Rethinking the Rationality Postulates for Argumentation-Based Inference

Henry PRAKKEN
Department of Information and Computing Sciences, Utrecht University and Faculty of Law, University of Groningen, The Netherlands

Abstract. Much research on structured argumentation aims to satisfy the rationality postulates of direct and indirect consistency and strict (deductive) closure. However, examples like the lottery paradox indicate that it is sometimes rational to accept sets of propositions that are indirectly inconsistent or not deductively closed. This paper proposes a variant of the *ASPIC*$^+$ framework that violates indirect consistency and full strict closure but satisfies direct consistency and restricted forms of strict closure and indirect consistency.

Keywords. Rational acceptance, Rationality postulates, Lottery paradox

1. Introduction

Much current work on structured argumentation (e.g. [6,9,2]) concerns the so-called rationality postulates of [1]. The idea is that argument extensions [4] should be closed under subarguments and that the sets of conclusions of all arguments in an extension should be directly consistent (no formulas that negate each other should be in the set), closed under strict (deductive) inference and indirectly consistent (the strict closure should be directly consistent). Most work on these postulates simply assumes that they should be satisfied, but examples like the lottery paradox [7] suggest that it may sometimes be rational to jointly accept indirectly inconsistent propositions or not to accept deductive consequences of acceptable propositions.

Imagine a fair lottery with one million tickets and just one prize. If the principle is accepted that it is rational to accept a proposition if its truth is highly probable, then for each ticket T_i it is rational to accept that T_i will not win while at the same time it is rational to accept that exactly one ticket will win. If we also accept that everything that deductively follows from a set of rationally acceptable propositions, then we have two rationally acceptable propositions that contradict each other: we can join all individual propositions $\neg T_i$ into a big conjunction $\neg T_1 \wedge \ldots \wedge \neg T_{1,000,000}$ with one million conjuncts, which contradicts the certain fact that exactly one ticket will win.

The problem does not only arise in precisely defined probabilistic settings (cf. [11]). First, non-statistical examples of the lottery paradox can easily be imagined. For example, for each arbitrary part of a complex machine we can rationally accept that it will not malfunction but at the same time we know that some part will at some point in time malfunction. Moreover, the problem arises in any model of 'fallible' rational acceptance. Rational acceptance is usually fallible, either because one starts from uncertain premises or

because one applies defeasible inferences. Now whenever a deductive inference is made from at least two 'fallible' pieces of information, the deductive inference can be said to aggregate the degrees of fallibility of the individual elements to which it is applied. This in turn means that the deductive inference may be weaker than either of these elements, so that a successful attack on the deductive inference does not necessarily imply a successful attack on one of the fallible elements to which it was applied.

In discussions of the lottery paradox several positions have been defended. for example, Pollock [10] argued that sets of rationally acceptable propositions should always be deductively closed. Moreover, in the lottery paradox he argued that for no ticket is it rational to accept that it will not win. However, this position is not quite self-evident: if propositions cannot be accepted even if their truth is highly probable, then many propositions that seem clearly acceptable would not formally come out as such. Others (including Kyburg [7]) reject the conjunction principle for rational acceptance, motivated by the fact that according to probability theory a conjunction of two highly probable propositions need not be highly probable. However, this also has its issues, since people often conjoin their beliefs, and regarding this as always irrational seems too strong. Therefore intermediate positions have also been considered. For example, Makinson [8] argues that (in the lottery example) conjunctions $\neg T_i \wedge \ldots \neg T_j$ are rationally acceptable for up to a particular (not too large) number of conjuncts. And [3] argue that examples like the lottery paradox are exceptional cases where strict closure fails since their underlying probability structure is uniform: no particular event is typical and randomness prevails. In this paper we want to explore whether such an intermediate position can be formalised in an argumentation setting. In doing so, we will make two assumptions.

First, problems like these do not arise when rational acceptance is seen as a matter of degree. In epistemology there is a debate whether rational acceptance is always a matter of degree or whether it makes sense to speak of full (though still possibly defeasible) acceptance [5]. Taking a stance in this debate goes beyond the scope of this paper but since the notion of full acceptance is in epistemology often defended, it makes sense to explore its consequences in an argumentation setting. This holds the more since most formal and computational models of argument model non-gradual notions of full acceptance.

Second, Pollock [10] also argued that what can be rationally accepted in the lottery paradox is that it is *highly probable* that it will win. At first sight, this approach would seem attractive, until one realises that if it is applied to the lottery example, it should be applied to many other examples of defeasible reasoning, since many of those arguably have an underlying probabilistic justification. So why require in the lottery example that the probability of a statement is expressed in the object language while not requiring this for, for instance, 'If P then usually Q' and 'P' defeasibly imply 'Q'? Accordingly, in this paper we will make a second assumption that is often adopted in formal and computational models of argument, namely, that the probability of statements is not expressed in the logical object language of a system but in its metalanguage, in the nonmonotonicity of its consequence notion. Just as the assumption that full acceptance is possible, this assumption is debatable, but both assumptions are widely adopted, which justifies this paper's aim to explore their logical consequences.

Summarising, the purpose of this paper is to formally investigate the relevance of examples like the lottery paradox for models of argumentation that model non-gradual notions of full acceptance and that express the probability of statements in the metalanguage in the nonmonotonicity of their consequence notion. In particular, we will explore

how the intermediate position can be formalised that conclusions of deductive inferences from fallibly acceptable propositions can but need not be rationally acceptable. We will argue that under the adopted assumptions the rationality postulate of direct consistency should be retained but that the postulates of indirect consistency and strict closure have to be weakened in general (although they may apply in special cases). We will carry out the investigations in terms of the *ASPIC*$^+$ framework, motivated by its generality: as shown earlier [12,9] it can be instantiated in many different ways and some of these ways capture other models of structured argumentation as special cases.

2. The *ASPIC*$^+$ framework

ASPIC$^+$ generates abstract argumentation frameworks in the sense of [4]. Formally, an **abstract argumentation framework** (AF) is a pair $(\mathcal{A}, \mathcal{D})$, where $\mathcal{A}$ is a set of *arguments* and $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation of *defeat*. We say that A *strictly defeats* B if A defeats B while B does not defeat A. A semantics for AFs returns sets of arguments called *extensions*, which are subsets of $\mathcal{A}$ with particular properties:

Definition 1 Let $(\mathcal{A}, \mathcal{D})$ be an AF. For any $X \in \mathcal{A}$, X is *acceptable* w.r.t. some $S \subseteq \mathcal{A}$ iff $\forall Y$ s.t. $(Y, X) \in \mathcal{D}$ implies $\exists Z \in S$ s.t. $(Z, Y) \in \mathcal{D}$. Let $S \subseteq \mathcal{A}$ be *conflict free*, i.e., there are no A, B in S such that $(A, B) \in \mathcal{D}$. Then S is: an *admissible* set iff $X \in S$ implies X is acceptable w.r.t. S; a *complete* extension iff $X \in S$ whenever X is acceptable w.r.t. S; a *preferred* extension iff it is a set inclusion maximal admissible set; the *grounded* extension iff it is the set inclusion minimal complete extension; a *stable* extension iff it is conflict-free and $\forall Y \notin S$, $\exists X \in S$ s.t. $(X, Y) \in \mathcal{D}$.
For $T \in \{$complete, preferred, grounded, stable$\}$, X is *sceptically* or *credulously* justified under the T semantics if X belongs to all, respectively at least one, T extension.

We next summarise *ASPIC*$^+$ as defined in [9]. It defines the notion of an abstract *argumentation system* as a structure consisting of a logical language $\mathcal{L}$ with negation, two sets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules, and a naming convention n in $\mathcal{L}$ for defeasible rules in order to talk about the applicability of defeasible rules in $\mathcal{L}$.

Definition 2 [**Argumentation systems**] An *argumentation system* is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a logical language with a unary negation connective $\neg$.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over wff in $\mathcal{L}$), such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- n is a partial function from $\mathcal{R}_d$ to $\mathcal{L}$.

We write $\psi = -\varphi$ just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$. Note that $-$ is not a connective in $\mathcal{L}$ but a function symbol in the metalanguage of $\mathcal{L}$.

ASPIC$^+$ leaves the choice of inference rules free. If desired, the strict rules can be based on a given deductive logic L by letting $\varphi_1, \ldots, \varphi_n \rightarrow \varphi \in \mathcal{R}_s$ iff $\varphi_1, \ldots, \varphi_n \vdash_L \varphi$. However, for simplicity this paper's examples will not encode full logics in $\mathcal{R}_s$.

Example 1 An example argumentation system is with $\mathcal{L} = \{p, \neg p, q, \neg q, r, \neg r, s, \neg s, t, \neg t, r_1, r_2, \neg r_1, \neg r_2\}$, $\mathcal{R}_s = \{p, r \rightarrow s; \neg s \rightarrow \neg r_1\}$, $\mathcal{R}_d = \{q \Rightarrow r; t \Rightarrow \neg s\}$ where $n(q \Rightarrow r) = r_1$ and $n(t \Rightarrow \neg s) = r_2$.

Definition 3 [**Consistency**] For any $S \subseteq \mathcal{L}$, let the *closure of S under strict rules*, denoted $Cl_{R_s}(S)$, be the smallest set containing S and the consequent of any strict rule in $\mathcal{R}_s$ whose antecedents are in $Cl_{R_s}(S)$. Then a set $S \subseteq \mathcal{L}$ is *directly consistent* iff $\nexists\ \psi, \varphi \in S$ such that $\psi = -\varphi$, and *indirectly consistent* iff $Cl_{R_s}(S)$ is directly consistent.

Example 2 In our example argumentation system, an example of a directly inconsistent set is $\{p, \neg p\}$ and an example of an indirectly inconsistent set is $\{p, r, \neg s\}$.

Definition 4 [**Knowledge bases**] A *knowledge base* in an $AS = (\mathcal{L}, \mathcal{R}, n)$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_n$ (the *axioms*) and $\mathcal{K}_p$ (the *ordinary premises*).

Arguments can be constructed from knowledge bases by applying inference rules. In what follows, for a given argument the function `Prem` returns all its premises, `Conc` returns its conclusion, `Sub` returns all its sub-arguments and `DefRules` and `TopRule` return, respectively, all defeasible rules and the last rule applied in the argument.

Definition 5 [**Arguments**] An *argument* A on the basis of a knowledge base $\mathcal{K}$ in an argumentation system $(\mathcal{L}, \mathcal{R}, n)$ is:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$; $\texttt{DefRules}(A)$ = $\emptyset$; $\texttt{TopRule}(A)$ = undefined.
2. $A_1, \ldots A_n \rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi \in \mathcal{R}_s$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$; $\texttt{Conc}(A) = \psi$; $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$; $\texttt{DefRules}(A)$ = $\texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$; $\texttt{TopRule}(A)$ = $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$.
3. $A_1, \ldots A_n \Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi \in \mathcal{R}_d$.
   $\texttt{Prem}(A)$, $\texttt{Conc}(A)$ and $\texttt{Sub}(A)$ are defined as in (2) while $\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi\}$ and $\texttt{TopRule}(A)$ = $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi$.

For any argument A, $\texttt{Prem}_\texttt{n}(A) = \texttt{Prem}(A) \cap \mathcal{K}_n$ and $\texttt{Prem}_\texttt{p}(A) = \texttt{Prem}(A) \cap \mathcal{K}_p$. An argument A is *infallible* if $\texttt{DefRules}(A) = \emptyset$ and $\texttt{Prem}(A) \subseteq \mathcal{K}_n$; otherwise it is *fallible*. For any set S of arguments, $\texttt{Conc}(S) = \{\varphi \mid \varphi = \texttt{Conc}(A) \text{ for some } A \in S\}$. We write $S \vdash \varphi$ if there exists a strict argument for φ with all premises taken from S.

Example 3 If our example argumentation system is combined with a knowledge base with $\mathcal{K}_n = \{p\}$ and $\mathcal{K}_p = \{q, t\}$, then the following arguments can be constructed, of which only A_1 is infallible:

$$
\begin{array}{llllll}
A_1 = & p & A_4 = & A_2 \Rightarrow r & A_7 = & A_5 \rightarrow \neg r_1 \\
A_2 = & q & A_5 = & A_3 \Rightarrow \neg s & & \\
A_3 = & t & A_6 = & A_1, A_4 \rightarrow s & &
\end{array}
$$

Arguments can be attacked in three ways: on an application of a defeasible rule, on the conclusion of such an application or on an ordinary premise.

Definition 6 [**Attack**] An argument A *attacks* an argument B iff A *undercuts* or *rebuts* or undermines B, where:

- A *undercuts* B (on B') iff $\texttt{Conc}(A) = -n(r)$ and $B' \in \texttt{Sub}(B)$ such that B''s top rule r is defeasible.
- A *rebuts* B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$.
- A *undermines* B (on φ) iff $\texttt{Conc}(A) = -\varphi$ for some $\varphi \in \texttt{Prem}(B) \cap \mathcal{K}_p$.

Example 4 In our running example A_6 rebuts A_5 and A_7 on A_5. Note that A_5 does not rebut A_6 since A_6 has a strict top rule. Furthermore, A_7 undercuts A_4 and A_6 on A_4.

Argumentation systems plus knowledge bases induce structured argumentation frameworks.

Definition 7 [**Structured Argumentation Frameworks**] Let AT be an *argumentation theory* $(AS, \mathcal{K})$. A *structured argumentation framework (*SAF*)* defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where $\mathcal{A}$ is the set of all finite arguments constructed from $\mathcal{K}$ in AS, $\preceq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y. A *c-structured argumentation framework (*c-SAF*)* is defined likewise except that $\mathcal{A}$ is the set of all finite arguments constructed from $\mathcal{K}$ with indirectly consistent set of premises.

The notion of *defeat* can then be defined as follows ($A \prec B$ is defined as usual as $A \preceq B$ and $B \not\preceq A$ and $A \approx B$ as $A \preceq B$ and $B \preceq A$).

Definition 8 [**Defeat**] A *defeats* B iff either A undercuts B; or A rebuts or undermines B on B' and $A \not\prec B'$.

Example 5 In our running example A_6 defeats A_5 unless $A_6 \prec A_5$. Furthermore, regardless of the argument ordering, A_7 defeats A_4 (and thus A_6).

Abstract argumentation frameworks are then generated from (c-)SAFs as follows:

Definition 9 [**Argumentation frameworks**] An *abstract argumentation framework (AF) corresponding to a* (c-)SAF = $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ is a pair $(\mathcal{A}, D)$ such that D is the defeat relation on $\mathcal{A}$ determined by (c-)SAF.

A nonmonotonic consequence notion can then be defined as follows. Let $T \in \{$complete, preferred, grounded, stable$\}$ and let $\mathcal{L}$ be from the AT defining $(c) - SAF$. A wff $\varphi \in \mathcal{L}$ is *sceptically T-justified* in $(c-)SAF$ if φ is the conclusion of a sceptically T-justified argument, and *credulously T-justified* in $(c-)SAF$ if φ is not sceptically T-justified and is the conclusion of a credulously T-justified argument.

[9] prove that for so-called 'well-defined' argumentation theories with so-called 'reasonable' argument orderings the extensions induced by Definition 9 satisfy all four rationality postulates of the rationality postulates of [1]. These and some related notions are defined as follows.

Definition 10 [**Well defined *(c-)SAF*s**] Let $AT = (AS, \mathcal{K})$ be an argumentation theory, where AS = $(\mathcal{L}, \mathcal{R}, n)$. We say that AT is:

- *closed under contraposition* iff for all $S \subseteq \mathcal{L}$, all $\varphi \in \mathcal{L}$ and all $\psi \in S$: if $S \vdash \varphi$, then $S \setminus \{\psi\} \cup \{\varphi'\} \vdash \psi'$ for all φ' such that $\varphi' = -\varphi$ and all ψ' such that $\psi' = -\psi$.
- *closed under transposition* iff if $\varphi_1, \ldots, \varphi_n \rightarrow \psi \in \mathcal{R}_s$, then for $i = 1 \ldots n$, $\varphi_1, \varphi_{i-1}, \psi', \varphi_{i+1}, \ldots, \varphi_n \rightarrow \varphi'_i \in \mathcal{R}_s$ for all φ'_i such that $\varphi'_i = -\varphi_i$ and all ψ' such that $\psi' = -\psi$.
- *axiom consistent* iff $\mathcal{K}_n$ is indirectly consistent.

If a *(c-)SAF* is defined by an AT that is axiom consistent and closed under contraposition or transposition, then the *SAF* is said to be *well defined*.

Henceforth, any *(c-)SAF* is assumed to be well defined.

Example 6 The argumentation theory in our running example is axiom consistent since $\{p\}$ is indirectly consistent. It can be made closed under contraposition or transposition by adding $p, \neg s \rightarrow \neg r$ and $r, \neg s \rightarrow \neg p$ and $r_1 \rightarrow s$ to $\mathcal{R}_s$.

We now define strict continuations of arguments slightly differently than in [9].[1]

Definition 11 [**Strict continuations**] The set of *strict continuations* of any set of arguments from $\mathcal{A}$ is the smallest set satisfying the following conditions:

1. Any argument A is a strict continuation of $\{A\}$.
2. If $A_1, \ldots, A_n$ and $S_1, \ldots, S_n$ are sets of arguments such that all A_i are a strict continuation of S_i and all of $B_1, \ldots, B_n$ are infallible arguments, then $A_1, \ldots, A_n, B_1, \ldots, B_n \rightarrow \varphi$ is a strict continuation of $S_1 \cup \ldots \cup S_n$.

Example 7 In our running example all arguments are strict continuations of themselves while A_6 is a strict continuation of $\{A_4\}$ and A_7 is a strict continuation of A_5.

Definition 12 [**Reasonable Argument Orderings**] An argument ordering $\preceq$ is *reasonable* iff:

1. i) $\forall A, B$, if A is infallible and B is fallible, then $B \prec A$;
   ii) $\forall A, B$, if B is infallible then $B \not\prec A$;
   iii) $\forall A, A', B$ such that A' is a strict continuation of $\{A\}$, if $A \not\prec B$ then $A' \not\prec B$, and if $B \not\prec A$ then $B \not\prec A'$ (i.e., applying strict rules to a set of arguments of which at most one is fallible does not weaken, resp. strengthen, arguments).
2. Let $\{C_1, \ldots, C_n\}$ be a finite subset of $\mathcal{A}$, and for $i = 1 \ldots n$, let $C^{+\setminus i}$ be some strict continuation of $\{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\}$. Then it is not the case that: $\forall i$, $C^{+\setminus i} \prec C_i$.

Example 8 In our running example, Conditions 1(i,ii) make that $A_1 \not\prec A_1$ and $A_i \prec A_1$ for all i such that $1 < i \leq 7$. Suppose we further have $A_5 \not\prec A_6$. Then by 1(iii) we also have $A_7 \not\prec A_6$. Suppose we also have $A_2 \not\prec A_7$; then by 1(iii) we also have $A_2 \not\prec A_5$. To illustrate Condition (2), let us temporarily move p from $\mathcal{K}_n$ to $\mathcal{K}_p$ and suppose $\mathcal{R}_s$ is closed under transposition. Then the following new arguments can be constructed:

$$A_8 = \quad A_1, A_5 \rightarrow \neg r \qquad A_9 = \quad A_4, A_5 \rightarrow \neg p$$

[1] The new definition is arguably simpler but does not affect the proofs of [9].

Note that A_6 strictly continues $\{A_1, A_4\}$, A_8 strictly continues $\{A_1, A_5\}$ and A_9 strictly continues$\{A_4, A_5\}$. Then we cannot have all of $A_6 \prec A_5$ and $A_8 \prec A_4$ and $A_9 \prec A_1$.

Finally, in some proofs below the notion of a maximum fallible subargument is used. The following definition improves the one of [9], which does not satisfy Lemma 11 below.

Definition 13 [**Maximal fallible subarguments**] For any argument A, the set $M(A)$ of *maximal fallible subarguments* of A is inductively defined as:

1. If $A \in \mathcal{K}_n$, then $M(A) = \emptyset$;
2. If $A \in \mathcal{K}_p$ or A has a defeasible top rule, then $M(A) = \{A\}$;
3. otherwise, i.e., if A is of the form $A_1, \ldots, A_n \rightarrow \varphi$, then $M(A) = M(A_1) \cup \ldots \cup M(A_n)$.

Example 9 In our running example we have that $M(A_1) = \emptyset$, $M(A_2) = \{A_2\}$, $M(A_3) = \{A_3\}$, $M(A_4) = M(A_6) = \{A_4\}$, $M(A_5) = M(A_7) = \{A_5\}$.

3. Changing the *ASPIC*$^+$ framework

We now reconsider the rationality postulates of [1] in light of our discussion in Section 1 and then propose a modified version of *ASPIC*$^+$. Our proposal applies to both sceptical and credulous justification (cf. Definition 1), since an extension can be seen as a set of arguments that a rational agent could accept. We will discuss the rationality postulates as applying to single extensions, but note that if they are satisfied for single extensions, they are easily provable for the intersection of all extensions (cf. [1,9]).

We first discuss the consistency and strict-closure postulates[2]. Direct consistency is not put into question by the lottery paradox or similar examples: it seems plainly irrational to simultaneously accept two propositions that negate each other. However, for strict closure and indirect consistency things are different. As discussed in Section 1, if a deductive inference is applied to at least two fallible subarguments, then it *aggregates* the 'amounts' of fallibility of its subarguments. This in turn means that the argument applying the deductive inference may be less preferred than either of these subarguments, so a successful attack on it does not imply a successful attack on one of these subarguments. Note that this line of reasoning does not apply to cases where a deductive inference is applied to at most one fallible element: then the amount of fallibility of the new argument is exactly the same as the amount of fallibility of the single fallible argument to which the deductive inference is applied. So we want to weaken the demand of strict closure to those subsets of an extension that contain at most one fallible argument. Combined with the wish to retain direct consistency, this implies a wish to restrict indirect consistency in the same way as strict closure.

We next discuss the changes in *ASPIC*$^+$. Consider the following modelling of the lottery paradox. Let $\mathcal{L}$ be a propositional language built from the set of atoms $\{T_i \mid 1 \leq i \leq 1,000,000\}$. Then let X denote a well-formed formula $X_1 \veebar \ldots \veebar X_{1,000,000}$ where $\veebar$ is exclusive or and where each X_i is of one of the following forms:

[2]For reasons of space, we do not formally list the postulates of [1] and leave the formulation of the new postulates implicit in the formal results of Section 4.

- If $i = 1$ then $X_i = T_1 \wedge \neg T_2 \wedge \ldots \wedge \neg T_n$
- If $i = n$ then $X_i = \neg T_1 \wedge \neg T_2 \wedge \ldots \wedge \neg T_{n-1} \wedge T_n$
- Otherwise $X_i = \neg T_1 \wedge \ldots \wedge \neg T_{i-1} \wedge T_i \wedge \neg T_{n+1} \wedge \ldots \wedge \neg T_n$

Next we choose $\mathcal{K}_p = \{\neg T_i \mid 1 \leq i \leq 1,000,000\}$, $\mathcal{K}_n = \{X\}$, $\mathcal{R}_s$ as consisting of all propositionally valid inferences from finite sets and $\mathcal{R}_d = \emptyset$.

We want to formalise an account of the paradox in which for each individual ticket the statement that it will not win is sceptically justified, in which the statement that exactly one ticket will win is sceptically justified and in which the justification status of conjunctions of statements that a ticket will not win depends on the size of the conjuncts. In this section we only discuss the first two demands; the last one will be discussed in Section 5. Our analysis does not depend on the choice of semantics. The following arguments are relevant for any i such that $1 \leq i \leq 1,000,000$.

$$\neg T_i \qquad \text{and} \qquad \neg T_1, \ldots, \neg T_{i-1}, \neg T_{i+1}, \ldots, \neg T_{1,000,000}, X \rightarrow T_i \text{ (call it } A_i\text{)}$$

This requires for all i that $A_i \prec \neg T_i$, to prevent A_i from defeating $\neg T_i$. This in turn requires that Condition (2) of Definition 12 of reasonable argument orderings is dropped, since it excludes such an argument ordering. On the other hand, Condition (1) of Definition 12 can be retained. In particular, Condition (1.iii) captures that applying a strict rule to the conclusion of a single argument A to obtain an argument A' does not change the 'preferedness' of A' compared to A. This is reasonable in general, since A and A' have exactly the same set of fallible elements (ordinary premises and/or defeasible inferences).

Finally, we need to allow rebutting attacks on strict-rule applications applied to at least two fallible subarguments, since otherwise A_i is not defeated and both A_i and $\neg T_i$ are justified, which violates direct consistency. However, such rebuttals should not be allowed on strict rules applied to just one fallible argument, since then strict closure and indirect consistency do for preferred and stable semantics not even hold for strict inferences from at most one fallible subargument. A counterexample is $\mathcal{R}_d = \mathcal{K}_n = \emptyset$, $\mathcal{R}_s = \{b \rightarrow \neg m, m \rightarrow \neg b\}$ and $\mathcal{K}_p = \{b, m\}$. Then $\{b, m\}$ is an admissible set [1].

Based on this analysis, *ASPIC*$^+$ is now adapted as follows. First, the definition of rebutting attack in Definition 6 is replaced with the following definition.[3]

Definition 14 [Semi-restricted rebut] A *rebuts* argument B (on B') iff for some $B' \in \texttt{Sub}(B)$ it holds that $\texttt{Conc}(A) = -\varphi$ and either:

1. B' is of the form $B_1, \ldots, B_n \Rightarrow \varphi$; or
2. B' is of the form $B_1, \ldots, B_n \rightarrow \varphi$ and $n \geq 2$ and at least two of $B_1, \ldots, B_n$ are fallible.

Example 10 In our running example A_5 does still not rebut A_6 since A_6 applies its strict top rule to just one fallible subargument. However, if p is moved from $\mathcal{K}_n$ to $\mathcal{K}_p$, then A_5 does rebut A_6.

Definition 8 of defeat then directly applies to the modified framework. Finally, argument orderings are from now on assumed to be *weakly reasonable* in that they satisfy Condition (1) of Definition 12.

[3] [1,2] investigate similar notions of rebutting attack. However, they allow rebuttals on strict rules applied to only one fallible argument and do not investigate weakened versions of the rationality postulates.

4. The new rationality postulates verified

We now verify that the changed *ASPIC*$^+$ framework satisfies [1]'s postulates of closure under subarguments and direct consistency plus the new postulates of 'restricted' strict closure and 'restricted' indirect consistency. The results and proofs are based on those of [9] but reformulated or adapted when needed. For ease of comparison the original numbering of [9] is retained. In fact, for c-SAFs the results can only be proven under the assumption that an argument's premises joined with $\mathcal{K}_n$ is consistent. Accordingly, the notion of a c-SAF is redefined as follows:

Definition 15 [**c-Structured Argumentation Frameworks redefined**] Let $AT = (AS, \mathcal{K})$ be an *argumentation theory*. A *c-structured argumentation framework (*c-SAF*)* defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where $\mathcal{A}$ is the set of all finite arguments constructed from $\mathcal{K}$ in AS such that for all $A \in \mathcal{A}$ it holds that $\texttt{Prem}(A) \cup \mathcal{K}_n$ is indirectly consistent, $\preceq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

Well-defined structured argumentation frameworks for *ASPIC*$^+$ with semi-restricted rebut and a weakly reasonable argument ordering are below denoted with $(c-)SAF^{sw}$, where $c - SAF^{sw}$'s are defined as in Definition 15.

Lemma 11 For any argument A: $\texttt{Conc}(M(A)) \cup \texttt{Prem}_n(A) \vdash \texttt{Conc}(A)$.

PROOF. By induction on the structure of arguments. The result is obvious if $A \in \mathcal{K}$ or $\texttt{TopRule}(A) \in \mathcal{R}_d$. If $\texttt{TopRule}(A) \in \mathcal{R}_s$, then by the induction hypothesis $\texttt{Conc}(A_i) \in Cl_{\mathcal{R}_s}(\texttt{Conc}(M(A_i)) \cup \texttt{Prem}_n(A_i))$ for all A_i $(1 \leq i \leq n)$. Since $\texttt{Prem}_n(A) = \texttt{Prem}_n(A_1) \cup \ldots \cup \texttt{Prem}_n(A_n)$, the result follows. QED

Proposition 8 For any argument A and fallible argument B that have contradictory conclusions: (1) A defeats B; or (2) some strict continuation $A+$ of A defeats B.

PROOF. If B has no strict top rule or a top rule applied to at least two fallible arguments, then clearly A defeats B. Otherwise, consider first systems closed under contraposition (Def. 10). By Lemma 11 it holds that $\texttt{Conc}(M(B)) \cup \texttt{Prem}_\texttt{n}(B) \vdash \texttt{Conc}(B)$. By contraposition, and since $\texttt{Conc}(A)$ and $\texttt{Conc}(B)$ contradict each other and $M(B) = \{B'\}$, we have that $\texttt{Prem}_\texttt{n}(B) \cup \texttt{Conc}(A) \vdash \varphi$ for some φ such that $\varphi = -\texttt{Conc}(B')$. Hence, one can construct a strict continuation $A+$ of A that concludes φ. Since by construction of $M(B)$ either B' is an ordinary premise or ends with a defeasible inference, $A+$ either undermines or rebuts B'. But then $A+$ also undermines or rebuts B.

For systems closed under transposition the existence of argument $A+$ is proven by straightforward generalisation of Lemma 6 of [1]. Then the proof is completed as above. In the case of c-SAFs, it must also be shown that $\texttt{Prem}(A+) \cup \mathcal{K}_n$ is indirectly consistent, which follows given $\texttt{Prem}(A+) \subseteq \texttt{Prem}(A) \cup \texttt{Prem}_\texttt{n}(B)$ and $\texttt{Prem}_n(B) \subseteq \mathcal{K}_n$, and $\texttt{Prem}(A) \cup \mathcal{K}_n$ is indirectly consistent by assumption.

2) Since $A+$ is a strict extension of A and B is a strict extension of B' and $A \not\preceq B$, we have $A+ \not\preceq B'$ by Condition (1c) of Definition 12, so $A+$ defeats B' and B. QED

Lemma 37 Let $(\mathcal{A}, \mathcal{C}, \preceq)$ be a $(c-)SAF^{sw}$. Let $A \in \mathcal{A}$ be a strict continuation of $S = \{A_1, \ldots, A_n\} \subseteq \mathcal{A}$ such that at most one member of S is fallible, and for $i = 1 \ldots n$, A_i is acceptable w.r.t. an admissible set $E \subseteq \mathcal{A}$. Then A is acceptable w.r.t. E.

PROOF. Let B be any argument defeating A. By Def. 6 of attack and Def. 14 of semi-restricted rebut, B attacks A by undercutting or rebutting on defeasible rules in A or undermining on an ordinary premise in A. Hence, by definition of strict continuations (Def. 11), it must be that B attacks A iff B attacks A_i for the unique fallible $A_i \in \{A_1, \ldots, A_n\}$. Either:

1) B undercuts A_i, and so by Def. 8, B defeats A_i, or:
2) B does not undercut A_i. Suppose $B \prec A'_i$. This contradicts B defeats A. Hence, B defeats A_i.

We have shown that if B defeats A then B defeats some $A_i \in S$. By assumption of A_i acceptable w.r.t. E and E being admissible, $\exists C \in E$ s.t. C defeats B. Hence, A is acceptable w.r.t. E. QED

Proposition 9 Let ($\mathcal{A}$, $\mathcal{C}$, $\preceq$) be a $c - SAF^{sw}$. If $A_1, \ldots, A_n$ are acceptable w.r.t. some admissible set $E \subseteq \mathcal{A}$ and at most one of $A_1, \ldots, A_n$ is fallible, then $\bigcup_{i=1}^{n} \texttt{Prem}(A_i) \cup \mathcal{K}_n$ is indirectly consistent.

PROOF. Suppose for contradiction otherwise and let S be any minimally indirectly inconsistent subset of $\bigcup_{i=1}^{n} \texttt{Prem}(A_i)$. Then for all $\varphi \in S$, $S \setminus \{\varphi\} \vdash \varphi'$ for all φ' such that $\varphi' = -\varphi$ and $S \setminus \{\varphi\}$ is indirectly consistent. Since at most one of $A_1, \ldots, A_n$ is fallible, we thus have for some A_i the set of ordinary premises $S = \{\varphi_1, \ldots, \varphi_m\} \subseteq \texttt{Prem}(A_i)$ (that must be non-empty given that $\mathcal{K}_n$ is indirectly consistent by assumption of axiom consistency (Def. 10)), that S is consistent but $S \cup \mathcal{K}_n$ is inconsistent. But this contradicts the fact that $\texttt{Prem}(A_i) \cup \mathcal{K}_n$ is indirectly consistent. QED

Theorem 12 [***Sub-argument Closure***] Let $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ be a *(c-)SAF* and E a complete extension of Δ. Then for all $A \in E$: if $A' \in \texttt{Sub}(A)$ then $A' \in E$.

PROOF. As in [9]. QED

Theorem 13 [***Restricted closure under Strict Rules***] Let $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ be a $(c-)SAF^{sw}$ and E a complete extension of Δ and let $S \subseteq E$ be such that at most one element of S is fallible. Then $\texttt{Conc}(S) = Cl_{R_s}(\texttt{Conc}(S))$.

PROOF. It suffices to show that any strict continuation X of S is in E. By Lemma 37, any such X is acceptable w.r.t. E. By Proposition 10 of [9], $E \cup \{X\}$ is conflict free. Hence, since E is complete, $X \in E$. Note that if Δ is a c-*SAF*, then Proposition 9 guarantees that $\texttt{Prem}(X) \cup \mathcal{K}_n$ is indirectly consistent. QED

Theorem 14 [***Direct Consistency***] Let $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ be a $(c-)SAF^{sw}$ and E a complete extension of Δ. Then $\{\texttt{Conc}(A) | A \in E\}$ is directly consistent.

PROOF. We show that if $A, B \in E$, $\texttt{Conc}(A) = -\texttt{Conc}(B)$, a contradiction results.

1. A is infallible, and: **1.1** if B is infallible, then this contradicts the assumption that $\mathcal{K}_n$ is consistent. **1.2** if B is fallible, and **1.2.1** B is an ordinary premise or has a defeasible top rule or has a strict top rule applied to at least two fallible subarguments, then A defeats B contradicting E is conflict free, or **1.2.2** B has a strict top rule applied to at most one fallible subargument (see **3** below).

2. A is fallible, and: **2.1** if B is infallible then either **2.1.1** A is an ordinary premise or has

a defeasible top rule or has a strict top rule applied to at least two fallible subarguments, in which case B defeats A, contradicting E is conflict free, or **2.1.2** A has a strict top rule applied to at most one fallible subargument (see **3** below); **2.2** if B is fallible and **2.2.1** B is an ordinary premise or has a defeasible top rule or has a strict top rule applied to at least two fallible subarguments, then either A defeats B or B defeats A, contradicting E is conflict free, or **2.2.2** B has a strict top rule applied to at most one fallible subargument (see **3** below).

3. Each of **1.2.2**, **2.1.2** and **2.2.2** describes the case where $X, Y \in E$, $\texttt{Conc}(X) = -\texttt{Conc}(Y)$, Y is fallible and has a strict top rule applied to at most one fallible subargument. In the case that Δ is a c-*SAF*, since $X, Y \in E$, then X, Y are acceptable w.r.t. E, and so by Proposition 9, $\texttt{Prem}(A) \cup \texttt{Prem}(B) \cup \mathcal{K}_n$ is indirectly consistent. By Proposition 8 there is a strict continuation $X+$ of X that defeats Y. By Lemma 37 $X+$ is acceptable w.r.t. E, and by Proposition 10 of [9], $E \cup \{X+\}$ is conflict free, contradicting $X+$ defeats Y. QED

Then Theorem 15 follows from Theorems 13 and 14.

Theorem 15 [***Restricted Indirect Consistency***] Let $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ be a $(c-)SAF^{sw}$ and E a complete extension of Δ and let $S \subseteq E$ be such that at most one element of S is fallible. Then $\texttt{Conc}(S)$ is indirectly consistent.

5. Conclusion

We first verify that the new variant of *ASPIC*$^+$is a middle ground between the extremes of Pollock and Kyburg in that whether a deductive consequence of multiple rationally acceptable propositions is also rationally acceptable depends on the specific example. The crucial element here is the argument ordering. Recall the modelling in Section 3 of the lottery paradox and assume that arguments have a numerical fallibility degree f, being the number of ordinary premises that they use. Next we define a 'bandwidth' for strict argument preference, by letting for any pair of fallible arguments A and B, $A \prec B$ iff $f(A) - f(B) > n$ for some natural number n. More sophisticated argument orderings may be possible but this one suffices to illustrate our point. Now if, for example, $n = 600,000$ and adopting preferred semantics for illustration, then all arguments for conjunctions $\neg T_i \wedge \ldots \wedge \neg T_j$ with fewer than 200,000 conjuncts strictly defeat their rebutting counterarguments and are thus in all preferred extensions, the arguments for conjunctions between $200,000$ and $800,000$ conjuncts defeat and are defeated by their rebutting counterarguments so are in some but not all preferred extensions, while the arguments with more than $800,000$ conjuncts are strictly defeated by their rebutting counterarguments so are not in any preferred extension.

We next conclude. In this paper we presented an argumentation-based notion of fallible rational acceptance according to which one can sometimes rationally accept sets of propositions that are indirectly inconsistent or not strictly closed. We proposed new rationality postulates capturing this idea and proposed a variant of *ASPIC*$^+$ that satisfies the new postulates while not satisfying their original versions. While we illustrated these ideas with a purely probabilistic example, the basic intuition is more general, being that an argument formed by strictly extending more than one fallible subargument has more

fallibility than each of the combined arguments alone. Therefore, the relevance of this paper is not confined to discussions of the lottery paradox but extends to any application of argumentation in which arguments can have multiple fallible elements.

Our approach captures the intermediate position that deductive inferences from multiple fallibly acceptable propositions can but need not be acceptable. The argumentation approach here provided a fresh logical perspective compared to other logical approaches. First, the truth-preserving nature of deductive inference rules is respected by allowing their application inside arguments as strict rules. A key observation here is that preservation of truth does not imply preservation of rational acceptability, since truth and rational acceptability are different things. A virtue of an argumentation approach is that it can naturally model this distinction, since the strict-closure postulate does not capture preservation of truth but preservation of rational acceptability. Second, argumentation can make a natural distinction between cases where strict closure and indirect consistency do and do not hold, since if an argument that applies a deductive inference to fallible subarguments is not rebutted on this inference or if none of its rebuttals are strong enough to defeat it, then this argument can still be acceptable. The notion of an argument ordering is crucial here, since it can make fine-grained distinctions between cases where applications of deductive inferences are and are not strong enough to survive attack.

Having said so, it remains to be investigated how argument orderings can be defined in principled ways. For example, can they help in modelling argumentation-based counterparts of [8]'s "lossy" inference rules, or [3]'s "big-step probabilities" (their attempt to distinguish between cases with and without uniform underlying probability structures)? Such investigations could shed further light on the relation between argumentation-based and other logical modellings of reasoning with uncertain information.

References

[1] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171:286–310, 2007.

[2] M. Caminada, S. Modgil, and N. Oren. Preferences and unrestricted rebut. In S. Parsons, N. Oren, C. Reed, and F. Cerutti, editors, *Computational Models of Argument. Proceedings of COMMA 2014*, pages 209–220. IOS Press, Amsterdam etc, 2014.

[3] D. Dubois, H. Fargier, and H. Prade. Ordinal and probabilistic representations of acceptance. *Journal of Artificial Intelligence Research*, 22:23–56, 2004.

[4] 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.

[5] R. Foley. Beliefs, degrees of belief, and the Lockean thesis. In F. Huber and C. Schmidt-Petri, editors, *Degrees of belief*, volume 342 of *Synthese Library*, pages 37–47. Springer, 2009.

[6] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical-logic arguments: postulates and properties. *Artificial Intelligence*, 175:1479–1497, 2011.

[7] H. Kyburg. *Probability and the Logic of Rational Belief.* Wesleyan U. P. Middletown, CT, 1961.

[8] D. Makinson. Logical questions behind the lottery and preface paradoxes: lossy rules for uncertain inference. *Synthese*, 186:511–529, 2012.

[9] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.

[10] J.L. Pollock. Defeasible reasoning. In J. Adler and L. Rips, editors, *Reasoning: Studies of Human Inference and its Foundations*, pages 451–470. Cambridge, Cambridge University Press, 2007b.

[11] D.L. Poole. The effect of knowledge on belief: Conditioning, specificity and the lottery paradox in default reasoning. *Artificial Intelligence*, 49:281–307, 1991.

[12] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.