Counterfactual Reasoning in Argumentation Frameworks

Chiaki SAKAMA [1]
Wakayama University, Japan

Abstract. In a formal argumentation framework, one is interested in whether a particular argument is accepted or not under argumentation semantics. When an argument A is accepted, on the other hand, one may ask a question "what if A were rejected?" We formulate such *counterfactual reasoning* in abstract argumentation frameworks. Based on Lewis's logic, we define two counterfactual conditionals in AF and investigate formal properties. We also argue counterfactual dependencies in AF and modal interpretation of AF in terms of counterfactual conditionals.

Keywords. argumentation framework, counterfactuals, causality

1. Introduction

A *counterfactual* is a conditional statement representing what would be the case if its premise were true (although it is not true in fact). For instance, "if Dung had not published the seminal paper on formal argumentation in 1995, argumentation in AI would not have been so popular as it is today" is a counterfactual statement. A formal model of counterfactuals is firstly developed by Stalnaker [27], in which he provides the *possible worlds semantics* for conditional sentences. Stalnaker interprets that a conditional sentence $\varphi > \psi$ is true if ψ is true at the world most like the actual world at which φ is true. Lewis [18,19] provides in-depth analyses of counterfactuals and causation. Since then counterfactuals have been studied in philosophy [22] and artificial intelligence [14], and have practically been used in cognitive psychology [8], political science [17] and social science [20].

Counterfactual reasoning is also used in human dialogue or argument in daily life. For instance, consider a dialogue in [24]:

> Paul: "My car is safe because it has an airbag."
> Olga: "I disagree that an airbag makes your car safe because newspapers recently reported on airbags expanding without cause."

In this dialogue, Paul's argument is refuted by Olga's counterargument. If Paul has no evidence to refute Olga's argument, he would accept her opinion. On the other hand, Paul may wonder "What if the news source is unreliable?" or may argue "If the news source were unreliable, then a car with an airbag would be safe", although he has no reason to

[1] Correspondence to: Department of Computer and Communication Sciences, Wakayama University, Sakaedani, Wakayama 640-8510, Japan; E-mail: sakama@sys.wakayama-u.ac.jp.

believe that the news source is unreliable. Such counterfactual arguments are used for challenging arguments by the opponent and requesting further evidence for justification.

In legal settings, it is a common practice to use counterfacuals for defeating assumptions that are incompatible with reality. For instance, suppose an argument by a defense in a court: "If the suspect killed the victim, the suspect would be at the victim's apartment at the time of the murder." Then the defense shows an evidence that the suspect was in fact at a party which was held far from the victim's apartment. Legal researches also study the effect of using counterfactuals on jurors' perception on the responsibility of an accused person.

> Counterfactual conditionals supposing the wrong to be absent in the antecedent, and deriving more desirable consequences than in reality in the consequent (e.g. *If that drunken driver had not neglected the stop sign, my client would have had her whole life in front of her*) evoke an emotional response from the jury, which leads them to attribute more responsibility, guilt and blame to the party responsible for the wrong. [21]

As such, counterfactuals are popularly used in dialogue, argument or dispute, while little attention has been paid on investigating counterfactuals in formal argumentation. The purpose of this paper is to provide an argumentation-theoretic interpretation of counterfactuals. An *argumentation framework* (AF) [11] represents knowledge about the world using arguments and attack relations over them. Given an argumentation framework, an argumentation semantics specifies which arguments are accepted or rejected. Accepting one argument may cause rejecting another argument and the other way round. Then one may ask "what if an accepted argument were rejected?" or "what if a rejected argument were accepted?" We formulate such counterfactual arguments in Dung's abstract argumentation framework and investigate formal properties.

The rest of this paper is organized as follows. Section 2 reviews basic notions of formal argumentation. Section 3 introduces counterfactual reasoning in AF and investigates formal properties. Section 4 argues counterfactual dependencies in argumentation frameworks. Section 5 discusses related issues and rounds off the paper.

2. Argumentation Framework

This section reviews formal argumentation frameworks which are in [9,11].

Definition 1 (argumentation framework) Let U be the universe of all possible *arguments*. An *argumentation framework* (AF) is a pair (Ar, att) where Ar is a finite subset of U and $att \subseteq Ar \times Ar$. An argument A *attacks* an argument B iff $(A, B) \in att$.

An argumentation framework (Ar, att) is associated with a directed graph in which vertices are arguments in Ar and directed arcs from A to B exist whenever $(A, B) \in att$.

Definition 2 (indirect attack/defend) Let $AF = (Ar, att)$ and $A, B \in Ar$.

- *A indirectly attacks B* if there is an odd-length path from A to B in a directed graph associated with AF.
- *A indirectly defends B* if there is an even-length (non-zero) path from A to B in a directed graph associated with AF.

Definition 3 (labelling) A *labelling* of $AF = (Ar, att)$ is a (total) function $\mathcal{L} : Ar \rightarrow \{ \texttt{in}, \texttt{out}, \texttt{undec} \}$.

When $\mathcal{L}(A) = \texttt{in}$ (resp. $\mathcal{L}(A) = \texttt{out}$ or $\mathcal{L}(A) = \texttt{undec}$) for $A \in Ar$, it is written as $\texttt{in}(A)$ (resp. $\texttt{out}(A)$ or $\texttt{undec}(A)$). In this case, the argument A is *accepted* (resp. *rejected* or *undecided*) in $\mathcal{L}$.

Definition 4 (complete labelling) A labelling $\mathcal{L}$ of $AF = (Ar, att)$ is a *complete labelling* if for each argument $A \in Ar$, it holds that:

- $\mathcal{L}(A) = \texttt{in}$ iff $\mathcal{L}(B) = \texttt{out}$ for every $B \in Ar$ such that $(B, A) \in att$.
- $\mathcal{L}(A) = \texttt{out}$ iff $\mathcal{L}(B) = \texttt{in}$ for some $B \in Ar$ such that $(B, A) \in att$.
- $\mathcal{L}(A) = \texttt{undec}$ iff $\mathcal{L}(A) \neq \texttt{in}$ and $\mathcal{L}(A) \neq \texttt{out}$.

We say that an argument A is *accepted* (resp. *rejected* or *undecided*) in AF if A is labelled $\texttt{in}$ (resp. $\texttt{out}$ or $\texttt{undec}$) in every complete labelling $\mathcal{L}$ of AF. Let $\texttt{in}(\mathcal{L}) = \{A \mid \mathcal{L}(A) = \texttt{in}\}$, $\texttt{out}(\mathcal{L}) = \{A \mid \mathcal{L}(A) = \texttt{out}\}$ and $\texttt{undec}(\mathcal{L}) = \{A \mid \mathcal{L}(A) = \texttt{undec}\}$.

Definition 5 (stable, semi-stable, grounded, preferred labelling) Let $\mathcal{L}$ be a complete labelling of AF. Then

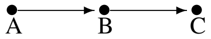
- $\mathcal{L}$ is a *stable labelling* iff $\texttt{undec}(\mathcal{L}) = \emptyset$.
- $\mathcal{L}$ is a *semi-stable labelling* iff $\texttt{undec}(\mathcal{L})$ is minimal wrt set inclusion among all complete labellings of AF.
- $\mathcal{L}$ is a *grounded labelling* iff $\texttt{in}(\mathcal{L})$ is minimal wrt set inclusion among all complete labellings of AF.
- $\mathcal{L}$ is a *preferred labelling* iff $\texttt{in}(\mathcal{L})$ is maximal wrt set inclusion among all complete labellings of AF.

A labelling $\mathcal{L}$ is *universally defined* if every AF has at least one $\mathcal{L}$. A complete (or semi-stable, grounded, preferred) labelling is universally defined, while a stable labelling is not. There is a one-to-one correspondence between the set $\texttt{in}(\mathcal{L})$ with a complete (resp. stable, semi-stable, grounded, preferred) labelling $\mathcal{L}$ of AF and a *complete* (resp. *stable*, *semi-stable*, *grounded*, *preferred*) *extension* of AF. In this paper, when we simply say "labelling" it means one of the five labellings introduced above.[2]

3. Counterfactual Reasoning in AF

Suppose $AF_1 = (\{A, B, C\}, \{(A, B), (B, C)\})$ which has the complete labelling $\{\texttt{in}(A), \texttt{out}(B), \texttt{in}(C)\}$. Then A and C are accepted, and B is rejected in AF_1. In this case, one can reason that:
"If A were rejected, then B would be accepted,"
"If A were rejected, then C would be rejected," or
"If B were accepted, then C would be rejected."
All these sentences are counterfactuals.

AF_1



To provide semantical grounds for those sentences, AF is modified to another AF' such that an argument A which is accepted in AF is rejected in AF'; or an argument B which is rejected in AF is accepted in AF'. There are several ways to construct such

[2] The result of this paper is directly extended to other argumentation semantics as well.

AF' from AF, and we choose AF' which is most similar to AF. To change $\mathtt{in}(A)$ in AF into $\mathtt{out}(A)$ in AF', we introduce a new argument X and an attack relation (X, A) to AF. The idea behind this modification is that if there exists a new argument X and an attack relation (X, A), the argument A will turn out. By contrast, to change $\mathtt{out}(A)$ in AF into $\mathtt{in}(A)$ in AF', we remove every attack relation (X, A) from AF. The idea behind this modification is that if no argument attacks A, the argument A will turn in. Such a modification is formally defined as follows.

Definition 6 (modification of AF) Let $AF = (Ar, att)$ and $A \in Ar$.

$$AF^c_{+A} = (Ar,\, att \setminus \{\, (X, A) \mid X \in Ar \,\}),$$
$$AF^c_{-A} = (Ar \cup \{X\}, att \cup \{\, (X, A)\}) \text{ where } X \in U \setminus Ar \text{ and } U \setminus Ar \neq \emptyset.$$

AF^c_{+A} and AF^c_{-A} are simply written AF^c if the argument A is clear in the context.

By definition, AF^c_{+A} is obtained by removing all attack relations $(X, A) \in att$ from AF. This makes $\mathcal{L}(A) = \mathtt{in}$ for every labelling $\mathcal{L}$, if any, in AF^c_{+A}. On the other hand, AF^c_{-A} is obtained by introducing a new argument X and an attack relation (X, A) to AF. This makes $\mathcal{L}(A) = \mathtt{out}$ for every labelling $\mathcal{L}$, if any, in AF^c_{-A}. The newly introduced argument X is not in AF but in the universe U of possible arguments.

Lewis [18] introduces two different types of counterfactual sentences. Given two different events φ and ψ, "*if it were the case that φ, then it would be the case that ψ*" (written $\varphi \,\Box\!\!\rightarrow\, \psi$) and "*If it were the case that φ, then it might be the case that ψ.*" (written $\varphi \,\Diamond\!\!\rightarrow\, \psi$). Here $\varphi \,\Box\!\!\rightarrow\, \psi$ implies $\varphi \,\Diamond\!\!\rightarrow\, \psi$. We consider similar types of counterfactuals in AF.

Definition 7 (counterfactuals in AF) Let $AF = (Ar, att)$, $A, B \in Ar$ and $\ell \in \{\mathtt{in}, \mathtt{out}\}$.

- $\mathtt{in}(A) \,\Box\!\!\rightarrow\, \ell(B)$ is *true* in AF if $\mathcal{L}(B) = \ell$ in every labelling $\mathcal{L}$ of AF^c_{+A}
- $\mathtt{in}(A) \,\Diamond\!\!\rightarrow\, \ell(B)$ is *true* in AF if $\mathcal{L}(B) = \ell$ in some labelling $\mathcal{L}$ of AF^c_{+A}
- $\mathtt{out}(A) \,\Box\!\!\rightarrow\, \ell(B)$ is *true* in AF if $\mathcal{L}(B) = \ell$ in every labelling $\mathcal{L}$ of AF^c_{-A}
- $\mathtt{out}(A) \,\Diamond\!\!\rightarrow\, \ell(B)$ is *true* in AF if $\mathcal{L}(B) = \ell$ in some labelling $\mathcal{L}$ of AF^c_{-A}

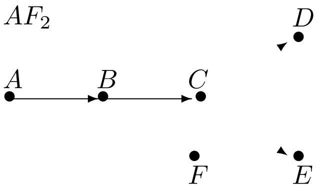
The "labelling" means one of the five labellings introduced in Section 2. The relations "$\ell_1(A) \,\Box\!\!\rightarrow\, \ell_2(B)$" and "$\ell_1(A) \,\Diamond\!\!\rightarrow\, \ell_2(B)$" are called *counterfactual conditionals* where ℓ_1 or ℓ_2 is either in or out. We do not consider counterfactual conditionals which include arguments with the labelling undec in this paper, and ℓ_i means either in or out throughout the paper. We call $\ell_1(A)$ the *antecedent* and $\ell_2(B)$ the *consequent* of the conditional.

Definition 8 (negation, etc) Let C_1 and C_2 be two counterfactual conditionals in AF.

- $\neg C_1$ is *true* in AF if C_1 is not true in AF.
- $C_1 \vee C_2$ is *true* in AF if C_1 or C_2 is true in AF.
- $C_1 \wedge C_2$ is *true* in AF if both C_1 and C_2 are true in AF.
- $C_1 \supset C_2$ is *true* in AF if $\neg C_1$ or C_2 is true in AF.

For simplicity's sake, "$\neg\, (\ell_1(A) \,\Box\!\!\rightarrow\, \ell_2(B))$" (or "$\neg\, (\ell_1(A) \,\Diamond\!\!\rightarrow\, \ell_2(B))$") is abbreviated as "$\ell_1(A) \,\Box\!\!\not\rightarrow\, \ell_2(B)$" (or "$\ell_1(A) \,\Diamond\!\!\not\rightarrow\, \ell_2(B)$") hereafter.

Example 1 Consider AF_2 illustrated in the right. It has the complete labelling: $\{\,\texttt{in}(A),\ \texttt{out}(B),\ \texttt{in}(C),\ \texttt{out}(D),\ \texttt{out}(E),\ \texttt{in}(F)\,\}$.
Then, the following counterfacuals hold in AF_2: "$\texttt{out}(A) \Box\!\!\rightarrow\ \texttt{in}(B)$" (If A were rejected then B would be accepted), "$\texttt{in}(B) \Box \rightarrow\ \texttt{out}(C)$" (If B were accepted then C would be rejected), "$\texttt{out}(C) \Diamond\!\!\rightarrow\ \texttt{in}(D)$" (If C were rejected then D might be accepted), "$\texttt{out}(C) \Diamond\!\!\rightarrow\ \texttt{in}(E)$" (If C were rejected then E might be accepted), etc.



By Definition 7, "$\texttt{in}(A) \Box\!\!\rightarrow\ \ell(B)$" (resp. "$\texttt{out}(A) \Box\!\!\rightarrow\ \ell(B)$") vacuously holds if AF^c_{+A} (resp. AF^c_{-A}) has no labelling. Such a situation happens under the stable labelling which is not universally defined. Otherwise, AF^c_{+A} (resp. AF^c_{-A}) has a labelling and A is in (resp. out) in every labelling of it. An argument B appearing in the consequent of counterfactual conditionals is labelled ℓ in *every* (resp. *some*) labelling of AF^c if the conditional operator is $\Box\!\!\rightarrow$ (resp. $\Diamond\!\!\rightarrow$). By definition, "$\ell_1(A) \Box\!\!\rightarrow\ \ell_2(B)$" implies "$\ell_1(A) \Diamond\!\!\rightarrow\ \ell_2(B)$" whenever AF^c has a labelling, but not the other way round.

In the rest of this section, we investigate formal properties of counterfactuals in AF. First, counterfactual conditionals are reflexive.

Proposition 1 *Let $AF = (Ar, att)$ and $A \in Ar$. Then $\ell(A) \Box\!\!\rightarrow\ \ell(A)$ is true in AF where $\ell \in \{\texttt{in},\ \texttt{out}\}$. $\ell(A) \Diamond\!\!\rightarrow\ \ell(A)$ is also true in AF whenever AF^c has a labelling.*

Proof: The relation $\texttt{in}(A) \Box\!\!\rightarrow\ \texttt{in}(A)$ (resp. $\texttt{out}(A) \Box\!\!\rightarrow\ \texttt{out}(A)$) holds in AF because $\mathcal{L}(A) = \texttt{in}$ (resp. $\mathcal{L}(A) = \texttt{out}$) in every labelling of AF^c_{+A} (resp. AF^c_{-A}). The results imply $\texttt{in}(A) \Diamond\!\!\rightarrow\ \texttt{in}(A)$ and $\texttt{out}(A) \Diamond\!\!\rightarrow\ \texttt{out}(A)$ whenever AF^c has a labelling. □

The antecedent of a counterfactual sentence is usually assumed false. However, counterfactual sentences with true antecedent may happen, for example, in a dialogue where the participants disagree on the truth of the antecedent. Borrowing an example from [18], one says: "If Caspar had come, it would have been a good party". Then the other replies "That is true; for he did, and it was a good party. You didn't see him because you spent the whole time in the kitchen, missing all the fun." According to Lewis, "a counterfactual with true antecedent is true iff the consequent is true". Formally, "if $\varphi \wedge \psi$ is true, then $\varphi \Box\!\!\rightarrow\ \psi$ is true" and "if $\varphi \wedge \neg\psi$ is true, then $\neg\,(\varphi \Box\!\!\rightarrow\ \psi)$ is true". This is also the case in counterfactuals in AF. In what follows, $\mathcal{L}^{(\prime)}$ means $\mathcal{L}$ or $\mathcal{L}'$.

Proposition 2 *Let $AF = (Ar, att)$, $A, B \in Ar$ and $\mathcal{L}^{(\prime)}$ a universally defined labelling.*

1. *If $\mathcal{L}(A) = \ell_1$ and $\mathcal{L}(B) = \ell_2$ in every labelling $\mathcal{L}$ of AF, then $\ell_1(A) \Box\!\!\rightarrow\ \ell_2(B)$ is true in AF.*
2. *If $\mathcal{L}(A) = \ell_1$ in every labelling $\mathcal{L}$ of AF and $\mathcal{L}'(B) \neq \ell_2$ in some labelling $\mathcal{L}'$ of AF, then $\ell_1(A) \Box\!\!\not\rightarrow\ \ell_2(B)$ is true in AF.*
3. *If $\mathcal{L}(A) = \ell_1$ in every labelling $\mathcal{L}$ of AF and $\mathcal{L}'(B) = \ell_2$ in some labelling $\mathcal{L}'$ of AF, then $\ell_1(A) \Diamond\!\!\rightarrow\ \ell_2(B)$ is true in AF.*
4. *If $\mathcal{L}(A) = \ell_1$ in every labelling $\mathcal{L}$ of AF and $\mathcal{L}'(B) = \ell_2$ in no labelling $\mathcal{L}'$ of AF, then $\ell_1(A) \Diamond\!\!\not\rightarrow\ \ell_2(B)$ is true in AF.*

Proof: (1) If $\mathcal{L}(A) = \texttt{in}$ (resp. $\mathcal{L}(A) = \texttt{out}$) in every labelling $\mathcal{L}$ of AF, then $\mathcal{L}'(A) = \texttt{in}$ (resp. $\mathcal{L}(A) = \texttt{out}$) in every labelling $\mathcal{L}'$ of AF^c_{+A} (resp. AF^c_{-A}). Since the modification from AF to AF^c does not change the labelling of A, it does not affect the labelling of B. Then $\mathcal{L}(B) = \ell_2$ in every labelling $\mathcal{L}$ of AF iff $\mathcal{L}'(B) = \ell_2$ in every labelling $\mathcal{L}'$ of AF^c. Hence, $\ell_1(A) \mathrel{\Box\!\!\rightarrow} \ell_2(B)$ holds in AF. The results of (2)–(4) are shown in similar ways. □

In AF_1 at the begging of this section, $\mathcal{L}(A) = \texttt{in}$ and $\mathcal{L}(B) = \texttt{out}$, so that "$\texttt{in}(A) \mathrel{\Box\!\!\rightarrow} \texttt{out}(B)$" holds in AF_1. On the other hand, Proposition 2 does not hold in general for labelling that is not universally defined.

By "if $\varphi \wedge \neg\psi$ is true, then $\neg\,(\varphi \mathrel{\Box\!\!\rightarrow} \psi)$ is true", *modus ponens* "if φ and $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ are true, then ψ is true" is valid in Lewis's logic. In AF, the next results hold by Proposition 2(2) and (4).

Proposition 3 *Let $AF = (Ar, att)$, $A, B \in Ar$ and $\mathcal{L}^{(')}$ a universally defined labelling.*

1. *If $\mathcal{L}(A) = \ell_1$ in every labelling $\mathcal{L}$ of AF and $\ell_1(A) \mathrel{\Box\!\!\rightarrow} \ell_2(B)$ is true in AF, then $\mathcal{L}'(B) = \ell_2$ in every labelling $\mathcal{L}'$ of AF.*
2. *If $\mathcal{L}(A) = \ell_1$ in every labelling $\mathcal{L}$ of AF and $\ell_1(A) \mathrel{\Diamond\!\!\rightarrow} \ell_2(B)$ is true in AF, then $\mathcal{L}'(B) = \ell_2$ in some labelling $\mathcal{L}'$ of AF.*

In Lewis's logic, the relation "$(\varphi \mathrel{\Box\!\!\rightarrow} \psi) \equiv \neg\,(\varphi \mathrel{\Diamond\!\!\rightarrow} \neg\psi)$" and "$(\varphi \mathrel{\Diamond\!\!\rightarrow} \psi) \equiv \neg\,(\varphi \mathrel{\Box\!\!\rightarrow} \neg\psi)$" hold. The following relations hold for counterfactuals in AF. In what follows, given $\ell \in \{\texttt{in}, \texttt{out}\}$, $\overline{\ell}$ is $\texttt{out}$ (resp. $\texttt{in}$) if ℓ is $\texttt{in}$ (resp. $\texttt{out}$).

Proposition 4 *Let $AF = (Ar, att)$ and $A, B \in Ar$.*

1. *If $\ell_1(A) \mathrel{\Box\!\!\rightarrow} \ell_2(B)$ is true in AF, then $\ell_1(A) \mathrel{\Diamond\!\!\not\rightarrow} \overline{\ell_2}(B)$ is true in AF.*
2. *If $\ell_1(A) \mathrel{\Diamond\!\!\rightarrow} \ell_2(B)$ is true in AF, then $\ell_1(A) \mathrel{\Box\!\!\not\rightarrow} \overline{\ell_2}(B)$ is true in AF.*

The converse relations also hold under stable labelling.

Proof: The if-parts are straightforward by definition. The converse also holds under stable labelling because if it is not the case that $\overline{\ell_2}(B)$ in some (resp. every) stable labelling of AF^c then $\ell_2(B)$ in every (resp. some) stable labelling of AF^c. □

The fact that the converse relations do not generally hold in Proposition 4 is due to the existence of arguments with the labelling $\texttt{undec}$. Lewis's logic does not satisfy the law of *conditional excluded middle*: $(\varphi \mathrel{\Box\!\!\rightarrow} \psi) \vee (\varphi \mathrel{\Box\!\!\rightarrow} \neg\psi)$, which is satisfied by Stalnaker's logic [27]. In AF it may happen that both "$\texttt{in}(A) \mathrel{\Box\!\!\not\rightarrow} \texttt{in}(B)$" and "$\texttt{in}(A) \mathrel{\Box\!\!\not\rightarrow} \texttt{out}(B)$". In Lewis's counterfactuals, if $(\varphi \mathrel{\Box\!\!\rightarrow} \psi)$ and $(\varphi \mathrel{\Box\!\!\rightarrow} \neg\psi)$ are both false, then $(\varphi \mathrel{\Diamond\!\!\rightarrow} \psi)$ and $(\varphi \mathrel{\Diamond\!\!\rightarrow} \neg\psi)$ are both true. In AF, on the other hand, the truth of both "$\texttt{in}(A) \mathrel{\Box\!\!\not\rightarrow} \texttt{in}(B)$" and "$\texttt{in}(A) \mathrel{\Box\!\!\not\rightarrow} \texttt{out}(B)$" implies the truth of both "$\texttt{in}(A) \mathrel{\Diamond\!\!\rightarrow} \texttt{in}(B)$" and "$\texttt{in}(A) \mathrel{\Diamond\!\!\rightarrow} \texttt{out}(B)$" under stable labelling. However, this implication does not hold in general because it may be that "$\texttt{in}(A) \mathrel{\Box\!\!\rightarrow} \texttt{undec}(B)$" holds. Such disagreement will be resolved if we allow arguments with the $\texttt{undec}$ labelling in counterfactual conditionals, but we do not pursue the issue further in this paper.

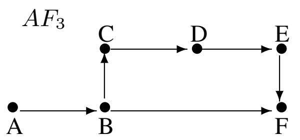
Lewis [18] argues three cases of *counterfactual fallacies* which distinguish counterfactual conditionals from the material conditional. The *fallacy of strengthening the antecedent* is the *invalid* inference pattern from $(\varphi \rightarrow \psi)$ to $(\varphi \wedge \chi \mathrel{\Box\!\!\rightarrow} \psi)$. The *fallacy of transitivity* is the *invalid* inference pattern from $(\chi \mathrel{\Box\!\!\rightarrow} \varphi)$ and $(\varphi \mathrel{\Box\!\!\rightarrow} \psi)$ to $(\chi \mathrel{\Box\!\!\rightarrow} \psi)$. The *fallacy of contraposition* is the *invalid* inference pattern from $(\varphi \mathrel{\Box\!\!\rightarrow} \psi)$ to $(\neg\psi \mathrel{\Box\!\!\rightarrow} \neg\varphi)$. These features are also the case in counterfactuals in AF.

Proposition 5 *Let $AF = (Ar, att)$ and $A, B, C \in Ar$.*

- *$\ell_1(A) \Box\!\!\rightarrow \ell_2(B)$ in AF does not imply $(\ell_1(A) \wedge \ell_3(C)) \Box\!\!\rightarrow \ell_2(B)$ in AF.*[3]
- *$(\ell_1(A) \Box\!\!\rightarrow \ell_2(B)) \wedge (\ell_2(B) \Box\!\!\rightarrow \ell_3(C))$ in AF does not imply $\ell_1(A) \Box\!\!\rightarrow \ell_3(C)$ in AF.*
- *$\ell_1(A) \Box\!\!\rightarrow \ell_2(B)$ in AF does not imply $\overline{\ell_2}(B) \Box\!\!\rightarrow \overline{\ell_1}(A)$ in AF.*

The above results also hold by replacing $\Box\!\!\rightarrow$ with $\Diamond\!\!\rightarrow$.

Example 2 Consider AF_3 illustrated in the right. It holds that



- "$\mathtt{in}(D) \Box\!\!\rightarrow \mathtt{in}(F)$" holds in AF_3 but "$(\mathtt{in}(B) \wedge \mathtt{in}(D)) \Box\!\!\not\rightarrow \mathtt{in}(F)$" in AF_3.
- "$(\mathtt{in}(B) \Box\!\!\rightarrow \mathtt{in}(D)) \wedge (\mathtt{in}(D) \Box\!\!\rightarrow \mathtt{in}(F))$" holds in AF_3 but "$\mathtt{in}(B) \Box\!\!\not\rightarrow \mathtt{in}(F)$" in AF_3.
- "$\mathtt{in}(E) \Box\!\!\rightarrow \mathtt{out}(F)$" holds in AF_3 but "$\mathtt{in}(F) \Box\!\!\not\rightarrow \mathtt{out}(E)$" in AF_3.

Counterfactuals are *nonmonotonic*, i.e., it may happen that $\varphi \Box\!\!\rightarrow \psi$ and $(\varphi \wedge \chi) \Box\!\!\rightarrow \neg\psi$. This is also the case in counterfactuals in AF. Nonmonotonicity implies the fallacy of strengthening the antecedent. Note that although contraposition of counterfactuals is invalid in general, inference by *modus tollens* on a counterfactual is valid, that is, $\neg\varphi$ is inferred from $(\varphi \Box\!\!\rightarrow \psi)$ and $\neg\psi$. Modus tollens also holds in counterfactuals in AF.

Proposition 6 *Let $AF = (Ar, att)$ and $A, B \in Ar$. If $\ell_1(A) \Box\!\!\rightarrow \ell_2(B)$ is true in AF and $\mathcal{L}(B) \neq \ell_2$ for any labelling $\mathcal{L}$ of AF, then $\mathcal{L}(A) \neq \ell_1$ for any labelling $\mathcal{L}$ of AF.*

Proof: Suppose $\mathtt{in}(A) \Box\!\!\rightarrow \mathtt{in}(B)$ holds in AF and $\mathcal{L}(B) \neq \mathtt{in}$ for any labelling $\mathcal{L}$ of AF. If $\mathcal{L}'(A) = \mathtt{in}$ for some labelling $\mathcal{L}'$ of AF, then the fact that B has the labelling $\mathtt{in}$ after removing any attack relation (X, A) in AF^c_{+A} implies that $\mathcal{L}'(B) = \mathtt{in}$ in the labelling $\mathcal{L}'$ of AF such that $\mathcal{L}'(A) = \mathtt{in}$. This contradicts the assumption that $\mathcal{L}(B) \neq \mathtt{in}$ for any labelling $\mathcal{L}$ of AF. Other cases are shown in similar ways. □

Clearly, modus tollens does not hold for the counterfactual conditional $\Diamond\!\!\rightarrow$. If $\ell_1(A) \Diamond\!\!\rightarrow \ell_2(B)$ is true in AF and $\mathcal{L}(B) \neq \ell_2$ for *some* labelling $\mathcal{L}$ of AF, then it does not necessarily hold that $\mathcal{L}(A) \neq \ell_1$ for any labelling $\mathcal{L}$ of AF.

Proposition 7 *Let $AF = (Ar, att)$ and $A, B, C \in Ar$. If $(\ell_1(A) \Box\!\!\rightarrow \ell_2(B)) \wedge (\ell_2(B) \Box\!\!\rightarrow \ell_1(A))$ is true in AF, then $(\ell_1(A) \Box\!\!\rightarrow \ell_3(C)) \supset (\ell_2(B) \Box\!\!\rightarrow \ell_3(C))$ is true in AF.*

Proof: We show the case of $\ell_1 = \ell_2 = \mathtt{in}$. Suppose that both $\mathtt{in}(A) \Box\!\!\rightarrow \mathtt{in}(B)$ and $\mathtt{in}(B) \Box\!\!\rightarrow \mathtt{in}(A)$ hold in AF. Then, A is labelled $\mathtt{in}$ iff B is labelled $\mathtt{in}$ in both AF^c_{+A} and AF^c_{+B}. Then, if $\mathcal{L}(C) = \ell_3$ in every labelling of AF^c_{+A} , it is also the case in AF^c_{+B}. Other cases are shown in similar ways. □

Proposition 7 does not hold for $\Diamond\!\!\rightarrow$. Two counterfactual conditionals are combined.

[3] We do not provide a formal definition of a counterfactual having conjunction in its antecedent, but the intended meaning is obvious. For instance, "$(\mathtt{in}(A) \wedge \mathtt{in}(B)) \Box\!\!\rightarrow \mathtt{in}(C)$" is true in AF if $\mathcal{L}(C) = \mathtt{in}$ in every labelling $\mathcal{L}$ of the argumentation framework which is obtained from AF by removing every attack relation attacking A or B.

Proposition 8 *Let $AF = (Ar, att)$ and $A, B, C \in Ar$. If $(\ell_1(A) \Box\!\!\rightarrow \ell_2(B)) \wedge (\ell_1(A) \Box\!\!\rightarrow \ell_3(C))$ is true in AF, then $\ell_1(A) \Box\!\!\rightarrow (\ell_2(B) \wedge \ell_3(C))$ is true in AF.*[4]

Proof: If B is labelled ℓ_2 and C is labelled ℓ_3 in every labelling of AF^c_{+A}, then $\ell_2(B) \wedge \ell_3(C)$ is true in every labelling of AF^c_{+A}. Hence, $(\mathtt{in}(A) \Box\!\!\rightarrow \ell_2(B)) \wedge (\mathtt{in}(A) \Box\!\!\rightarrow \ell_3(C))$ imply $\mathtt{in}(A) \Box\!\!\rightarrow (\ell_2(B) \wedge \ell_3(C))$. Similarly, it is shown that $(\mathtt{out}(A) \Box\!\!\rightarrow \ell_2(B)) \wedge (\mathtt{out}(A) \Box\!\!\rightarrow \ell_3(C))$ imply $\mathtt{out}(A) \Box\!\!\rightarrow (\ell_2(B) \wedge \ell_3(C))$. □

Such combination property does not hold for $\Diamond\!\!\rightarrow$. In Example 1, both "$\mathtt{in}(B) \Diamond\!\!\rightarrow \mathtt{in}(D)$" and "$\mathtt{in}(B) \Diamond\!\!\rightarrow \mathtt{in}(E)$" hold in AF_2, but "$\mathtt{in}(B) \Diamond\!\!\not\rightarrow (\mathtt{in}(D) \wedge \mathtt{in}(E))$".

4. Counterfactual Dependencies

A counterfactual sentence may be true even if there is no causal dependency between the antecedent and the consequent. In fact, "if $\varphi \wedge \psi$ is true, then $\varphi \Box\!\!\rightarrow \psi$ is true", so that the counterfactual conditionals do not require any causal relation between φ and ψ. Lewis argues that "we do know that causation has something or other to do with counterfactuals" [19] and defines counterfactual dependencies between sentences. Formally, an event φ *depends causally* on another event ψ iff both $(\psi \Box\!\!\rightarrow \varphi)$ and $(\neg\psi \Box\!\!\rightarrow \neg\varphi)$ hold. This definition is captured in AF as follows.

Definition 9 (counterfactual dependencies) Let $AF = (Ar, att)$ and $A, B \in Ar$. Then, $\ell_1(A) \overset{c}{\Box\!\!\rightarrow} \ell_2(B)$ is *true* in AF if both $\ell_1(A) \Box\!\!\rightarrow \ell_2(B)$ and $\overline{\ell_1}(A) \Box\!\!\rightarrow \overline{\ell_2}(B)$ hold in AF where $\ell_1, \ell_2 \in \{\mathtt{in}, \mathtt{out}\}$. In this case, we say that $\ell_2(B)$ *causally depends on* $\ell_1(A)$. The relation "$\ell_1(A) \overset{c}{\Box\!\!\rightarrow} \ell_2(B)$" is called a *counterfactual dependency*.

By definition, $\ell_1(A) \overset{c}{\Box\!\!\rightarrow} \ell_2(B)$ implies $\ell_1(A) \Box\!\!\rightarrow \ell_2(B)$, but not vice versa. In Example 1, "$\mathtt{out}(A) \Box\!\!\rightarrow \mathtt{in}(F)$" holds in AF_2, but "$\mathtt{out}(A) \overset{c}{\Box\!\!\not\rightarrow} \mathtt{in}(F)$", for example. Unlike $\Box\!\!\rightarrow$, the result of Proposition 2(1) does not hold for $\overset{c}{\Box\!\!\rightarrow}$ in general. Like $\Box\!\!\rightarrow$, the relation $\overset{c}{\Box\!\!\rightarrow}$ is reflexive but not transitive. Strengthening the antecedent or contraposition is invalid, while modus ponens and modus tollens are valid. Propositions 7 and 8 also hold for $\overset{c}{\Box\!\!\rightarrow}$. In AF an argument causally depends on another argument if there is a directed path between those arguments. Formally, the following relations hold.

Proposition 9 *Let $AF = (Ar, att)$, $A, B \in Ar$ and $\ell \in \{\mathtt{in}, \mathtt{out}\}$.*

- *If $\ell(A) \overset{c}{\Box\!\!\rightarrow} \ell(B)$ is true in AF, then A indirectly defends B.*
- *If $\ell(A) \overset{c}{\Box\!\!\rightarrow} \overline{\ell}(B)$ is true in AF, then A indirectly attacks B.*

Proof: Suppose that $\mathtt{in}(A) \overset{c}{\Box\!\!\rightarrow} \mathtt{in}(B)$ holds in AF. Then both $\mathtt{in}(A) \Box\!\!\rightarrow \mathtt{in}(B)$ and $\mathtt{out}(A) \Box\!\!\rightarrow \mathtt{out}(B)$ hold in AF. Then B is labelled $\mathtt{in}$ in every labelling of AF^c_{+A}, and B is labelled $\mathtt{out}$ in every labelling of AF^c_{-A}. Such a change of labelling happens only when there is a directed path from A to B. If A does not indirectly defends B, every path from A to B is an indirect attack relation. Since $\mathtt{in}(B)$ in AF^c_{+A}, for any argument $X \in Ar$ such that $(X, B) \in att$, there is an argument $Y \in Ar$ such that $(Y, X) \in att$ and $\mathtt{in}(Y)$ in AF^c_{+A}. Then Y indirectly defends B, but A does not indirectly defend

[4] We do not provide the definition of a counterfactual having conjunction in its consequent, but the intended meaning is obvious.

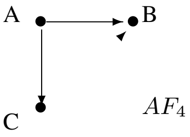
Y (i.e., every path from A to B is indirectly attacking). In this case, making A $\texttt{out}$ in AF^c_{-A} does not change the labelling of B and B is labelled $\texttt{in}$ in every labelling of AF^c_{-A}. Contradiction. The other cases are shown in similar ways. □

The converse of Proposition 9 does not hold in general. In Example 2, B indirectly defends F in AF_3, but $\texttt{in}(B) \overset{c}{\Box\!\!\not\rightarrow} \texttt{in}(F)$. Note that Lewis does not use the conditional operator $\Diamond\!\!\rightarrow$ for defining causal dependencies between events. We can also observe that the relation $\Diamond\!\!\rightarrow$ is inappropriate for defining causal dependencies between arguments. For instance, consider $AF = (\{A, B, C\}, \{(B, C), (C, B)\})$. Then, both $\texttt{in}(A) \Diamond\!\!\rightarrow \texttt{in}(B)$ and $\texttt{out}(A) \Diamond\!\!\rightarrow \texttt{out}(B)$ hold, while there is no connection between A and B.

Lewis distinguishes "causal dependencies" from "causation". He says:

> Causal dependence among actual events implies causation. If c and e are two actual events such that e would not have occurred without c, then c is a cause of e. But I reject the converse. Causation must always be transitive; causal dependence may not be; so there can be causation without causal dependence. [19, p. 563][5]

In the context of AF, indirect attack/defend relations are transitively combined and considered a kind of "causation". By contrast, counterfactual dependencies are not transitive (cf. Example 2). Proposition 9 shows that counterfactual dependencies imply indirect attack/defend relations between arguments, which conforms to Lewis's view that causal dependencies imply causation. Different causes may interact. Suppose AF_4 illustrated in the right. Then, "$\texttt{in}(A) \Box\!\!\rightarrow \texttt{out}(B)$" but "$\texttt{out}(A) \Box\!\!\not\rightarrow \texttt{in}(B)$" imply "$\texttt{in}(A) \overset{c}{\Box\!\!\not\rightarrow} \texttt{out}(B)$". Likewise, "$\texttt{in}(C) \Box\!\!\rightarrow \texttt{out}(B)$" but "$\texttt{out}(C) \Box\!\!\not\rightarrow \texttt{in}(B)$" imply "$\texttt{in}(C) \overset{c}{\Box\!\!\not\rightarrow} \texttt{out}(B)$". Thus, $\texttt{out}(B)$ causally depends on neither $\texttt{in}(A)$ nor $\texttt{in}(C)$. In this AF_4, the argument A attacks B and, at the same time, indirectly defends B via C. The argument B is rejected because the argument A is accepted in AF_4. Thus, the existence of A is the actual cause of rejecting B. On the other hand, if A were rejected then C would be accepted, which results in rejecting B. In this sense, the argument C is considered a "potential" alternative cause of rejecting B. Lewis distinguishes the actual cause from potential causes using the term "preemption" as follows:



> Suppose that c_1 occurs and causes e; and that c_2 also occurs and does not cause e, but would have caused e if c_1 had been absent. Thus c_2 is a potential alternative cause of e, but is preempted by the actual cause c_1. [19, p. 567]

In AF_4, "$\texttt{in}(C) \Box\!\!\rightarrow \texttt{out}(B)$" represents that $\texttt{in}(C)$ is a potential cause of $\texttt{out}(B)$ but is *preempted* by the actual cause $\texttt{in}(A)$.

5. Discussion

Lewis [18] argues modal interpretation of sentences in terms of counterfactuals. Let $\bot$ be a sentential constant *false* at every possible world. Then, $\Box\varphi \overset{def}{=} (\neg\varphi \Box\!\!\rightarrow \bot)$ and $\Diamond\varphi \overset{def}{=} \neg\Box\neg\varphi$. We can construct a similar interpretation of arguments using counterfactuals. Let $\bot$ be an inconsistent argument which is always false. Given an argument A, define

[5] There are arguments against the transitivity of causation [16].

Table 1.

	Axioms	Propositions in this paper
(A1)	all truth-functional tautologies	
(A2)	$(p > q) \wedge (p > r) \supset (p \supset q \wedge r)$	Prop. 8
(A3)	$p > \top$	universality of semantics
(A4)	$p > p$	Prop. 1
(A5)	$((p > q) \wedge (q > p)) \supset ((p > r) \supset (q > r))$	Prop. 7
(A6)	$p \wedge q \supset (p > q)$	Prop. 2
(A7)	$(p > q) \supset (p \supset q)$	(not addressed)
(A8)	$(p > r) \wedge (q > r) \supset (p \vee q > r)$	(not addressed)
(A9)	$(p > q) \wedge \neg(p > \neg r) \supset (p \wedge r > q)$	(not addressed)

$$\Box\ell(A) \overset{def}{=} \bar{\ell}(A) \mathrel{\Box\!\!\rightarrow} \mathtt{in}(\bot) \quad \text{and} \quad \Diamond\ell(A) \overset{def}{=} \ell(A) \mathrel{\Box\!\!\not\rightarrow} \mathtt{in}(\bot).$$

For instance, "$\Box\mathtt{in}(A) = \mathtt{out}(A) \mathrel{\Box\!\!\rightarrow} \mathtt{in}(\bot)$" (an argument A is necessarily accepted iff inconsistency would be accepted if A were rejected). We use such argumentative reasoning in daily life. For example, put the argument A as "$\sqrt{2}$ is an irrational number". Then, the validity of the argument is proven by showing inconsistency under the assumption that $\sqrt{2}$ were a rational number. A modal interpretation of arguments provided above introduces yet another connection between argumentation and modal logic, which is different from existing studies such as [3,9,15].

Lewis [18] introduces an axiomatic system of counterfactuals, which is reformulated by Gärdenfors [13] as (A1)–(A9) in Table 1 where $>$ means a conditional operator. Gärdenfors shows that the above nine axioms together with two inference rules, (R1) Modus Ponens and (R2) "if $q \supset r$ is a theorem then $(p > q) \supset (p > r)$ is a theorem", provide the same logic of conditionals as Lewis's counterfactuals. Some correspondences between the axioms and the propositions presented in this paper are presented in Table 1. In the table, if we interpret the conditional "$\mathtt{in}(A) \mathrel{\Box\!\!\rightarrow} \top$" or "$\mathtt{out}(A) \mathrel{\Box\!\!\rightarrow} \top$" as the consistency (i.e., existence of an extension) of AF^c_{+A} or AF^c_{-A}, the axiom (A3) also holds in argumentation semantics which is universally defined. Due to space limitations, we do not address properties corresponding to (A7)–(A9) in this paper. It would be an interesting research topic to formulate counterfactuals using an instantiated AF in propositional logic and verify those axioms.

Computational complexity of counterfactual reasoning in AF is derived by Definition 7. Given an argumentation framework AF, the problem of deciding whether a counterfactual conditional $\mathtt{in}(A) \mathrel{\Box\!\!\rightarrow} \ell(B)$ (resp. $\mathtt{in}(A) \mathrel{\Diamond\!\!\rightarrow} \ell(B)$) holds or not in AF is equivalent to deciding whether $\mathcal{L}(B) = \ell$ in every (resp. some) labelling $\mathcal{L}$ of AF^c_{+A}. Likewise, the problem of deciding whether a counterfactual conditional $\mathtt{out}(A) \mathrel{\Box\!\!\rightarrow} \ell(B)$ (resp. $\mathtt{out}(A) \mathrel{\Diamond\!\!\rightarrow} \ell(B)$) holds or not in AF is equivalent to deciding whether $\mathcal{L}(B) = \ell$ in every (resp. some) labelling $\mathcal{L}$ of AF^c_{-A}. Therefore, complexities of counterfactual reasoning under the operator $\Box\!\!\rightarrow$ (resp. $\Diamond\!\!\rightarrow$) are equivalent to those of skeptical (resp. credulous) reasoning of an argument under argumentation semantics. Then we can apply the complexity results of argumentation semantics reported in [12].

As stated in the introduction, counterfactual reasoning is widely used in dialogue or dispute, so the proposed framework has potential application to realizing counterfactual reasoning in dialogue systems based on formal argumentation. Counterfactuals are also used in diagnosis in which assumptions are introduced for explaining the observed misbehavior of a device [14]. For instance, we experimentally know that if a car did not start then the car battery might be dead. The situation is represented by $\mathtt{in}(A) \Diamond\!\!\rightarrow \mathtt{in}(B)$ where A is "A car does not start" and B is "The car battery is dead". One morning I found that my car does not start. Then, by $\mathtt{in}(A)$ and the above counterfactual sentence, I conclude by modus ponens that $\mathtt{in}(B)$ might be the case. Thus, counterfactuals could be used for analytic tools in AF. Counterfactual reasoning would have connection to *dishonest reasoning*. When one wants to have a desired outcome which would be achieved not by telling true belief but by false belief, one has an incentive to *lie*. In this case, one would reason counterfactually. For instance, suppose a child, Susie, who watches TV in the living room. Mom asks whether she did her homework. Susie considers that Mom would permit her watching TV if she finished her homework while she did not finish it. The situation is represented by the counterfactual conditional $\mathtt{in}(A) \Box\!\!\rightarrow \mathtt{in}(B)$ with $A =$ "Susie finishes her homework" and $B =$ "Mon permits watching TV". To have the desired result B, Susie lies to her Mom: "Yes, I did my homework". Dishonest reasoning is used in a *debate game* which provides an abstract model of debates between two players [26]. In debate games, two players have their own argumentation frameworks and each player builds claims to refute the opponent. A player may provide false or inaccurate arguments as a tactic to win the game. Counterfactual reasoning would be used for building false arguments in a game.

We finally remark some related works. Booth $et\,al.$ [6] introduce *conditional acceptance functions* that account for dynamic aspects of argument evaluation. Using the function, one can reason counterfactually that "what if an argument A was not accepted?". They show that such counterfactual reasoning is useful to distinguish argumentation frameworks which have different topological structures but have the same extensions [7]. They characterize counterfactuals as nonmonotonic inference in a manner different from ours. Counterfactual reasoning is closely related to theory change and belief revision [13,22]. Rotstein $et\,al.$ [25] study argumentation theory change in abstract argumentation framework. They introduce argument change operators which expand/contact the set of arguments to warrant a particular argument. Baumann and Brewka [1] consider the problem of modifying AF in a way that a desired set of arguments becomes an extension. Compared with these studies, our interest is not in the change of AF to warrant or enforce desired arguments, but we focus on the effect of counterfactual change of a particular argument. Boella $et\,al.$ [4,5] consider the effect of adding/removing arguments or attack relations under the grounded semantics. Their primary interest is on the invariance of the argumentation semantics when arguments or attack relations have been added/removed. This is in contrast with counterfactuals in AF which consider possible changes by introducing arguments and removing attack relations. Cayrol $et\,al.$ [10] study the effect of an addition of an argument on the outcome of the grounded/preferred semantics. They analyze how extensions of an AF change by adding a new argument that may interact with existing arguments. The study focuses on dynamics of extensions in AF, rather than causal changes between particular arguments. Bochman [2] studies a connection between argumentation and causal reasoning. He argues that causal reasoning is viewed as a kind of argumentation by interpreting that "A produces B iff $\neg B$ attacks A." This is different from our formulation of causal dependencies in AF. He does not argue how counterfactuals are characterized in his logic. Pearl [23] introduces a *causal model* which encodes propositional sentences that deal with causal relationships. A causal model is associated with a directed graph called a *causal diagram* which is used for representing and reasoning with counterfactuals [23,16]. The vertices in causal diagrams are propositional variables whose truth values are determined by a set of functions defined over those variables. Arcs in causal diagrams represent input-output relations between variables in those functions. Like argumentation graphs, causal diagrams are used for structural analyses of counterfactuals, while two graphs encode the problem in different ways.

References

[1] R. Baumann and G. Brewka. Expanding argumentation frameworks: enforcing and monotonicity results. In: *Proc. 3rd COMMA, Frontiers in AI and Applications* 216, pp. 75–86, IOS Press, 2010.

[2] A. Bochman. Propositional argumentation and causal reasoning. In *Proc. IJCAI-05*, pp. 388–393, 2005.

[3] G. Boella, J. Hulstijn, and L. van der Torre. A logic of abstract argumentation. *Argumentation in Multi-Agent Systems, Lecture Notes in Artificial Intelligence* 4049, pp. 29–41, Springer, 2006.

[4] G. Boella, S. Kaci, and L. van der Torre. Dynamics in argumentation with single extensions: abstract principles and the grounded extension. In: *Proc. ECSQARU-09*, LNAI 5590, pp. 107–118, 2009.

[5] G. Boella, S. Kaci, and L. van der Torre. Dynamics in argumentation with single extensions: attack refinement and the grounded extension (extended version). *Argumentation in Multi-Agent Systems, Lecture Notes in Artificial Intelligence* 6057, pp. 150–159, Springer, 2010.

[6] R. Booth, S. Kaci, T. Rienstra, L. van der Torre. Conditional acceptance functions. In: *Proc. 4th COMMA, Frontiers in AI and Applications* 245, pp. 470–477, IOS Press, 2012.

[7] R. Booth, S. Kaci, T. Rienstra, L. van der Torre. Monotonic and nonmonotonic inference for abstract argumentation. In: *Proc. Int'l Florida AI Research Society Conference*, pp. 597–602, 2013.

[8] R. M. J. Byrne. *The Rational Imagination: How People Create Alternatives to Reality*. MIT Press, 2005.

[9] M. Caminada and D. Gabbay. A logical account of formal argumentation. *Studia Logica* 93:109–145, 2009.

[10] C. Cayrol, F. Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in abstract argumentation frameworks: adding an argument. *J. Artificial Intelligence Research* 38:49–84, 2010.

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

[12] P. E. Dunne and M. Wooldridge. Complexity of abstract argumentation. In: I. Rahwan and G. R.. Simari (eds.), *Argumentation in Artificial Intelligence*, pp. 85–104, Springer, 2009.

[13] P. Gärdenfors. Conditionals and changes of belief. *Acta Philosophica Fennica* 30:381–404, 1978.

[14] M. Ginsberg. Counterfactuals. *Artificial Intelligence* 30:35–79, 1986.

[15] D. Grossi. Argumentation in the view of modal logic. *Argumentation in Multi-Agent Systems, Lecture Notes in Artificial Intelligence* 6614, pp. 190–208, Springer, 2011.

[16] C. Hitchcock. The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy* 98(6):273–299, 2001.

[17] G. King and L. Zeng. The dangers of extreme counterfactuals. *Political Analysis* 14(2): 131–159, 2005.

[18] D. Lewis. *Counterfactuals*. Blackwell Publishing, 1973.

[19] D. Lewis. Causation. *Journal of Philosophy* 70:556–567, 1973.

[20] S. L. Morgan and C. Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, 2007.@

[21] N. Nivelle. Counterfactual conditionals in argumentative legal language in Dutch. *Pragmatics* 18(3):469–490, 2008.

[22] D. Nute and C. B. Cross. Conditional logic. In: D. M. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, 2nd edition*, vol. 4, pp. 1–98, Kluwer Academic, 2002.

[23] J. Pearl. *Causality: models, reasoning, and inference*. Cambridge University Press, 2000.

[24] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *J. Logic and Computation* 15(6):1009–1040, 2005.

[25] N. D. Rotstein, M. O. Moguillansky, M. A. Falappa, A. J. García, and G. R. Simari. Argument theory change: revision upon warrant. In: *Proc. 2nd Int'l Conf. Computational Models of Argument, Frontiers in AI and Applications* 172, pp. 336–347, IOS Press, 2008.

[26] C. Sakama. Dishonest arguments in debate games. In: *Proc. 4th Int'l Conf. Computational Models of Argument, Frontiers in AI and Applications* 245, pp. 177–184, IOS Press, 2012.

[27] R. Stalnaker. A theory of conditionals. In: N. Rescher (ed.), *Studies in Logical Theory*, Blackwell, 1968.