Relevance and Minimality in Systems of Defeasible Argumentation

Johannes C. Flieger*

Department of Computing,
Imperial College of Science,
Technology and Medicine,
London SW7 2BZ,
United Kingdom

October 2002

Abstract

We present a metalogical characterisation of relevance for systems of defeasible argumentation and use it to define the notion of a relevant argument system. We employ a variant of the idea (influential in linguistics and philosophy) that communication and cognition are governed by a trade-off between opposing demands of informational sufficiency and economy of means; the notion of informational sufficiency is modelled in terms of satisfying a query associated with a topic of argumentation, while the notion of economy is based on proof-theoretic minimality. The resulting system of relevant argumentation is able to handle fallacies of relevance, such as the paradoxes of material implication, even when the underlying deductive system is a classical rather than a relevance logic.

*I would like to thank Marek Sergot, Francesca Toni, Robyn Carston, Alessio Lomuscio, Antonio de Freitas and Yochanan Cordovero for encouragement, guidance, and helpful discussions regarding the topics of this paper.

1 Introduction

In this paper, we present a metalogical characterisation of relevance for systems of defeasible argumentation based on a proof-theoretic notion of minimality.

In argumentation, irrelevance is undoubtedly a defect and relevance a virtue. Indeed, relevance is of fundamental importance in human reasoning in general, and is an issue that arises, in some form, in practically any area involving the modelling of communication and reasoning among humans or between humans and machines, including information retrieval, information extraction, automated text summarisation, machine learning, reasoning under uncertainty, natural language processing, intelligent interface design and query-answering systems.

In the logical tradition, the issue of relevance has largely revolved around ways of dealing with the (so-called) paradoxes of material and strict implication, which are paradigmatic instances of the more general class of *fallacies of relevance*:

> The archetype of fallacies of relevance is $A \rightarrow .B \rightarrow A$, which would enable us to infer that Bach wrote the Coffee Cantata from the premiss that the Van Allen belt is doughnut-shaped—or indeed from any premiss you like. [AB75, p. 30]

A natural response is to conclude that classical logic is unsuitable as a basis for formalising relevant reasoning, and that it should be extended and/or modified to yield a system with the desired properties. This view is most prominently represented by the various relevance logics, viz., logical systems designed to ensure that the fallacies of relevance do not arise, and which indeed extend and deviate from classical logic in various ways.[1]

In this paper, we develop a theoretical characterisation of argumentational relevance comprising

- a definition of topic-relevance for arguments;

[1]For a succint overview of the central systems of relevance logic, see Dunn [Dun83]. A brief discussion of their philosophical significance and the challenge they present to the classical notion of validity, see Haack [Haa78, ch. 10]. Historically, the development of such systems owes much to the work of Anderson and Belnap [AB75, ABD92], but has antecedents in the logical and mathematical tradition, especially the investigations of Church [Chu51], Moh [Moh50], and Ackermann [Ack56, Ack58].

- a definition of topic-relevance for sentences;
- an account of how relevance propagates in an argument system;
- a definition of the relation of relevant attack (and therefore relevant defeat);
- a characterisation of a relevant argument system.

We also address the issue of whether an argument system, if it is to avoid the so-called 'fallacies of relevance', ought to be based on a relevance logic. We shall see that this is not necessary: a system of relevant argumentation based on a classical (or subclassical) logic can handle the paradoxes of implication, for example, without denying their status as theorems.

This paper is organised as follows. Section 2 describes the components of a simple argument system and section 3 reviews certain basic properties of natural deduction derivations. In section 4 we address the issue of topic relevance in arguments. Section 5 considers the notion of how topic relevance propagates in argument systems, and defines the notion of 'relevant attack', which we use in section 6 as the basis for defining a relevant argument system. Section 7 addresses the paradoxes of implication from an argumentational standpoint. Section 8 discusses related work, and section 9 concludes.

2 Argument Systems

Argumentation systems attempt to model patterns of defeasible reasoning in which arguments for and against a certain claim are advanced and evaluated. Such systems typically comprise the following five components:[2]

- An underlying logical language.
- A definition of an argument.
- A definition of conflict between arguments.
- A definition of the defeat relation between arguments.
- A definition of the justificational status of an argument.

The distinction between an abstract argument system and a concrete one depends on the degree to which the above components are specified. We shall refer to a system that abstracts away from internal

[2]Cf. Prakken and Vreeswijk [PV01].

details of specific arguments (treating them as primitives), as well as the temporal order in which arguments are introduced, as an *abstract* argument system:[3]

Definition 1 (Abstract argument system) *An (abstract) argument system $\mathcal{S}$ is a pair*

$$\mathcal{S} = \langle \mathcal{A}, \longrightarrow \rangle \tag{1}$$

where $\mathcal{A}$ is a set of arguments, and $\longrightarrow$ is a relation, the 'attack' relation, holding between pairs of arguments in $\mathcal{A}$.

Thus defined, abstract argument systems can be represented naturally as directed graphs, in which arguments are represented by nodes in the graph and the attack relation by a set of directed arcs.

In contrast to this essentially *static* notion of argument systems, we can conceive of argumentation as a dynamic, dialectic process of attacking and defending propositions of interest: an (idealised) dialogue in which one agent attempts to persuade another of the truth (or falsity) of some state of affairs. The two perspectives are complementary, not contradictory: the dynamic process of argumentation can be viewed as the unfolding over time of an abstract argument system.[4]

2.1 Deductive Framework

An argument system typically contains a deductive system at its core, although it may also contain defeasible rules of an inductive, abductive, or analogical nature. For the underlying logic, we assume a Prawitz-style natural deduction system.[5]

Definition 2 (Deductive system) *A deductive system $\mathfrak{S}$ is a pair*

$$\mathfrak{S} = \langle \mathfrak{L}, \mathfrak{R} \rangle$$

such that

[3] For discussion of abstract argument systems, see Dung [Dun95], Kowalski and Toni [KT96], and Vreeswijk [Vre97].

[4] The correspondence is many-one: a given argument exchange defines a unique (up to isomorphism) abstract argument system, but a given argument system may allow for for several possible temporal unfoldings.

[5] Cf. Prawitz [Pra65]; details of the exposition are based on Tennant [Ten78] and Van Dalen [vD94].

- $\mathfrak{L}$ *is a formal language consisting of countably many formulae, and*
- $\mathfrak{R}$ *is a set of inference rules, containing the basis rule*

$$\mathrm{d}_{\mathfrak{L}}(\varphi, \varphi, \{\varphi\}) \textit{ for all } \varphi \in \mathfrak{L},$$

and a further set of rules of the form

$$\left\{ \begin{array}{c} \mathrm{d}_{\mathfrak{L}}(\mathcal{D}_1, \varphi_1, \Gamma_1) \\ \vdots \\ \mathrm{d}_{\mathfrak{L}}(\mathcal{D}_n, \varphi_n, \Gamma_n) \\ \\ \text{and} \\ \\ \mathrm{k}(\mathcal{D}_1, \ldots, \mathcal{D}_n, \varphi) \end{array} \right\} \Longrightarrow \mathrm{d}_{\mathfrak{L}}(\frac{\mathcal{D}_1 \cdots \mathcal{D}_n}{\varphi}, \varphi, \mathrm{f}(\mathcal{D}_1, \ldots, \mathcal{D}_n, \varphi))$$

where k *is an effectively determinable condition on* $\mathcal{D}_1, \ldots, \mathcal{D}_n$*, and* φ*, and* f *is the effectively determinable set of premises on which the conclusion* φ *depends.*

We take a derivation to be a (finite) tree of sentences of $\mathfrak{L}$. Since the derivation rules in $\mathfrak{R}$ have the form of clauses in an inductive definition, each deductive system $\langle \mathfrak{L}, \mathfrak{R} \rangle$ defines a corresponding set of derivations: $\mathfrak{D}_{\langle \mathfrak{L}, \mathfrak{R} \rangle} = \{\mathcal{D} \mid \mathrm{d}_{\mathfrak{L}}(\mathcal{D}, \varphi, \Gamma) \textit{ for } \varphi \in \mathfrak{L}, \Gamma \subseteq \mathfrak{L}\}$.

Convention 1

- *Where* $\mathrm{f}(\mathcal{D}_1, \ldots, \mathcal{D}_n, \varphi)$ *is simply the union of the premises* $\Gamma_1 \cup \ldots \cup \Gamma_n$ *of the subordinate derivations* $\mathcal{D}_1, \ldots, \mathcal{D}_n$*, then the rule of inference may be represented graphically as*

$$\frac{\varphi_1 \quad \cdots \quad \varphi_n}{\varphi}$$

where $\varphi_1, \ldots, \varphi_n$ *are the respective conclusions of* $\mathcal{D}_1, \ldots, \mathcal{D}_n$.

- *Where* $\mathrm{k}(\mathcal{D}_1, \mathcal{D}_2, \varphi)$ *is the null condition, then it may be omitted.*
- *If we want to make explicit mention of the premises and conclusion of a derivation, we may write*

$$\begin{array}{c} \Gamma \\ \mathcal{D} \\ \varphi \end{array}$$

instead of just $\mathcal{D}$.

- *If $\mathcal{D}$ is a derivation, then* $\dfrac{\begin{matrix}[\gamma]^i \\ \mathcal{D}\end{matrix}}{\varphi}\, i$ *is a derivation with zero or more occurrences of γ cancelled (the index i marks the step in the derivation which licenses the cancellation). We shall refer to the (uncancelled) premises of a derivation $\mathcal{D}$ by* $\mathrm{prem}(\mathcal{D})$.

- *If* $\mathrm{d}(\mathcal{D}, \varphi, \Gamma)$, $\gamma \in \Gamma$, *and* $\mathrm{d}(\mathcal{D}', \gamma, \Delta)$, *then* $\begin{matrix}\mathcal{D}' \\ (\gamma) \\ \mathcal{D}\end{matrix}$ *represents the derivation obtained by writing the derivation $\mathcal{D}'$ above each initial (uncancelled) occurrence of γ in $\mathcal{D}$; if $\gamma \notin \Gamma$, then* $\begin{matrix}\mathcal{D}' \\ (\gamma) \\ \mathcal{D}\end{matrix}$ *is just $\mathcal{D}$ itself.*

Example 1 The classical rule of $\wedge$-introduction is associated with the following conditions:

$$\left\{\begin{matrix} \mathrm{d}(\mathcal{D}_1, \varphi_1, \Gamma_1) \\ \\ \mathrm{d}(\mathcal{D}_2, \varphi_2, \Gamma_2) \\ \\ \textit{and} \\ \\ \mathrm{k}_{\wedge I}(\mathcal{D}_1, \mathcal{D}_2, (\varphi_1 \wedge \varphi_2)) \end{matrix}\right\} \Longrightarrow \mathrm{d}(\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{(\varphi_1 \wedge \varphi_2)}, (\varphi_1 \wedge \varphi_2), \mathrm{f}_{\wedge I}(\mathcal{D}_1, \mathcal{D}_2, (\varphi_1 \wedge \varphi_2)))$$

where $\mathrm{k}_{\wedge I}(\mathcal{D}_1, \mathcal{D}_2, (\varphi_1 \wedge \varphi_2))$ is the null condition and $\mathrm{f}_{\wedge I}(\mathcal{D}_1, \mathcal{D}_2, (\varphi_1 \wedge \varphi_2)) = \Gamma_1 \cup \Gamma_2$. By the conventions adopted above, the rule may be stated graphically as

$$\frac{\varphi_1 \qquad \varphi_2}{\varphi_1 \wedge \varphi_2}$$

Definition 3 (Subderivation) *A derivation $\mathcal{D}'$ is a subderivation of a derivation $\mathcal{D}$, written $\mathcal{D}' \sqsubseteq \mathcal{D}$, iff $\mathcal{D}'$ is a subtree of $\mathcal{D}$. If $\mathcal{D}' \sqsubseteq \mathcal{D}$ and $\mathcal{D}' \not\equiv \mathcal{D}$, then we refer to $\mathcal{D}'$ as a proper subderivation of $\mathcal{D}$, written $\mathcal{D}' \sqsubset \mathcal{D}$.*

It is often useful to speak of derivability, without having to specify the exact structure of a derivation:

Definition 4 ($\vdash$) $\Gamma \vdash \varphi$ *means that there exists a derivation with conclusion φ and with all (uncancelled) hypotheses in* Γ. $\mathrm{Th}(\Gamma)$ *is the set* $\{\varphi \in \mathfrak{L} | \Gamma \vdash \varphi\}$.

We say that θ is *derivable* from Γ.

Remark 1

- Because all deductions have finite length, every deductive system $\mathfrak{S} = \langle \mathfrak{L}, \mathfrak{R} \rangle$ is *compact*: whenever $\Delta \vdash \theta$, then $\Delta_0 \vdash \theta$ for some finite subset $\Delta_0 \subseteq \Delta$.
- A deductive system $\mathfrak{S}$ is said to be *monotonic* iff $\Delta \subseteq \Delta'$ implies $\mathrm{Th}(\Delta) \subseteq \mathrm{Th}(\Delta')$.

We shall confine our discussion in the sequel to propositional systems whose notion of logical consequence is monotonic; non-monotonicity is reflected in the possible change of status of an argument (say, from 'justified' to 'overruled'). That is, new premises or hypotheses do not invalidate arguments (in the sense of rendering the conclusion underivable from the premises), but simply give rise to counterarguments that may attack and perhaps defeat the original.

2.2 Identity Conditions for Arguments

Just as there are different levels of abstraction possible in argument systems, the definition of an argument admits of different levels of abstraction.

At the most abstract level, an argument is considered an unanalysed primitive of the system, defined exclusively in terms of the *external* attack relations it enters into (where the attack relation itself is taken as primitive).[6]

At the concrete level, an argument can be defined as a derivation in the underlying logical system:

Definition 5 (Argument) *If Δ is a (possibly inconsistent) subset of $\mathfrak{L}$, then an* argument *over Δ is a derivation* $\begin{matrix}\mathcal{D}\\ \varphi\end{matrix}$ *such that* $\mathrm{prem}(\mathcal{D}) \subseteq \Delta$*. We refer to the set* $\mathrm{prem}(\mathcal{D})$ *as the* premises *or* grounds *of the argument, and to φ as its* conclusion*.*

In cases where the discussion does not depend on the internal structure of an argument, we shall take the liberty of specifying an argument as an ordered pair of the form $\langle \textit{grounds}, \textit{conclusion} \rangle$.

Where the grounds of an argument are a subset of a database Δ, we shall speak of the database *supporting* the argument:

[6]Cf. Dung [Dun95].

Definition 6 (Support) *A database* Δ supports *an argument* $\langle \Gamma, \varphi \rangle$, *written* $\Delta \Vdash \langle \Gamma, \varphi \rangle$, *if* $\langle \Gamma, \varphi \rangle$ *is an argument over* Δ.

Intuitively, the database Δ is a set of sentences of $\mathfrak{L}$ that constitutes the 'common ground' between the agents participating in the argumentation process.[7]

2.3 Topics of Argumentation

Arguments are usually about something (even if only implicitly), and what counts as relevant in an argument is dependent on what the argument is about. We view arguments as centered around *topics*, which are simply the issue (or set of issues) that the argument is intended to resolve. One way of modelling aboutness is to treat the topic of the argument as a set of mutually exclusive claims, which in the simplest case reduces to a claim-counterclaim set consisting of a proposition and its contrary. It is then possible to define relevant arguments as those which provide total or partial support for some member of the set. We shall employ a prefix question mark, e.g., $?\psi$, to denote topics.[8] In this paper, we shall only consider topics consisting of a (propositional) sentence and its contrary:

Definition 7 *For all sentences* ψ *in* $\mathfrak{L}$: $?\psi \triangleq \{\psi, \bar{\psi}\}$.

The choice of the question mark is motivated by the similarity between the model of a topic as a claim-counterclaim set and the model of a query as the set of answers that could satisfy it. Thus, we can just as well think of arguments as competing answers to a hypothetical query. In view of this, we will speak interchangeably of an argument *addressing* a topic or *satisfying* a query:

Definition 8 *An argument* $\langle \Gamma, \varphi \rangle$ directly addresses *a topic* $?\psi$ *(or:* satisfies *a query* $?\psi$*) if* $\varphi \in\ ?\psi$ *(i.e.,* $\varphi \in \{\psi, \bar{\psi}\}$*).*

2.4 Basic Relations Between Arguments

We define the basic attack relation between individual arguments as follows:

[7] Cf. Fox *et al.* [FKA92], Krause *et al.* [K$^+$95].

[8] The scope of the '?' sign is understood to be the *largest* sentence occurring to its right.

Definition 9 (Attack) *An argument* $\begin{array}{c}\mathcal{D}\\ \gamma\end{array}$ *attacks an argument* $\begin{array}{c}\mathcal{D}'\\ \varphi\end{array}$ *if* $\bar{\gamma}$ *occurs in* $\begin{array}{c}\mathcal{D}'\\ \varphi\end{array}$*.*

Where $\varphi = \bar{\gamma}$*, the attack is referred to as a* conclusion attack*, and where* $\bar{\gamma} \in \mathrm{prem}(\begin{array}{c}\mathcal{D}'\\ \varphi\end{array})$*, the attack is referred to as an* assumption attack*.*

We distinguish weak and strong notions of the binary 'defeat' relation, defined over the set of arguments:

Definition 10 (Defeat) *An argument* a defeats *an argument* b*, written* $a \;\underline{\gg\!\!-}\; b$*, iff* a *attacks* b *and is not weaker than* b*.*

Definition 11 (Strict defeat) *An argument* a strictly defeats *an argument* b*, written* $a \gg\!\!- b$*, iff* a *attacks* b *and* a *is stronger than* b*.*

Since the rules of our deductive system are assumed to be valid, the defeasibility of an argument is ultimately a matter of the defeasibility of the grounds it rests upon.[9] That is, the database is assumed to contain premises of varying 'strength' (degree of confirmation, entrenchment). For the purposes of the present discussion, it suffices to assume two degrees of strength for premises, 'defeasible' and 'undefeasible'.

2.5 The Status of Arguments

Arguments may attack not only other arguments, but also themselves; and since an argument is only as strong as itself, a self-attacking argument is *ipso facto* a self-defeating one.

Definition 12 (Self-defeat) *An argument* $\langle \Gamma, \varphi \rangle$ *is self-defeating iff* $\Gamma \vdash \alpha, \bar{\alpha}$ *for some* $\alpha \in \mathfrak{L}$*.*

That is, an argument a is self-defeating if its premises can be used to construct two mutually attacking arguments.[10] Arguments that

[9] We will consider only purely deductive systems; that is, we will not discuss argument systems which contain defeasible rules of an inductive or analogical nature.

[10] Note that it is not required that a itself explicitly contain a contradiction.

are self-defeating are often referred to as 'trivial' or 'inconsistent', and non-self-defeating ones as 'non-trivial' or 'consistent'.

While the notion of self-defeat is *local* to a given argument, the notions of 'overruling' and 'justification' are *global* in nature, in that they are sensitive to other arguments that may be present in an argument system:

Definition 13 (Overruled argument) *An argument is* overruled *if it is not justified and either it is self-defeating, or it or one of its proper subarguments is defeated by a justified argument.*

Definition 14 (Justified argument) *An argument a is* justified *if*

1. *a is not self-defeating; and*
2. *All proper subarguments of a are justified; and*
3. *All arguments defeating a are overruled.*

3 Argument Minimality

3.1 Superfluous Premises

A derivation, as we have defined it, may contain superfluous hypotheses. This is unobjectionable from the perspective of truth-preservation: given monotonicity, we do not lose a proof merely by expanding the set of premises.

The opposition to superfluous premises is rather a pragmatic one: in the automated reasoning literature, the distaste for superfluous premises is largely motivated by concerns of computational economy or efficiency; this is also the case for linguistic and philosophical theories which view cognitive activity in economic terms, as geared towards the effective deployment of limited cognitive resources.[11] However, there are also specifically argument-theoretic grounds for keeping superfluous premises to a minimum, having to do with the wider range of attacks that superfluous premises invite.

Consider the following argument, call it a, that establishes the conclusion p:

[11] Cf. Sperber and Wilson [SW95].

$$\dfrac{\dfrac{p \qquad\qquad q}{p \wedge q}\ (\wedge I)}{p}\ (\wedge E)$$

Now suppose that we have a justified counterargument $b \equiv \langle \Gamma, \neg q \rangle$ such that b attacks and defeats subargument $\langle \{q\}, q \rangle$ of a; then b defeats a, in virtue of defeating one of its subarguments. Similarly for the case where we have a justified counterargument $c \equiv \langle \Gamma', \neg(p \wedge q) \rangle$ such that c attacks and defeats subargument $\langle \{p, q\}, (p \wedge q) \rangle$ of a. [12]

Now, the proponent of a could attempt to defend the argument by advancing arguments d, e such that d defeats b and e defeats c; but it is clear that the vulnerability to these attacks, and therefore the need to defend against them, can be traced to the dependency on the superfluous premise q.

We can distinguish between premises that are essential and those that are not as follows:

Definition 15 (Essential premise of an argument) *A premise* $\gamma \in \Gamma$ *is said to be* essential *to an argument* $\langle \Gamma, \varphi \rangle$ *if* $\Gamma - \{\gamma\} \not\vdash \varphi$.

We now define relative and absolute notions of what we shall refer to as 'premise minimality':

Definition 16 (Relative minimality of premises) *An argument* $a \equiv \langle \Gamma, \varphi \rangle$ *is premise minimal relative to an argument* $b \equiv \langle \Gamma', \varphi \rangle$*, written* $a \prec_\pi b$*, if* $\Gamma \subset \Gamma'$.

Definition 17 (Premise-minimal argument) *An argument* $a \equiv \langle \Gamma, \varphi \rangle$ *is premise minimal, written* $\emptyset \prec_\pi a$*, if for all* $\gamma \in \Gamma : \Gamma - \{\gamma\} \not\vdash \varphi$.

Informally, an argument is premise minimal relative to another if it has fewer inessential premises, and is premise-minimal *simpliciter* if all its grounds are essential.

[12] For illustrative purposes, let p stand for 'John is at home' and q for 'John is at the office'. The argument can then be read idiomatically as 'John is at home, John is at the office, so John is both at home and at the office; therefore, John is at home'. The counterargument to q might be 'John is not at the office (because his office is empty)', whereas the counterargument to $\neg(p \wedge q)$ might be 'John is not both at home and at the office (because these are different locations, and he cannot be in two places at once)'.

3.2 Superfluous Derivation Steps

So far we have not considered the internal structure of arguments, characterising them abstractly as pairs of the form $\langle$*grounds*, *conclusion*$\rangle$. While this is adequate for many cases, the structure of the derivation is also a factor to consider in evaluating the minimality of an argument. Consider the following derivations:

Example 2

1. $$\dfrac{\dfrac{[p]^1}{p \to p}\,(\to I),1 \qquad p}{p}\,(\to E)$$

2. $$\dfrac{\dfrac{p \qquad p}{p \wedge p}\,(\wedge I)}{p}\,(\wedge E)$$

3. p

The derivations in the above example all depend on exactly the same set of (uncancelled) hypotheses, viz. p itself; yet it is clear that the former two contain superfluous steps when compared to the latter. Indeed, derivations exhibiting an arbitrary degree of superfluousness are easy to construct; for example, by repeated introduction of connectives that are immediately eliminated:

Example 3

$$\dfrac{\dfrac{\dfrac{\dfrac{[p]^1}{p \to p}\,(\to I),1}{p \to (p \to p)}\,(\to I) \qquad p}{p \to p}\,(\to E) \qquad p}{p}\,(\to E)$$

Thus there would appear to be some intuitive grounds for distinguishing among derivations not only in terms of the minimality of their hypotheses, but also according to the manner in which the derivation is carried out.

3.3 Cuts and Reduction

We have taken derivations to be tree structures, each of whose nodes is an occurrence of a formula of $\mathfrak{L}$. Intuitively, a formula occurrence is eliminable if it could be omitted while allowing for the derivation of the same conclusion from (at most) the same set of premises.

This criterion can be explicated in terms of a conversion from a derivation which contains a given formula occurrence, to a derivation with the same premises and conclusion which omits that occurrence: if such a conversion can be carried out successfully, then the formula occurrence is superfluous. We shall say that a formula occurrence (token) is *inessential*, or *eliminable*, if it can be cut in this way, and that a formula (type) is non-essential, or eliminable, if all its occurrences are.

We shall assume that the conditions under which conversions between derivations can take place are specified by the Cut metarule (that is, the underlying deductive system is closed under Cut):

Definition 18 (Cut rule)

$$\frac{\Gamma \vdash \gamma \qquad\qquad \Gamma' \cup \{\gamma\} \vdash \varphi}{\Gamma' \cup \Gamma \vdash \varphi}\ \text{Cut}$$

In the particular case where $\Gamma \vdash \gamma$ and $\Gamma \subseteq \Gamma'$, from $\Gamma' \cup \{\gamma\} \vdash \varphi$ we can infer the existence of a derivation $\Gamma' \vdash \varphi$ from which the premise γ has been eliminated.

The cut metarule cannot be expressed as a derivation rule within a natural deduction system.[13] Moreover, in itself it does not provide an algorithm for converting one derivation into another; it merely allows us to infer that such a derivation exists. Of course, given the finiteness of the deductive system, the derivations will be recursively enumerable; so, given closure under Cut, there is a highly trivial procedure for finding a derivation of $\Gamma' \cup \Gamma \vdash \varphi$ given that $\Gamma \vdash \gamma$ and $\Gamma' \cup \{\gamma\} \vdash \varphi$: simply search through all the derivations until we find a derivation of $\Gamma' \cup \Gamma \vdash \varphi$.

We shall refer to the eliminated formula γ as the *cut formula*, the derivation $\Gamma \vdash \gamma$ as the 'ticket', and the step of replacing $\Gamma' \cup \{\gamma\} \vdash \varphi$ by a derivation $\Gamma' \vdash \varphi$ as a *conversion*.

[13]It is a rule *over* derivations, not a rule *within* a derivation. For discussion, see Girard [GLT89], Troelstra and Schwichtenberg [TS00].

Convention 2

- *$\mathcal{D} \dot{\leadsto} \mathcal{D}'$ means that $\mathcal{D}$ converts to $\mathcal{D}'$ (in one step), while $\mathcal{D} \leadsto \mathcal{D}'$ means that there is a finite sequence of conversions $\mathcal{D} = \mathcal{D}_0 \dot{\leadsto} \mathcal{D}_1 \dot{\leadsto} \ldots \dot{\leadsto} \mathcal{D}_{n-1} = \mathcal{D}'$. We let $\mathcal{D} \bar{\leadsto} \mathcal{D}'$ mean that $\mathcal{D} \leadsto \mathcal{D}'$ or $\mathcal{D} = \mathcal{D}'$.*
- *A sequence of conversions is called a* reduction sequence*. A derivation $\mathcal{D}$ is said to be* irreducible *if there is no $\mathcal{D}'$ such that $\mathcal{D} \leadsto \mathcal{D}'$.*

Assuming *some* conversion procedure is available[14], we define relative and absolute notions of derivational minimality in terms of reducibility:

Definition 19 (Relative minimality of derivations) *An argument a is derivation-minimal compared to an argument b, written $a \prec_\delta b$, iff $b \leadsto a$.*

Definition 20 (Derivation-minimal argument) *An argument a is derivation-minimal, written $\emptyset \prec_\delta a$, iff there is no argument b such that $a \dot{\leadsto} b$.*

Informally, an argument a is derivationally minimal relative to an argument b if b is reducible to a, and is derivation-minimal *simpliciter* if it is irreducible.[15]

For deductive systems that conform to the Cut rule, we then have the following basic results concerning minimality:

Proposition 1 *Given an argument $\langle \Gamma \cup \{\gamma\}, \varphi \rangle$ where $\Gamma \vdash \gamma$, then γ is an inessential premise.*

[14]It is clearly preferable to have a *non-trivial* procedure for conversion based on criteria *local* to a given derivation. In systems of natural deduction, there are certain readily identifiable formula occurrences in derivations that can always be cut, namely, those that are both the immediate conclusion of an introduction rule and the major premise of an elimination rule ('IE-formulae'). Basing our conversion procedure on the elimination of IE-formulae would make $\dot{\leadsto}$ non-trivial, and make $\leadsto$ both *well-founded* and *confluent* (or *Church-Rosser*)—for proofs and discussion, see Tennant [Ten78], Girard [GLT89], Troelstra and Schwichtenberg [TS00].

[15]Note that premise minimality is just a special case of derivational minimality: the case where none of the formulae which occur at the leaf nodes of the derivation tree can be completely eliminated.

Proof Trivial, from the definitions of argument, essential premise, and the Cut rule. $\square$

In particular, when γ is a tautology, then $\Gamma \vdash \gamma$ for all Γ, so we have

Corollary 1 *If $\vdash \gamma$, then γ is never essential.*

In other words, a tautology is never required as an essential ground (*uncancelled* premise) of an argument.[16]

3.4 Premise Minimality and Derivational Minimality Contrasted

From the standpoint of both *attacking power* (the range of arguments that an argument can attack) and *robustness* (i.e., susceptibility to attack), arguments with the same grounds and same conclusion are arguably equivalent, regardless of the structure of the derivation. However, arguments that differ in the premises they employ may well be vulnerable to substantially different attacks, even if they have the same conclusion.

Example 4

1. $$\dfrac{\dfrac{p \qquad\qquad q}{p \wedge q}\,(\wedge I)}{p}\,(\wedge E) \qquad \stackrel{\cdot}{\rightsquigarrow} \qquad p$$

2. $$\dfrac{\dfrac{p \qquad\qquad p}{p \wedge p}\,(\wedge I)}{p}\,(\wedge E) \qquad \stackrel{\cdot}{\rightsquigarrow} \qquad p$$

3. p

All three derivations have the same conclusion; the first two are identical in structure, but differ in their grounds, while the last two have identical grounds but differ in structure. All three derivations have the same normal form (indeed, the first two derivations normalise in virtue of precisely the same conversion rule).

Seen as arguments, all the above derivations attack the same class of arguments (in virtue of having the same conclusion). But in terms

[16]The uninformativeness of tautologies is well-known to students of logic; we find it reassuring that our definition accords with this intuition.

of susceptibility to attack, the correct grouping is between 1 and 3, not 1 and 2. Derivation 2 is susceptible to attacks to which both 1 and 3 are not (e.g., on q and $(p \wedge q)$), whereas 1 and 3 are susceptible to *substantially* the same range of attacks.[17]

It is not hard to see why this is so. First, additional tokens of a formula type do not affect the attack properties of an argument; only the absolute presence or absence of a formula type does (an attack on a sentence φ attacks *all* the occurrences of φ). Second, given the assumption that the participants in the argument share the same underlying logic, and that the rules of this logic are not defeasible, it is clear that any attack on the conclusion of any individual step in the derivation can be transformed into an attack on (one or more of) the ultimate premises of that derivation.

Since the identity of an argument is largely, if not completely, a matter of the attack and defeat relations it participates in, this provides a justification of the (common) practice of representing arguments as pairs of the form $\langle \mathit{grounds}, \mathit{conclusion} \rangle$: for it is possible to view structurally different derivations of the same conclusion from the same set of premises as different instances of the *same* argument (unlike derivations of the same conclusion from different sets of premises, which are generally instances of *different* arguments even if they are structurally identical).

Although it is reasonable to conclude from this argument that premise-minimality is the more significant argument-theoretic notion, the notion of derivational minimality is still useful: for example, even if it is true that an attack on $(p \wedge p)$ in derivation 1 can be trivially converted into an attack on p in derivation 3, it is nevertheless not *technically* an attack on p, given our definition of attack. Thus the attack relation recognises a distinction between the two derivations that premise-minimality ignores. Derivational minimality, on the other hand, being a more fine-grained notion than premise-minimality, does provide a way of distinguishing between 1 and 3.

4 Relevance

It is common to apply the term '(ir)relevant' to both arguments (e.g., "The prosecution's argument is totally irrelevant") and sentences (e.g.,

[17] Argument 1 is technically vulnerable to an attack on $p \wedge p$, not just an attack on p; however, any attack on $p \wedge p$ can be trivially transformed into an attack on p.

"That statement was relevant"). Since our principal concern is with relevance to an argumentation topic, we shall concentrate on two relevance relations in particular: argument-topic relevance and sentence-topic relevance. We shall motivate and define each notion separately, and then examine how they are related.

The notation $\rho_\Delta(x, y)$ is to be read 'x is relevant to y relative to Δ', where x is either a sentence or an argument, y is a topic, and the subscript Δ denotes a database.[18]

4.1 Argument-Topic Relevance

Intuitively, an argument that directly addresses the topic of argumentation (that is, one that answers the corresponding query) is relevant to that topic. Hence, as a first approximation, we take the relevance relation between an argument and a topic to be based on the relation of directly addressing the topic (definition 8 above):

Definition 21 (Direct relevance of argument to topic) *An argument a is directly relevant to a topic $?\psi$ relative to a database Δ, written $\rho_\Delta(a, ?\psi)$, if*

1. *$\Delta \Vdash a$; and*
2. *a is not overruled; and*
3. *a directly addresses $?\psi$.*

Note that this definition allows an argument to contain superfluous premises and/or derivation steps and yet be considered relevant; this is because it seems excessive to say that an argument that satisfies the topic of the argument is nevertheless completely irrelevant, simply on the grounds that it contains a single superfluous premise or derivation step.

But just as it seems excessive to say that an argument containing a single superfluous premise or derivation step is completely irrelevant, it is also seems excessive to make no distinction whatsoever,

[18]Although we shall employ the same symbol, ρ, for both argument-topic relevance and sentence-topic relevance, it should be noted that these are distinct relations which might be more scrupulously represented by ρ^{arg}, ρ^{sent}, for example. Confusion is unlikely to arise, however, as the identity of the relation can easily be determined from the type of the first argument, so we shall persist in our chosen usage in the interests of avoiding notational overload.

relevance-wise, between arguments that contain any number of superfluous premises (derivation steps) and those that do not.

One way to accommodate both intuitions is to acknowledge varying degrees of relevance, such that relevance is not simply a classificatory concept, but also a comparative one.[19]

4.2 Degrees of Relevance

In defining the boundary between the relevant and the irrelevant, we risk going wrong in two distinct ways: the demarcation may be too strict, denying the existence of borderline cases, or it may be too vague, resulting in a messy border. The ideal criterion is one that is sharp, but which can also be relaxed in a clear and precise manner.[20]

One way to order arguments with the same conclusion with regard to relevance is in terms of their relative minimality. Thus, instead of saying that the argument $\langle\{p, p \to p\}, p\rangle$ is completely irrelevant to the topic $?p$ by virtue of containing a single superfluous premise, we might instead concede that it is relevant, but less relevant than the argument $\langle\{p\}, p\rangle$. Similar reasoning applies to the case of arguments containing superfluous derivation steps.

We have seen that premise minimality is arguably the more appropriate notion for providing identity conditions for arguments; however, we shall employ the more general notion of derivational minimality in our definition of a (partial) relevance ordering on arguments. The reason for this is that it provides a notion of eliminability that is more attuned to the level of granularity of the attack relation:

Definition 22 (Comparative topic relevance of arguments) *Given relevant arguments $a \equiv \langle\Gamma, \varphi\rangle$ and $b \equiv \langle\Gamma', \varphi\rangle$, we say that a is more relevant than b, written $a \succ_R b$, if $a \prec_\delta b$.*

Definition 23 (Maximal topic relevance) *An argument a is* maximally topic relevant *to a query $?\psi$ relative to a database Δ if and only if*

1. *a is relevant to $?\psi$ relative to Δ; and*
2. *a is a minimal argument.*

[19] Cf. Sperber and Wilson [SW95, p. 123ff.].

[20] Similar desiderata apply to criteria for, say, formal definitions of grammaticality: a good criterion of 'grammatical sentence' is one that provides a natural measure of degrees of grammaticality.

4.3 Sentence-Topic Relevance

Consider a topic $?\varphi$ and database Δ, and arbitrary $\gamma \in \Delta$. We let φ^* stand for a sentence φ or its complement $\bar{\varphi}$.

Clearly, if $\gamma \in \{\varphi, \bar{\varphi}\}$, then it is relevant, as it directly addresses the topic. We do not need to consider any other sentences in Δ.

When $\gamma \notin \{\varphi, \bar{\varphi}\}$, however, then there may be cases where the relevance of γ depends on what other sentences are contained in Δ.

For example, if the topic is $?q$, γ is p, and $\Delta = \{p, r\}$, then there is no relation of relevance between p and $?q$. On the other hand, relative to $\Delta' = \{p, p \to q\}$, p is relevant to the topic (since we can construct an argument for q); similarly for the database $\Delta'' = \{p, p \to \neg q\}$. So, given a query $?\varphi$, γ is made relevant by the presence of a formula of the form $\gamma \to \varphi^*$. E.g., if $?\varphi$ is *Is John at home?*, then a sentence such as *John's car is in the driveway* constitutes a relevant item of information given a conditional such as *If his car is in the driveway, then John is at home.*

In these cases, the determining factor of whether a sentence is relevant to the topic of an argument is its potential for occurring as part of an argument that satisfies the query which that topic corresponds to. We can generalise this as follows:[21]

Definition 24 (Relevance of sentence to topic) *A sentence γ is relevant to topic $?\psi$, written $\rho_\Delta(\gamma, ?\psi)$, if and only if there exists a non-overruled argument* $\begin{matrix}\mathcal{D}\\ \varphi\end{matrix}$ *such that* $\Delta \Vdash \begin{matrix}\mathcal{D}\\ \varphi\end{matrix}$ *and*

1. $\begin{matrix}\mathcal{D}\\ \varphi\end{matrix}$ *directly addresses $?\psi$; and*

2. γ occurs in $\begin{matrix}\mathcal{D}\\ \varphi\end{matrix}$ *; and*

3. γ is not eliminable in $\begin{matrix}\mathcal{D}\\ \varphi\end{matrix}$ *.*

That is, a sentence γ is relevant to a topic $?\psi$ relative to database Δ iff γ is an *essential* (ineliminable) formula in *some* argument for ψ or its contrary, ψ, constructible from that database.

The notions of sentence-topic relevance and argument-topic relevance are related by the following result concerning minimal arguments:

[21] This definition, like definition 21 above, is provisional, for reasons that will be made clear in sections 5 and 6.

Proposition 2 *Given topic* $?\varphi$*, database* Δ*, and minimal argument* a *such that* $\rho_\Delta(a, ?\varphi)$*, then* $\rho_\Delta(\gamma, ?\varphi)$ *for all* γ *occurring in* a*.*

Proof Trivial. Given a minimal argument a such that $\rho_\Delta(a, ?\varphi)$, it follows by the definition of minimality that there is no argument x such that $a \dot{\rightsquigarrow} x$. Consider now an arbitrary sentence γ occurring in a: if γ is eliminable, then there is an argument b such that $a \dot{\rightsquigarrow} b$, contradicting the minimality of a; so γ is not eliminable. But then we have $\rho_\Delta(\gamma, ?\varphi)$, since a satisfies $?\varphi$, and as we are arbitrary in γ, this holds for all γ occurring in a. $\square$

It should be noted, however, that the converse does not hold; that is, given $\rho_\Delta(\gamma, ?\varphi)$ for all γ occurring in a, it does *not* necessarily follow that $\rho_\Delta(a, ?\varphi)$. This is simply because the relevance of each of the γ may be due to their essential role in some argument *other* than a itself:

Example 5 Consider a database $\Delta = \{p, q, p \to q\}$ and a topic $?q$. Then arguments a and b are both relevant, supported by Δ, and moreover minimal:

a q

b $\dfrac{p \to q \quad p}{q}\ (\to E)$

Thus we have, in particular, $\rho_\Delta(q, ?q)$, $\rho_\Delta(p, ?q)$ and $\rho_\Delta(p \to q, ?q)$. But we also have the following derivation, call it c:

c $\dfrac{q}{p \to q}\ (\to I)$

Then $\Delta \Vdash c$, and c is minimal; moreover, every sentence occurring in c is relevant to $?q$ (q itself in virtue of argument a, $p \to q$ in virtue of argument b). However, argument c itself is not directly relevant to the topic.

5 Propagation of Relevance

The definition of argument-topic relevance we have employed thus far is deficient. For a start, it is *incomplete*: it only covers cases of direct relevance, that is, cases where the argument under consideration directly addresses the topic of argumentation; so while it may describe *sufficient* conditions for an argument to be relevant to a topic, it certainly does not provide *necessary* ones. Given an argument a that is relevant to the topic of argumentation, other arguments may become relevant as a result which do not fall under the definition as given: first, the proper subarguments of a relevant argument are potentially topic-relevant in virtue of their contribution to addressing the topic of argumentation, though they do not themselves directly address the topic; second, counterarguments are also candidates for being relevant, as they may affect the status of a topic-relevant argument.

Our aim in this section is to characterise this set of 'indirectly' relevant arguments, by providing an account of how topic-relevance propagates through an argument system (to subarguments and to counterarguments). However, consideration of this issue will also bring to the fore an aspect of the definition which is incorrect (or at least *inexact*), and which will prompt us to revise it in section 6 below.

First, note that, given a relevant argument a and an argument b such that b attacks (and perhaps even defeats) a, argument b may be relevant to the topic even if it does not address it directly. For example, let p stand for *John's car is at the office*, q for *John's car is at his house*, r for *John is at home*; let the topic be $?r$, i.e., the issue of whether John is at home, and suppose that we have database $\Delta = \{p, q, p \to \neg q, q \to r\}$. Then we have the following arguments, denoted by a and b:

$$\boldsymbol{a} \quad \frac{q \to r \qquad q}{r} \; (\to E)$$

$$\boldsymbol{b} \quad \frac{p \to \neg q \qquad p}{\neg q} \; (\to E)$$

Now b attacks a, and potentially undermines the answer to the query that a provides; intuitively, b certainly counts as relevant because it has the potential to affect the outcome of the argument. However, b itself does not address the topic directly, in that it does not itself settle the issue of whether r is the case; rather, b derives its topic relevance from the fact that it attacks a topic-relevant argument, namely a. And if we had an argument c that attacks b (thus potentially reinstating a), then c would be topic-relevant as well, in virtue of attacking an attacker of a (potentially reinstating a).

Thus, we have here a simple example of topic relevance propagating 'backwards' along the attack relation (from the attacked to the attacker). It would be convenient if the standard attack relation provided a straightforward guarantee of propagation; call this the *simple propagation model*. According to this model, once we have determined the set of arguments that are directly relevant to the topic (by the standard definition of attack), topic relevance then propagates to all arguments that attack any member of this set. Unfortunately, this is not the case: not all attacks on topic-relevant arguments are themselves topic-relevant, as we shall see.

5.1 Relevant Attacks

Recall that the notion of attack in concrete argumentation systems (as opposed to abstract ones, in which the attack relation is taken as given) is standardly construed in terms of a contradiction in the underlying logic: a attacks b if the conclusion of a contradicts some sentence occurring in b. Our definition of topic relevance, on the other hand, tolerates the presence of superfluous premises and subarguments; we justified this on the grounds that it would be excessive to classify an argument as irrelevant on the basis of a single superfluous subargument, proposing instead a partial ordering on relevant arguments based on a (proof-theoretic) notion of relative minimality. But this means that, given a topic-relevant argument, we do not have a guarantee that all of its subarguments are relevant; and if not all of its subarguments are relevant, then neither are all of its counterarguments, since an attack may target precisely such a superfluous subargument.

Example 6 Consider arguments a and b in the context of a database $\Delta = \{p, q, \neg q\}$ and topic $?p$:

a $$\dfrac{\dfrac{p \qquad\qquad q}{p \wedge q}\,(\wedge I)}{p}\,(\wedge E)$$

b $\neg q$

Then a is topic relevant. Moreover, $\Delta \Vdash b$ and $b \longrightarrow a$. However, b can hardly be considered relevant to the topic, even though it technically undermines argument a, for the simple reason that q is completely superfluous.

Indeed, if the attacking arguments themselves contain superfluous subarguments, then the chain of arguments may lead further and further into irrelevance (see figure 1). There is a need, therefore, to distinguish between attacks on the basis of the topic-relevance of the subarguments they target. Given that the problem of propagation arises from the presence of superfluous elements in the derivation, we start by defining the notion of an essential subargument:

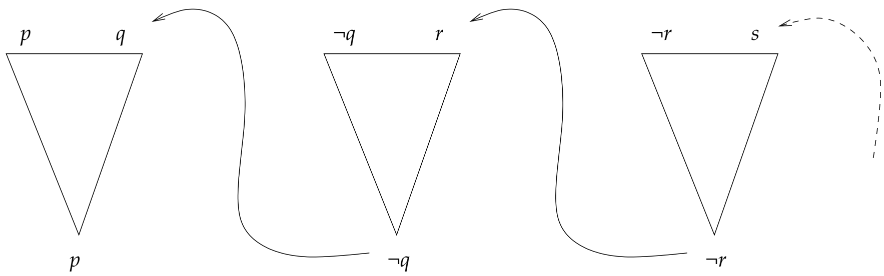

Figure 1: chain of irrelevant attacks

Definition 25 (Essential subargument) *Let a be an argument. Then a subargument $a' \sqsubseteq a$ is an* essential subargument *of a, written $a' \sqsubseteq_\rho a$, iff the conclusion of a' is ineliminable in a.*

We then define the notion of topic-relevant attack as:

Definition 26 (Relevant attack) *Let a be an argument. Then b* relevantly attacks *a, written $b \xrightarrow{\rho} a$, iff*

1. *$b \longrightarrow a'$; where*
2. *a' is an essential subargument of a.*

That is, a relevant attack on a is an attack on an essential subargument of a. We can now define the notions of relevant subargument and relevant counterargument directly in terms of essential subargument and relevant attack:

Definition 27 (Relevant subargument) *Let a be an argument, Δ a database, and $?\varphi$ a topic. Then*

$$\rho_\Delta(a, ?\varphi) \text{ and } b \sqsubseteq_\rho a \Longrightarrow \rho_\Delta(b, ?\varphi).$$

Definition 28 (Relevant counterargument) *Let a be an argument, Δ a database, and $?\varphi$ a topic. such that . Then*

$$\rho_\Delta(a, ?\varphi) \text{ and } b \xrightarrow{\rho} a \Longrightarrow \rho_\Delta(b, ?\varphi).$$

In other words, an essential subargument of a topic-relevant argument is itself topic-relevant, as is an argument that relevantly attacks a topic-relevant argument.

6 Relevant Argument Systems

Given the notion of relevant attack, $\xrightarrow{\rho}$, we can define a *relevant argument system* by specifying the remaining components that together make up an argumentation system, viz., the defeat relation, and the notions of overruled and justified argument.

We extend the notion of relevance to the defeat relation as follows:

Definition 29 (Relevant defeat) *Argument a* relevantly defeats *argument b, written $a \succeq\!\!\!\!\succ_\rho b$, iff a relevantly attacks b and is not weaker than b.*

Definition 30 (Strict relevant defeat) *Argument a* strictly and relevantly defeats *argument b, written $a \succ\!\!\!\succ_\rho b$, iff a relevantly attacks b and a is stronger than b.*

We can also refine our notion of self-defeat:

Definition 31 (Relevant self-defeat) *An argument $\langle\Gamma, \varphi\rangle$ is* relevantly self-defeating *iff $\Gamma' \vdash \alpha, \alpha$ for some $\alpha \in \mathfrak{L}$, where $\Gamma' \subseteq \Gamma$ is the set of essential premises of $\langle\Gamma, \varphi\rangle$.*

That is, an argument is relevantly self-defeating if its set of essential premises is not conflict-free.

Definition 32 (Relevantly overruled arguments) *An argument is* relevantly overruled *if it is not relevantly justified and either it is relevantly self-defeating, or it or one of its proper subarguments is relevantly defeated by a relevantly justified argument.*

Definition 33 (Relevantly justified arguments) *An argument a is* relevantly justified *iff*

1. *a is not relevantly self-defeating; and*
2. *All proper subarguments of a are relevantly justified; and*
3. *All arguments defeating a are relevantly self-defeating, or have at least one proper subargument that is relevantly overruled.*

These defitions are such that a given argument may be self-defeating, but not *relevantly* self-defeating, or overruled (defeated) but not *relevantly* overruled (defeated). The argument in figure 2, for example, is self-defeating because $\{p, q, \neg q\} \vdash q, \neg q$; however, it is not *relevantly* self-defeating, because the essential premise set is $\{p\}$, and $p \nvdash \alpha, \neg\alpha$ for any $\alpha \in \mathfrak{L}$. Clearly, it is relevant self-defeat that is of greater

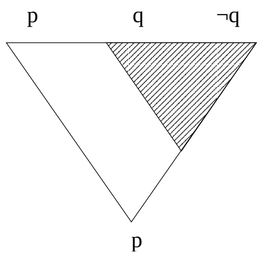

Figure 2: self-defeating but not relevantly self-defeating argument

consequence, for a self-defeating but not relevantly self-defeating argument can be salvaged (e.g., by eliminating the offending superfluous premises). The notion of topic-relevant attack thus enables us to distinguish two different notions of defeat, one of which is clearly of greater significance than the other.

The final issue we need to address is the reference to the notion of overruled argument in definitions 21 and 24 for topic relevance (of arguments and sentences respectively). In order to bring those definitions into line with our notions of relevant argumentation, it remains to substitute 'relevantly overruled' for 'overruled' throughout:

Definition 34 (Revised version of definition 21) *An argument a is directly relevant to a topic $?\psi$ relative to a database Δ, written $\rho_\Delta(a, ?\psi)$, if*

1. *$\Delta \Vdash a$; and*

2. *a is not* relevantly *overruled; and*

3. *a directly addresses* $?\psi$.

Definition 35 (Revised version of definition 24) *A sentence* γ *is relevant to topic* $?\psi$*, written* $\rho_\Delta(\gamma, ?\psi)$*, if and only if there exists a* *non*-relevantly-*overruled argument* $\begin{array}{c}\mathcal{D}\\ \varphi\end{array}$ *such that* $\Delta \Vdash \begin{array}{c}\mathcal{D}\\ \varphi\end{array}$ *and*

1. $\begin{array}{c}\mathcal{D}\\ \varphi\end{array}$ *directly addresses* $?\psi$*; and*

2. γ *occurs in* $\begin{array}{c}\mathcal{D}\\ \varphi\end{array}$ *; and*

3. γ *is not eliminable in* $\begin{array}{c}\mathcal{D}\\ \varphi\end{array}$ *.*

6.1 Propagation Among Minimal Arguments

Are there any conditions under which the simple propagation model suffices?

The answer is yes: if we restrict ourselves to maximally relevant arguments, then *all* attacks preserve topic relevance, and the distinction between attacks and relevant attacks collapses (and consequently the distinction between defeat and relevant defeat).

It is easy to see why. We defined maximally topic-relevant arguments as those which are derivationally minimal. So if an argument a is maximally relevant, then every sentence occurring in a is essential (i.e., ineliminable), whence it follows that every one of its subarguments is essential (since the conclusion of any such subargument must be essential). Thus every attack on a is *a fortiori* an attack on an essential subargument of a (possibly the conclusion of a itself). Hence every attack is a relevant attack.

Thus we have the following:

Proposition 3 *Given arguments* a, b *where* a *is maximally relevant, then* $b \xrightarrow{\rho} a$ *iff* $b \longrightarrow a$.

Thus there are considerable benefits to a system composed only of minimal arguments. Arguments will be maximally robust, and maximally relevant. Only the simple propagation model is required, since the standard attack relation preserves topic-relevance.

However, this is easier to demand in theory than in practice, if only because in complex arguments superfluous premises may creep in undetected.[22] If our goal is to model the kinds of argumentation that occur in moderately realistic settings (e.g., legal argumentation), then we should arguably prefer a definition that can accommodate less-than-maximally-relevant arguments, while recognising that these fall short of the ideal.

7 Paradoxes of Material Implication

We now consider how our argument-theoretic definition of relevance handles the paradoxes of material implication (we assume the underlying logic is classical, and '$\rightarrow$' is interpreted as material implication).

Theorem 1 (Paradoxes of material implication)

1. $\neg\gamma \vdash \gamma \rightarrow \varphi$
2. $\varphi \vdash \gamma \rightarrow \varphi$
3. $\vdash (\gamma \wedge \neg\gamma) \rightarrow \varphi$
4. $\vdash \gamma \rightarrow (\varphi \vee \neg\varphi)$

Suppose we have a conditional ($\gamma \rightarrow \varphi$) and a topic $?\varphi$. From an informational standpoint, it can be argued that if we know that the consequent is true, then the conditional becomes less relevant, because any application of the conditional will not enable us to derive any more information regarding ψ than we already possess: that is, knowing the truth value of the consequent *supresses* the relevance of the conditional. For example, if I already *know* with certainty that John is at home, then the conditional *If his car is in the driveway, then John is at home* is no longer relevant.

Moreover, if we do not know that the antecedent φ is true (and especially if we know that the antecedent is *false*), then any conditional in which φ occurs as the antecedent is less relevant: the conditional may be true, but we cannot apply it. For example, if I know that John's car is not in the driveway, the conditional *If his car is in the driveway, then John is at home* is not relevant to me: the conditional supplies no additional knowledge. Therefore, not knowing that the

[22] Indeed, even in a highly idealised model, ensuring argument minimality may turn out to be prohibitively expensive computationally (Francesca Toni, personal communication).

antecedent of the conditional is true (or knowing that it is false) also suppresses the relevance of the conditional.

The point regarding material conditionals with contingently false antecedents or true consequents is this: such conditionals are irrelevant, because they cannot be applied to yield non-trivial information about their consequents. In the case of strict conditionals with *necessarily* false antecedents or true consequents, the situation is even starker: such conditionals can *never* be applied to give new knowledge. The conditionals may be true, but they are informationally sterile.

Suppose we are interested in a topic $?\varphi$, and have a database Δ such that $\neg\gamma \in \Delta$. Then we have an argument

$$a \quad \equiv \quad \cfrac{\cfrac{\cfrac{\neg\gamma \qquad\qquad [\gamma]^1}{\bot}\ \bot I}{\varphi}\ EFSQ}{\gamma \to \varphi}\ (\to I),1$$

such that $\Delta \Vdash a$, which we abbreviate to

$$a \quad \equiv \quad \begin{array}{c} \neg\gamma \\ \mathcal{D} \\ \gamma \to \varphi \end{array}$$

Given that φ occurs as the consequent of the conditional, we might try to extend the derivation into an argument a' for φ. But in order to detach the consequent, we require the antecedent as a minor premise, so a' would then have to be of the general form

$$\cfrac{\begin{array}{c} \neg\gamma \\ \mathcal{D} \\ \gamma \to \varphi \end{array} \qquad \begin{array}{c} \mathcal{D}' \\ \gamma \end{array}}{\varphi}\ (\to E)$$

which is self-defeating, since $\mathrm{prem}(a') \vdash \neg\gamma, \gamma$. Thus a' is not relevant.

Consider, on the other hand, a database Δ such that $\varphi \in \Delta$. Then we have an argument

$$a \quad \equiv \quad \dfrac{\varphi}{\gamma \to \varphi}\ (\to I)$$

such that $\Delta \Vdash a$. Given that φ occurs as the consequent of the conditional, we might try to extend the derivation into an argument a' for φ. Again, in order to detach the consequent, we require the antecedent as a minor premise, so a' would then have the general form

$$\dfrac{\dfrac{\varphi}{\gamma \to \varphi}\ (\to I) \qquad \begin{array}{c}\mathcal{D}\\ \gamma\end{array}}{\varphi}\ (\to E)$$

which is consistent, but can be objected to on the grounds that it is not minimal (since we already had $\varphi \in \Delta$), even though the argument as a whole is still relevant by our definition.[23] Note that in any case the *sentence* $(\gamma \to \varphi)$ is not relevant, as it is eliminable via reduction.

Also of interest, because of their connection to circular patterns of argumentation, are theorems of the form $(\varphi \to \varphi)$ of the form, which are valid not only classically, but also in the 'mainstream' relevantist systems **R** and **E**:[24]

Example 7 Consider the following database Δ relative to topic $?\varphi$:

$$\begin{array}{l}\varphi \to \varphi\\ \varphi\end{array}$$

Then $\Delta \Vdash \langle\{\varphi, \varphi \to \varphi\}, \varphi\rangle$, so $\langle\{\varphi, \varphi \to \varphi\}, \varphi\rangle$ is relevant, but we also have $\Delta \Vdash \langle\{\varphi\}, \varphi\rangle$ where $\langle\{\varphi\}, \varphi\rangle \succ_R \langle\{\varphi, \varphi \to \varphi\}, \varphi\rangle$, so $\langle\{\varphi, \varphi \to \varphi\}, \varphi\rangle$ is not maximally relevant. $\varphi \to \varphi$ is eliminable from $\langle\{\varphi, \varphi \to \varphi\}, \varphi\rangle$, and since there are no other non-self-defeating arguments for (or against) φ supported by Δ (unless φ is a tautology, in which case it is inferrable without premises), it follows that $\varphi \to \varphi$ is eliminable from all arguments relevant to $?\varphi$, hence $\varphi \to \varphi$ is not relevant to $?\varphi$.

Alternatively, suppose that Δ' is the database consisting solely of

[23] However, if we were to adopt a strict criterion of relevance based on absolute minimality, then the argument would be irrelevant.

[24] Cf. Anderson and Belnap [AB75].

$$\varphi \rightarrow \varphi$$

Then $(\varphi \rightarrow \varphi)$ is again not relevant to $?\varphi$, in this case because the antecedent is not a member of Δ' (if it should happen that the antecedent, while not explicitly a member of Δ', is itself a tautology, then it would be inferrable without the aid of $(\varphi \rightarrow \varphi)$).

Suppose, finally, that our database, call it Δ'', happens to contain

$$(\varphi \rightarrow \varphi) \rightarrow \varphi$$
$$\varphi \rightarrow \varphi$$

Here again, since $\vdash \varphi \rightarrow \varphi$, we have $(\varphi \rightarrow \varphi) \rightarrow \varphi \vdash \varphi$, so $\varphi \rightarrow \varphi$ is eliminable from the premises.

Our definition classifies $(\varphi \rightarrow \varphi)$ as irrelevant to the topic $?\varphi$; therefore, an argument in which it occurs is less than maximally relevant.[25] However, the *validity* of $(\varphi \rightarrow \varphi)$ is not questioned. Thus we have a way of characterising circular arguments as ill-formed that respects the underlying (classical) logic. By contrast, the relevance logic **S** of Meyer, Martin and Dwyer avoids circular arguments, but at the cost of excluding all theorems of the form $(\varphi \rightarrow \varphi)$.[26]

The point to note here is that irrelevance in no way implies logical invalidity:[27] statements may be logically valid but irrelevant. Most relevance logics instead modify the logical system precisely in order to deny logical validity to 'paradoxical' classical tautologies such as $(\perp \rightarrow \varphi)$ or $(\varphi \rightarrow \top)$.

8 Related Work

The general idea that communication is governed by two opposing principles enjoining informational sufficiency and linguistic economy has a long pedigree in linguistics and philosophy.[28] The idea that relevance itself can be defined in terms of these two antinomic aspects

[25] And if we were to adopt a strict criterion of relevance, circular arguments would be irrelevant *simpliciter*.

[26] Cf. Meyer *et al.* [MMD83].

[27] Nor, indeed, does logical validity imply relevance.

[28] See the first section of Horn [Hor93] for a brief discussion of the history and influence of this idea in linguistic theory.

has been proposed by Cooper [Coo71] in the context of information retrieval, and by Sperber and Wilson [SW95] in cognitive linguistics.[29] The accounts differ in how they interpret 'informativeness' and 'economy': Cooper models informativeness in terms of satisfying a user's informational needs (as manifested in a query to the information retrieval system), whereas Sperber and Wilson relate informativeness to the number of 'cognitive effects' an utterance gives rise to (modelled as a special kind of context-dependent logical inference); Cooper's notion of economy is one of *form* (like ours, it is based on minimality of proof), whereas Sperber and Wilson's is one of *process* (economy of cognitive effort).

The definition we have presented in this paper is closer to Cooper's; for example, our characterisation of topic relevance for sentences is very similar to the definition of relevance among sentences proposed in Cooper [Coo71], where the relevance of a sentence φ to another sentence ψ is analysed in terms of it making an ineliminable contribution to proving ψ or its negation $\neg\psi$. Our account differs sharply from Cooper's, however, in acknowledging degrees of relevance, which Cooper explicitly rejects.[30]

Our pragmatic response to the paradoxes of implication is inspired by the defence of the material conditional due to Grice [Gri89] and others (e.g. Fogelin [Fog78]). However, Grice's theory of conversation makes appeal to an unexplicated notion of relevance, a fact that makes the account potentially question-begging vis-à-vis relevance logic; as some relevance logicians have noted,[31] relevance logic can be seen as an attempt to explicate the very notion that Grice takes for granted. It is interesting that the post-Gricean theory of Sperber and Wilson, which *does* explicitly address the issue of the inferential mechanism underlying relevance, proposes a logical mechanism (the 'deductive device') that is even more restrictive than relevance logic, as it excludes *all* introduction rules for logical connectives, as well as EFSQ.[32] By contrast, the account of argument-theoretic relevance we have presented

[29] A concern with economy can also be discerned in relevance logic, where it manifests itself as the requirement that all premises be 'genuinely used' in the course of the proof—see Anderson and Belnap [AB75].

[30] Moreover, Cooper's definition is vulnerable to the paradoxes of implication (which he does not discuss).

[31] Cf. Dunn [Dun83].

[32] For a description of the deductive device and arguments for its psychological plausibility, see Sperber and Wilson [SW95, pp. 93–103].

avoids the question-begging aspect of Grice's theory, while allowing the underlying logic to be classical.

9 Conclusion

We have provided a basic definition for a relevant argument system, comprising

- a definition of topic-relevance for arguments;
- a definition of topic-relevance for sentences;
- an account of how relevance propagates in an argument system;
- a definition of the relation of relevant attack (and therefore relevant defeat).
- a characterisation of a relevant argument system.

The ultimate basis for our definition is a variant of the idea that communication and cognition are governed by a trade-off between opposing demands of informational sufficiency and economy of means; we have modelled the notion of informational sufficiency in terms of satisfying a query associated with a topic of argumentation, while we have based the notion of economy on that of proof-theoretic minimality. The resulting system of relevant argumentation is able to handle fallacies of relevance, such as the paradoxes of implication, using a classical rather than a relevance logic as the underlying deductive system.[33] The validity of the paradoxes is not challenged; rather, their unacceptability stems from the fact that they represent argumentational dead-ends: any attempt to expand the derivation of the 'paradoxical' implication into an argument for the consequent results in an argument that is either self-defeating or redundant.

References

[AB75] A. R. Anderson and N. D. Belnap. *Entailment: The Logic of Relevance and Necessity*, volume 1. Princeton University Press, Princeton, NJ, 1975.

[33] Note that the metalogical character of the definition means that it is compatible with a range of logical systems—including relevance logics.

[ABD92] A. R. Anderson, N. D. Belnap, and J. M. Dunn. *Entailment: The Logic of Relevance and Necessity*, volume 2. Princeton University Press, Princeton, NJ, 1992.

[Ack56] W. Ackermann. Begründung einer strengen Implikation. *The Journal of Symbolic Logic*, 21:113–128, 1956.

[Ack58] W. Ackermann. Über die Beziehung zwischen strikter und strengen Implikation. *Dialectica*, 12:213–222, 1958.

[Chu51] A. Church. The weak theory of implication. In Menne-Wilhelmy-Angsil, editor, *Untersuchungen zum Logikkalkül und der Logik der Einzelwissenschaften*, pages 22–37. Kommissions-Verlag Karl Alber, Munich, 1951.

[Coo71] W. S. Cooper. A definition of relevance for information retrieval. *Information Storage and Retrieval*, 7:19–37, 1971.

[Dun83] J. M. Dunn. Relevance logic and entailment. In F. Guenthner and D. Gabbay, editors, *Handbook of Philosophical Logic*, volume 3. Reidel, Dordrecht, 1983.

[Dun95] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[FKA92] J. Fox, P. Krause, and S. Ambler. Arguments, contradictions and practical reasoning. In *Proceedings of the 10th European Conference on Artificial Intelligence (ECAI-92)*, pages 623–627, Vienna, Austria, 1992.

[Fog78] R. J. Fogelin. *Understanding Arguments*. Harcourt, Brace, Jovanovich, New York, 1978.

[GLT89] J-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, 1989.

[Gri89] P. Grice. *Studies in the Ways of Words*. Harvard University Press, Cambridge, Mass., 1989.

[Haa78] S. Haack. *Philosophy of Logics*. Cambridge University Press, Cambridge, 1978.

[Hor93] L. Horn. Economy and redundancy in a dualistic model of natural language. In S. Shore and M. Vilkuna, editors, *1993 Yearbook of the Linguistic Association of Finland*, pages 33–72. 1993.

[K^{+}95] P. Krause et al. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11:113–131, 1995.

[KT96] R. A. Kowalski and F. Toni. Abstract argumentation. *Artificial Intelligence and Law*, 4:275–296, 1996.

[MMD83] R.K. Meyer, E.P. Martin, and R. Dwyer. The fundamental S-theorem—a corollary. *Notre Dame Journal of Formal Logic*, 24:509–516, 1983.

[Moh50] Shaw-Kwei Moh. The deduction theorems and two new logical systems. *Methodos*, 2:56–75, 1950.

[Pra65] D. Prawitz. *Natural Deduction: A Proof Theoretical Study.* Stockholm, 1965.

[PV01] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic.* Kluwer Academic, Boston, MA, 2nd edition, 2001.

[SW95] D. Sperber and D. Wilson. *Relevance: Communication and Cognition.* Blackwell, Oxford, 2nd edition, 1995. First edition published in 1986.

[Ten78] N. Tennant. *Natural Logic.* Edinburgh University Press, Edinburgh, 1978. Reprinted with corrections in 1990.

[TS00] A. S. Troelstra and H. Schwichtenberg. *Basic Proof Theory.* Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2nd edition, 2000.

[vD94] D. van Dalen. *Logic and Structure.* Springer, 3rd edition, 1994.

[Vre97] G. A. W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.