_{13}, \mathbf{T}'_{13} \rangle$ with $\mathbf{H}'_{13} = \{a, b\}$ and $\mathbf{T}'_{13} = \{a, b\}$. Here, both a and b are supportable (and supported) w.r.t. $E_{\mathcal{J}_{13}} = \{a, b\}$.

The following result shows that, indeed, this correspondence holds for any EBAF:

Theorem 4
Given some finitary $\mathbf{EF}$, there is a one-to-one correspondence between its stable extensions and the equilibrium models of $\mathcal{C}_{EF}(\mathbf{EF})$ such that

(i) if $\mathcal{I}$ is an equilibrium model of $\mathcal{C}_{EF}(\mathbf{EF})$, then $E_{\mathcal{I}}$ is a stable extension of $\mathbf{EF}$,
(ii) if E is a stable extension of $\mathbf{EF}$ and $\mathcal{I}$ is a total interpretation such that $T^{+}_{\mathcal{I}} = Sup(E)$ and $T^{-}_{\mathcal{I}} = Def(E)$, then $\mathcal{I}$ is an equilibrium model of $\mathcal{C}_{EF}(\mathbf{EF})$.

5 Translation of $\mathcal{C}_{LP}$-program to regular programs

In this section, we show how $\mathcal{C}_{LP}$-programs can be translated into regular ASP programs. An important practical consequence of this fact is that current state-of-the-art ASP solvers (Faber *et al.* 2008; Gebser *et al.* 2012) can be applied to $\mathcal{C}_{LP}$-programs. Let us introduce such a translation as follows:

Definition 9
Given a $\mathcal{C}_{LP}$-program P, by δP we denote the result of

1. replacing every positive literal a in the body of a rule by $a \wedge \neg{\sim}a$,
2. replacing every negative literal *not* a in the body of a rule by $\neg a \vee (a \wedge {\sim}a)$,
3. replacing all occurrences of $\Rightarrow$ by $\rightarrow$.

Proposition 12
Any $\mathcal{C}_{LP}$-program P and interpretation $\mathcal{I}$ satisfy: $\mathcal{I} \models^{+} P$ iff $\mathcal{I} \models^{+} \delta P$.

Proposition 12 shows how we can translate any $\mathcal{C}_{LP}$-program into an equivalent theory that does not use the new connectives *not* and $\Rightarrow$. The result of the translation in Definition 9 is almost a standard logic program, but for two points. First, strong negation has to be understood in a paraconsistent way, so an atom can be true and false at the same time. This can be addressed by using new auxiliary atoms to represent strongly negated atoms.[6] Second, step 2 introduces a disjunction in the body, which is not allowed in the standard syntax of logic programs. This can be addressed in polynomial-time also

[6] In fact, modern solvers already allow the use of explicit negation and their implementation is done by using new auxiliary atoms to represent strongly negated atoms. However, solvers also include a constraint of the form $a \wedge {\sim}a \rightarrow \bot$ for every atom a. This would remove the non-consistent answer sets, something we have to avoid to obtain paraconsistent answer sets.

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

by using auxiliary atoms (similar to Tseitin 1968). The following definition addresses these two issues.

Definition 10
Given a $\mathcal{C}_{LP}$-program P, by τP we denote the result of applying the following transformations to δP:

1. replacing every explicit literal of the form $\sim a$ by a fresh atom $\tilde{a}$,
2. adding rules $a' \leftarrow \neg a$ and $a' \leftarrow a \wedge \tilde{a}$ for each atom $a \in At$ with a' a new fresh atom, and
3. replacing each occurrence of $\neg a \vee (a \wedge \tilde{a})$ in the body of any rule by a'.

Given a total interpretation $\mathcal{I}$, we also denote by $\tau\mathcal{I}$ an interpretation that, for every atom $a \in At$, satisfies:

1. $\tau\mathcal{I} \not\models^{-} a$
2. $\tau\mathcal{I} \models^{+} a$ iff $\mathcal{I} \models^{+} a$
3. $\tau\mathcal{I} \models^{+} \tilde{a}$ iff $\mathcal{I} \models^{-} a$
4. $\tau\mathcal{I} \models^{+} a'$ iff either $\mathcal{I} \not\models^{+} a$ or both $\mathcal{I} \models^{+} a$ and $\mathcal{I} \models^{-} a$.

Proposition 13
Any $\mathcal{C}_{LP}$-program P and total interpretation $\mathcal{I}$ satisfy that $\mathcal{I}$ is an equilibrium model of P iff τI an equilibrium model of τP.

The result of Definition 10 is a standard logic program. Proposition 13 shows that we can use this translation in combination with standard ASP solvers to obtain equilibrium for $\mathcal{C}_{LP}$-program and stable extensions of all the AFs considered in this paper. The second consequence of this translation is that deciding whether there exists any stable extension of some $\mathcal{C}_{LP}$-program is in Σ_2^{P} in general and in NP if the program is normal (Dantsin *et al.* 2001). This complexity results are tight because hardness follows from Corollary 1 and the hardness results for finding answer sets for these classes of programs (Dantsin *et al.* 2001). Therefore, deciding whether there exists any stable extension of some $\mathcal{C}_{LP}$-program is Σ_2^{P}-complete in general and NP-complete for normal $\mathcal{C}_{LP}$-programs. Furthermore, this result directly applies to EBAFs so that deciding whether there exists any stable extension is NP-complete.

6 Discussion

LP and AFs are two well-established KRR formalisms for dealing with nonmonotonic reasoning (NMR). In particular, Answer Set Programming (ASP) is an LP paradigm, based on the stable model semantics, which has raised as a preeminent tool for practical NMR with applications in diverse areas of AI including planning, reasoning about actions, diagnosis, abduction and beyond (Baral 2003; Brewka *et al.* 2011). On the other hand, one of the major reasons for the success of AFs is their ability to handle conflicts due to inconsistent information.

Here, we have shown that both formalisms can be successfully accommodated in Nelson's constructive logic. In fact, it is easy to see that by rewriting attacks using definition (7), the translation of any AF becomes a normal $\mathcal{C}_{LP}$-program. For instance, by rewriting the attack (13), we obtain the equivalent formula:

$$penguinTweety \Rightarrow \sim\!flyTweety, \tag{14}$$

which is a $\mathcal{C}_{LP}$-rule. In this sense, we can consider $\mathcal{C}_{SF}(\mathbf{EF}_1)$ in Example 9 as an alternative representation of the Tweety scenario in LP. Note that both the unique equilibrium model $\mathcal{I}_9$ of program P_9 (Example 5) and the unique equilibrium model $\mathcal{I}_{12}$ of this program satisfy:

$$\begin{array}{ll} \mathcal{I}_9 \not\models flyTweety & \mathcal{I}_{12} \not\models flyTweety \\ \mathcal{I}_9 \models not\ flyTweety & \mathcal{I}_{12} \models not\ flyTweety. \end{array}$$

In other words, in both programs we conclude that Tweety cannot fly. However, there are a couple of differences between these two representations. First, in contrast with $\mathcal{I}_9$, we have that $\mathcal{I}_{12}$ is not consistent: $\mathcal{I}_{12} \models^+ flyTweety$ and $\mathcal{I}_{12} \models^+ \sim\!flyTweety$. Second and perhaps more interestingly, in $\mathcal{C}_{SF}(\mathbf{EF}_1)$, the "normality" of the statement "birds can fly" does not need to be explicitly represented. Instead, this normality is implicitly handled by the strong closed word assumption **CW**, which resolves the contradictory evidence for *flyTweety* by regarding it as false. In this sense, $\mathcal{C}_{LP}$-programs and AFs can be seen as two different syntaxes of the same formalism based on the principles **NC** and **CW** highlighted in the introduction. In addition, another principle of this formalism is the fact that evidence must be founded or justified: this clearly shows up in normal LP and EBAFs where true literals can be computed by some recursive procedure, but also in Dung's AFs where, as we have seen, defeat can be understood as a proof of falsity.

Regarding practical aspects, we can use $\mathcal{C}_{LP}$-programs as a unifying formalism to deal with both logic programs and AFs. This directly allows to introduce variables in AFs through the use of grounding. Going further, full first-order characterizations of AFs can be provided by applying the same principles to first-order constructive logic (full first-order characterization of consistent logic programs has been already provided by Pearce and Valverde 2004). Besides, constructive logic immediately provides an interpretation for other richer syntaxes like the use of disjunctive targets in Collective Argumentation (Bochman 2003) or the use of arbitrary propositional formulas to represent attacks in Abstract Dialectical Frameworks (Brewka and Woltran 2010; Brewka *et al.* 2013).

7 Conclusion and future work

We have formalized the principles **NC** and **CW** in Nelson's constructive logic and shown that this is a conservative extension of logic programs which allow us to reason with contradictory evidence. Furthermore, this allows us to translate argumentation frameworks in a modular way and using the object language such that attacks and supports become connectives in logic using the object level. As a consequence, we can combine both formalisms in an unifying one and use proof methods from the logic or answer set solver to reason about it.

Regarding future work, an obvious open topic is to explore how other argumentation semantics can be translated into the logic. For instance, the relation between the

https://doi.org/10.1017/S1471068422000114 Published online by Cambridge University Press

complete semantics for AFs, three-valued stable models semantics for LP (Przymusinski 1991; Y. *et al.* 2009) and partial equilibrium logic (Cabalar *et al.* 2007) suggest that our framework can be extended to cover other semantics such as the complete and preferred. Similarly, the relation between the paracoherent semantics for AFs (Amendola and Ricca 2019) and semiequilibrium models (Amendola *et al.* 2016) suggest a possible direction to capture this semantics using the object level. It will also be interesting to see the relation with the semistable semantics for AFs (Caminada *et al.* 2012). The relation with other AFs extensions such as Collective Argumentation (Bochman 2003), Abstract Dialectical Frameworks (Brewka and Woltran 2010; Brewka *et al.* 2013) or Recursive Argumentation Frameworks (Barringer *et al.* 2005; Modgil 2009; Gabbay 2009; Baroni *et al.* 2011; Cayrol *et al.* 2016; 2021) is also a direction worth exploring. Another important open questions are studying how the principles **NC** and **CW** stand in the context of paraconsistent logics (da Costa 1974) and paraconsistent logic programming (Blair and Subrahmanian 1989); and studying the notion of strong equivalence (Lifschitz *et al.* 2001; Oikarinen and Woltran 2011) in this logic and evidence-based frameworks.

Competing interests

The authors declare none.

References

Akama, S. 1987. Constructive predicate logic with strong negation and model theory. *Notre Dame Journal of Formal Logic 29,* 1 (12), 18–27.

Almukdad, A. and Nelson, D. 1984. Constructible falsity and inexact predicates. *The Journal of Symbolic Logic 49,* 1, 231–233.

Amendola, G., Eiter, T., Fink, M., Leone, N. and Moura, J. 2016. Semi-equilibrium models for paracoherent answer set programs. *Artificial Intelligence 234*, 219–271.

Amendola, G. and Ricca, F. 2019. Paracoherent answer set semantics meets argumentation frameworks. *Theory and Practice of Logic Programming 19,* 5–6, 688–704.

Amgoud, L., Cayrol, C. and Lagasquie-Schiex, M.-C. 2004. On the bipolarity in argumentation frameworks. In *NMR 2004, Proceedings*, J. P. Delgrande and T. Schaub, Eds. 1–9.

Arieli, O. and Caminada, M. W. 2013. A qbf-based formalization of abstract argumentation semantics. *Journal of Applied Logic 11,* 2, 229–252.

Baral, C. 2003. *Knowledge Representation, Reasoning and Declarative Problem Solving.*

Baroni, P., Cerutti, F., Giacomin, M. and Guida, G. 2011. AFRA: argumentation framework with recursive attacks. *International Journal of Approximate Reasoning 52,* 1, 19–37.

Barringer, H., Gabbay, D. and Woods, J. 2005. Temporal dynamics of support and attack networks: From argumentation to zoology. In *Mechanizing Mathematical Reasoning.* LNAI, vol. 2605. Springer Verlag, 59–98.

Besnard, P. and Doutre, S. 2004. Checking the acceptability of a set of arguments. In *10th International Workshop on Non-Monotonic Reasoning (NMR 2004), Whistler, Canada, June 6–8, 2004, Proceedings*, J. P. Delgrande and T. Schaub, Eds. 59–64.

Besnard, P., Doutre, S. and Herzig, A. 2014. Encoding argument graphs in logic. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems - 15th International Conference, IPMU 2014, Montpellier, France, July 15-19, 2014, Proceedings,* *Part II*, A. Laurent, O. Strauss, B. Bouchon-Meunier and R. R. Yager, Eds. Communications in Computer and Information Science, vol. 443. Springer, 345–354.

Blair, H. and Subrahmanian, V. 1989. Paraconsistent logic programming. *Theoretical Computer Science 68,* 2, 135–154.

Bochman, A. 2003. Collective argumentation and disjunctive logic programming. *Journal of Logic and Computation 13,* 3, 405–428.

Brewka, G., Eiter, T. and Truszczynski, M. 2011. Answer set programming at a glance. *Communications of the ACM 54,* 12, 92–103.

Brewka, G., Strass, H., Ellmauthaler, S., Wallner, J. P. and Woltran, S. 2013. Abstract dialectical frameworks revisited. In *IJCAI 2013, Proceedings*, F. Rossi, Ed. IJCAI/AAAI, 803–809.

Brewka, G. and Woltran, S. 2010. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010*, F. Lin, U. Sattler and M. Truszczynski, Eds. AAAI Press.

Cabalar, P., Fandinno, J., Fariñas del Cerro, L., Pearce, D. and Valverde, A. 2017. On the properties of atom definability and well-supportedness in logic programming. In *EPIA 2017, Proceedings*, E. C. Oliveira, J. Gama, Z. A. Vale and H. L. Cardoso, Eds. Springer, 624–636.

Cabalar, P., Odintsov, S., Pearce, D. and Valverde, A. 2007. Partial equilibrium logic. *Annals of Mathematics and Artificial Intelligence 50,* 3–4, 305–331.

Caminada, M., Carnielli, W. and Dunne, P. 2012. Semi-stable semantics. *Journal of Logic and Computation 22,* 5, 1207–1254.

Caminada, M., Sá, S., Alcântara, J. and Dvorák, W. 2015. On the equivalence between logic programming semantics and argumentation semantics. *International Journal of Approximate Reasoning 58*, 87–111.

Caminada, M. W. A. and Gabbay, D. M. 2009. A logical account of formal argumentation. *Studia Logica 93,* 2 (11), 109.

Cayrol, C., Cohen, A. and Lagasquie-Schiex, M.-C. 2016. Towards a new framework for recursive interactions in abstract bipolar argumentation. In *Proceedings of COMMA*, 191–198.

Cayrol, C., Fandinno, J., Fariñas del Cerro, L. and Lagasquie-Schiex, M. 2021. Valid attacks in argumentation frameworks with recursive attacks. *Annals of Mathematics and Artificial Intelligence 89*, 1573–7470.

Cayrol, C., Fandinno, J., Fariñas del Cerro, L. and Lagasquie-Schiex, M.-C. 2018. Argumentation Frameworks with Recursive Attacks and Evidence-Based Supports. Rapport de recherche IRIT/RR–2018–01–FR, IRIT, Universite Paul Sabatier, Toulouse. 1.

Cayrol, C., Fandinno, J., Fariñas del Cerro, L. and Lagasquie-Schiex, M.-C. 2018. Argumentation frameworks with recursive attacks and evidence-based supports. In *FoIKS 2018, Proceedings.*

da Costa, N. 1974. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic 15*, 497–510.

Dantsin, E., Eiter, T., Gottlob, G. and Voronkov, A. 2001. Complexity and expressive power of logic programming. *ACM Computing Surveys 33,* 3, 374–425.

Doutre, S., Herzig, A. and Perrussel, L. 2014. A dynamic logic framework for abstract argumentation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20–24, 2014*, C. Baral, G. D. Giacomo and T. Eiter, Eds. AAAI Press.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence 77,* 2, 321–358.

Dvorák, W., Gaggl, S. A., Linsbichler, T. and Wallner, J. P. 2015. Reduction-based approaches to implement modgil's extended argumentation frameworks. In *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation – Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, T. Eiter, H. Strass, M. Truszczynski and S. Woltran, Eds. Lecture Notes in Computer Science, vol. 9060. Springer, 249–264.

Dvorák, W., Gaggl, S. A., Wallner, J. P. and Woltran, S. 2011. Making use of advances in answer-set programming for abstract argumentation systems. In *Applications of Declarative Programming and Knowledge Management – 19th International Conference, INAP 2011, and 25th Workshop on Logic Programming, WLP 2011, Vienna, Austria, September 28–30, 2011, Revised Selected Papers*, H. Tompits, S. Abreu, J. Oetsch, J. Pührer, D. Seipel, M. Umeda and A. Wolf, Eds. Lecture Notes in Computer Science, vol. 7773. Springer, 114–133.

Dvorák, W., Szeider, S. and Woltran, S. 2012. Abstract argumentation via monadic second order logic. In *Scalable Uncertainty Management - 6th International Conference, SUM 2012, Marburg, Germany, September 17–19, 2012. Proceedings*, E. Hüllermeier, S. Link, T. Fober and B. Seeger, Eds. Lecture Notes in Computer Science, vol. 7520. Springer, 85–98.

Faber, W., Pfeifer, G., Leone, N., Dell'Armi, T. and Ielpa, G. 2008. Design and implementation of aggregate functions in the DLV system. *Theory and Practice of Logic Programming 8,* 5–6, 545–580.

Fandinno, J. and Fariñas del Cerro, L. 2018. Constructive logic covers argumentation and logic programming. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018*, M. Thielscher, F. Toni and F. Wolter, Eds. AAAI Press, 128–137.

Gabbay, D. and Gabbay, M. 2016. The attack as intuitionistic negation. *Logic Journal of the IGPL 24,* 5, 807–837.

Gabbay, D. M. 2009. Semantics for higher level attacks in extended argumentation frames part 1: Overview. *Studia Logica 93,* 2, 357.

Gabbay, D. M. and Gabbay, M. 2015. The attack as strong negation, part i. *Logic Journal of the IGPL 23*, 881–941.

Gebser, M., Kaufmann, B. and Schaub, T. 2012. Conflict-driven answer set solving: From theory to practice. *Artificial Intelligence 187–188*, 52–89.

Gelfond, M. and Lifschitz, V. 1988. The stable model semantics for logic programming. In *Logic Programming: Proceedings of the Fifth International Conference and Symposium (Volume 2)*.

Gelfond, M. and Lifschitz, V. 1991. Classical negation in logic programs and disjunctive databases. *New Generation Computing 9,* 3/4, 365–386.

Grossi, D. 2011. Argumentation in the view of modal logic. In *Argumentation in Multi-Agent Systems*, P. McBurney, I. Rahwan and S. Parsons, Eds. Springer Berlin Heidelberg, 190–208.

Gurevich, Y. 1977. Intuitionistic logic with strong negation. *Studia Logica 36,* 1-2, 49–59.

Kamide, N. and Wansing, H. 2015. *Proof Theory of N4-Related Paraconsistent Logics*. College Publications London.

Karacapilidis, N. and Papadias, D. 2001. Computer supported argumentation and collaborative decision making: The hermes system. *Information Systems 26,* 4, 259–277.

Lifschitz, V., Pearce, D. and Valverde, A. 2001. Strongly equivalent logic programs. *ACM Transactions on Computational Logic 2,* 4, 526–541.

López-Escobar, E. 1972. Refutability and elementary number theory. *Indagationes Mathematicae (Proceedings) 75,* 4, 362 – 374.

Markov, A. 1953. A constructive logic.

Modgil, S. 2009. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence 173,* 9–10, 901–934.

NELSON, D. 1949. Constructible falsity. *Journal of Symbolic Logic 14,* 1 (03), 16–26.

NELSON, D. 1959. Negation and separation of concepts in constructive systems. *Constructivity in Mathematics*, 208–225.

NIELSEN, S. H. AND PARSONS, S. 2007. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Argumentation in Multi-Agent Systems*, N. Maudet, S. Parsons and I. Rahwan, Eds. Berlin, Heidelberg, 54–73.

NIEVES, J. C., CORTÉS, U. AND OSORIO, M. 2008. Preferred extensions as stable models. *TPLP 8,* 4, 527–543.

ODINTSOV, S. AND RYBAKOV, V. 2015. Inference rules in Nelson's logics, admissibility and weak admissibility. *Logica Universalis 9,* 1, 93–120.

ODINTSOV, S. P. 2005. The class of extensions of Nelson's paraconsistent logic. *Studia Logica 80,* 2(8), 291–320.

ODINTSOV, S. P. AND PEARCE, D. 2005. Routley semantics for answer sets. In *LPNMR 2005, Proceedings*, C. Baral, G. Greco, N. Leone and G. Terracina, Eds. Springer, 343–355.

OIKARINEN, E. AND WOLTRAN, S. 2011. Characterizing strong equivalence for argumentation frameworks. *Artificial Intelligence 175,* 14-15, 1985–2009.

OREN, N. AND NORMAN, T. 2008. Semantics for evidence-based argumentation. In *COMMA 2008, Proceedings.*, P. Besnard, S. Doutre and A. Hunter, Eds. 276–284.

OSORIO, M., PÉREZ, J. A. N. AND ARRAZOLA, J. 2005. Safe beliefs for propositional theories. *Annals of Pure and Applied Logic 134,* 1, 63–82.

PEARCE, D. 1996. A new logical characterisation of stable models and answer sets. In *NMELP 1996, Selected Papers*, J. Dix, L. M. Pereira and T. C. Przymusinski, Eds. Springer, 57–70.

PEARCE, D. 2006. Equilibrium logic. *Annals of Mathematics and Artificial Intelligence 47,* 1–2, 3–41.

PEARCE, D. AND VALVERDE, A. 2004. Towards a first order equilibrium logic for nonmonotonic reasoning. In *JELIA 2004, Proceedings*, J. J. Alferes and J. A. Leite, Eds. Lecture Notes in Computer Science, vol. 3229. Springer, 147–160.

POLBERG, S. AND OREN, N. 2014. Revisiting support in abstract argumentation systems. Tech. rep., TU Wien, Institut for Informatics.

PRZYMUSINSKI, T. 1991. Three-valued nonmonotonic formalisms and semantics of logic programs. *Artificial Intelligence 49,* 1-3, 309–343.

RASIOWA, H. 1969. N-lattices and constructive logic with strong negation.

REITER, R. 1980. A logic for default reasoning. *Artificial Intelligence 13,* 1–2, 81–132.

ROUTLEY, R. 1974. Semantical analyses of propositional systems of Fitch and Nelson. *Studia Logica 33,* 3, 283–298.

SAKAMA, C. AND INOUE, K. 1995. Paraconsistent stable semantics for extended disjunctive programs. *Journal of Logic and Computation 5,* 3, 265–285.

THOMASON, R. H. 1969. A semantical study of constructible falsity. *Mathematical Logic Quarterly 15,* 16–18, 247–257.

TONI, F. AND SERGOT, M. 2011. *Argumentation and Answer Set Programming.* Springer Berlin Heidelberg, Berlin, Heidelberg, 164–180.

TSEITIN, G. 1968. On the complexity of derivation in the propositional calculus. *Zapiski nauchnykh seminarov LOMI 8*, 234–259.

VAN GELDER, A., ROSS, K. A. AND SCHLIPF, J. S. 1991. The well-founded semantics for general logic programs. *Journal of the ACM (JACM) 38,* 3, 619–649.

VERHEIJ, B. 2003a. Deflog: On the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation 13,* 3, 319–346.

VERHEIJ, B. 2003b. Deflog: On the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation 13,* 3, 319–346.

Vorob'ev, N. 1952. A constructive propositional calculus with strong negation. In *Doklady Akademii Nauk SSR*. Vol. 85. 465–468.

Wu, Y. and Caminada, M. 2010. A labelling-based justification status of arguments.

Caminada, M. and Gabbay, D. 2009. Complete extensions in argumentation coincide with 3-valued stable models in logic programming. *Studia Logica 93,* 2–3, 383–403.