An Argument-based Approach to Reasoning with Clinical Knowledge

Nikos Gorogiannis[a], Anthony Hunter[a], Matthew Williams[a,b]

[a]*Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK*
[b]*Mount Vernon Hospital, Northwood, Middlesex, HA6 2RN, UK*

Abstract

Better use of biomedical knowledge is an increasingly pressing concern for tackling challenging diseases and for generally improving the quality of healthcare. The quantity of biomedical knowledge is enormous and it is rapidly increasing. Furthermore, in many areas it is incomplete and inconsistent. The development of techniques for representing and reasoning with biomedical knowledge is therefore a timely and potentially valuable goal. In this paper, we focus on an important and common type of biomedical knowledge that has been obtained from clinical trials and studies. We aim for (1) a simple language for representing the results of clinical trials and studies; (2) transparent reasoning with that knowledge that is intuitive and understandable to users; and (3) simple computation mechanisms with this knowledge in order to facilitate the development of viable implementations. Our approach is to propose a logical language that is tailored to the needs of representing and reasoning with the results of clinical trials and studies. Using this logical language, we generate arguments and counterarguments for the relative merits of treatments. In this way, the incompleteness and inconsistency in the knowledge is analysed via argumentation. In addition to motivating and formalising the logical and argumentation aspects of the framework, we provide algorithms and computational complexity results.

Key words: knowledge representation, argumentation, inconsistency, clinical knowledge, biomedical knowledge

1. Introduction

Within many scientific fields, especially those employing imprecise measurements, statistics or imperfect modelling, it is common for incomplete and inconsistent knowledge to arise. Medicine and more specifically the provision of healthcare, is a prime example of such a field with the added complication that knowledge is discovered continuously and in very high volume. For example, according to Index Medicus [15], within the last year there were approximately 1200 peer-reviewed publications in the

form of reviews, meta-analyses or reports on clinical trials, on the subject of breast cancer. This information is, of course, invaluable in medical practice, and therefore, practitioners would have to read and understand huge volumes of information. Since this is clearly not practical, experts try to distill and reconcile the state-of-the-art published results into a definitive and coherent form, producing *systematic reviews* and *meta-analyses*. However, these necessarily suffer from the same weaknesses that primary publications do, i.e., that reviews and even more so meta-analyses take time to perform and therefore lag behind the state-of-the-art they are intended to capture; and that they do not scale well with the increasing rate of publication of new results, as they require a substantial amount of painstaking work by experts. For these reasons, systems that help with summarising and analysing such volumes of information in an efficient way, are appealing. Research on constructing and studying such systems, which we will loosely call *decision support systems*, has been ongoing and several systems are being developed [16]. However, none of these address the particular area of interest in this paper or the requirements we set out later in this section.

The area that our paper focuses on is the knowledge reported by clinical trials and studies, two of the primary methods for generating new clinical knowledge. Such publications form a significant input to the process of delivering healthcare, either through the setting of clinical guidelines or through directly informing the practitioner who makes healthcare-related decisions. The adoption of *evidence-based medicine*, i.e., the increasing requirement on guiding medical practice on the basis of the best available evidence of clinical effectiveness, makes the task of absorbing and understanding the literature even more pressing for clinicians. To our knowledge few, if any, decision support systems provide help in this area. Our contribution is an attempt to formalise some of the problems encountered in this process and to suggest a solution in the form of a framework for reasoning about clinical studies, and for constructing arguments for and against the use of treatments with specific patient classes and in relation to specific clinical outcome indicators.

The basic unit of knowledge in our framework is the result of a clinical study or trial. To elucidate what such a result represents, we describe in general terms how a typical (two-arm randomised) clinical trial comparing two treatments is performed. A group of people conforming to well-defined entry criteria (e.g., premenopausal women with post-operative early-stage breast cancer) is sampled and that sample is randomly split into two roughly equal groups. Group one is administered treatment T_1 which is usually either a form of placebo, or the treatment held as the standard of care at the time of the trial. Group two is given treatment T_2, a new treatment whose efficacy against T_1 the trial intends to establish. After all the members of both groups have either been treated for a predefined period of time (e.g., 5 years), or for whom a well-defined event has occurred (e.g., reappearance of disease or death), the trial ends and the efficacy of the two treatments is compared on the basis of a statistical indicator (e.g., relative risk) measuring a specific clinical outcome (e.g, disease-free survival). Below, we give as an example of such a clinical study result an excerpt from [10], a report on a trial comparing two treatments for patients with a particular kind of breast cancer.

> [. . .] patients with axillary lymph node-negative, estrogen receptor-positive
> breast cancer [. . .] chemotherapy plus tamoxifen resulted in significantly

better disease-free survival than tamoxifen alone (90% for methotrexate, fluorouracil, and tamoxifen (MFT) versus 85% for tamoxifen (P = .01); [. . .]

We will return to this example in the next section.

Having a language for formally expressing such results allows the creation of repositories of results and opens up the way for exploiting such repositories. We envisage the following ways in which such a system would be of benefit.

- A clinician can easily find publications related to a specific patient, a disease, or a treatment.

- A medical scientist can locate areas of inconsistency within the literature and thus locate research-worthy scientific topics.

- A health authority professional can construct arguments for and against the use of a treatment, as part of an evidence-based process of creating clinical guidelines.

- A doctor can generate argument and counter-argument structures that are useful in communicating to the patient the pros and cons of a treatment regimen. In this way, the patient and their preferences can be more directly involved in the decision making process.

For these benefits to arise, however, any approach aiming to address the problem set out above will have to satisfy certain requirements. We list below those requirements that we feel are essential; the list below is, of course, not exhaustive, but we believe that the following points deserve special mention.

Simple language. The language used to express knowledge should be simple and easily understandable by the user, even if the same knowledge can be expressed in richer, more powerful or better known formal languages.

Transparent reasoning. The complexity, uncertainty and potential conflicts inherent in clinical knowledge makes the success of systems that generate a single, "correct" decision, highly unlikely. Hence, systems that analyse the available knowledge and present the different possible results to the user, are preferable. This also leads to the requirement that the process by which the system constructs any structure, be it an argument, set of arguments and counter-arguments, or a logical conclusion, is simple, transparent and verifiable by the user.

Simple computation. As this framework is intended to be the basis of real, functional systems, and the computational complexity of reasoning is usually high, we should aim to restrict the language as much as possible so as to allow for optimising the corresponding algorithms.

These requirements have a number of consequences as to the characteristics of possible solutions. The first and third requirements imply a preference for a specialised and simplified language which rules out more expressive languages such as, e.g., first-order logic. In addition, there should be a fairly straightforward process for integrating a new

piece of information in a repository of results, allowing the scaling up of the system to a potentially large amount of data.

We also feel that the second desideratum naturally leads to systems that model domain-specific common-sense reasoning rather than systems that, as far as their users are concerned, 'magically' produce the 'right' answer. While a lot of clearly significant work exists on non-monotonic, default, paraconsistent and other logics, we feel that by aiming to represent some of the ways clinicians use knowledge produced by clinical studies, we are naturally led to a simple, monotonic, credulous logic.

Of course, using a monotonic, credulous logic will naturally give rise to inconsistency. For the purposes of analysis, such inconsistency is perhaps best suited to an argumentation framework, again reflecting how experts discuss and analyse results in order to arrive at a course of action, or a resolution of conflicting studies. For these reasons, we will develop an argumentation system as part of our approach, that allows users to examine and analyse conflicting sets of clinical study results. Argumentation is an emerging and promising research are providing methods for handling conflicting information (for reviews see [7, 20, 5, 6]).

The structure of the rest of this paper is as follows. We begin by discussing our proposed language for capturing results of clinical studies in Section 2. Then, in Section 3 we will develop the logic $\langle \mathcal{R}, \Vdash \rangle$ that aims at representing some of the inferences experts draw from such knowledge. The decidability of inference in this logic, as well as some algorithmic aspects are examined in Section 4. An upper bound for the computational complexity of inference is presented in Section 5. Using this logic as a basis, an argumentation system is defined in Section 6, along with an examination of several decision problems and structural properties. We conclude and discuss further avenues of study in Section 7.

2. A Language for Clinical Trial Results

To motivate our discussion of the language required for expressing clinical trial results, we will use the excerpt from [10], presented in the previous section and shown below, as a running example. We also note here that even though the terms *clinical study* and *clinical trial* are not equivalent, lacking a term that subsumes both we will use them interchangeably from now on.

> [...] patients with axillary lymph node-negative, estrogen receptor-positive breast cancer [...] chemotherapy plus tamoxifen resulted in significantly better disease-free survival than tamoxifen alone (90% for methotrexate, fluorouracil, and tamoxifen (MFT) versus 85% for tamoxifen [P = .01]; [...]

Clearly, the patients are the central subjects of clinical studies and therefore it is crucial to be able to describe succinctly and unambiguously the patient class involved in a clinical trial. This requirement already enjoys widespread interest, particularly in the development and use of *medical ontologies*, such as SNOMED [17]. Such ontologies provide a language for capturing patient characteristics (among other things) as well as logical machinery for answering queries such as whether one patient class is a subset

4

of another. Logics employed by ontologies are typically description logics, and these provide the necessary inference tools (see [2] which includes a chapter on medical ontologies).

However, in the quest for computational efficiency, multiple description logics have been defined by subtly varying their expressiveness and, by implication, the complexity of the related decision problems. As such, choosing a particular description logic can be an unnecessary constraint when we want to discuss a general system in which the ontology is but one part. For these reasons we will treat patient classes as simple sets, effectively using them as a boolean algebra. Therefore, for simplicity we will use a propositional logic whose letters correspond to basic patient characteristics. Our results will be largely generalisable to description logics as well, and we discuss this in Section 7.

Definition 1. *Let $\mathcal{L}_P$ be the language of a propositional logic with the usual connectives $\vee, \wedge, \neg, \Rightarrow$, a symbol for falsum $\bot$, and an inference relation $\vdash$. The propositional letters of $\mathcal{L}_P$ represent patient characteristics. A **patient class** is any formula $\phi \in \mathcal{L}_P$.*

We will use the lower case Greek letters $\phi, \psi, \chi, \ldots$ to denote patient classes from now on. We will also assume that $\vdash$ implicitly uses as premises the knowledge available on the structure of the patient classes. So, for example, if the patient classes $\phi, \psi \in \mathcal{L}_P$ are disjoint (e.g., with ϕ standing for the class of men and ψ for the class of women), we expect that $\phi \wedge \psi \vdash \bot$. We will also denote the length of a patient class ϕ as $|\phi|$, and we will understand this as the number of occurrences of propositional letters in ϕ.

In the excerpt above, the phrase "*patients with axillary lymph node-negative, estrogen receptor-positive breast cancer*" delineates the patient class in question. Using LN^- to denote "*axillary lymph node-negative*", ER^+ to denote "*estrogen receptor-positive*" and BC to denote "*breast cancer*", this patient class would be written as $LN^- \wedge ER^+ \wedge BC$ and its length would be 3.

The next component of our language is a way to denote specific treatments. Again, medical ontologies cater for this task by providing complex categories and relationships between treatments, regimes, substances used, and other characteristics. We would like to focus on the interrelationships between treatments and therefore, for simplicity, we will elide such complexity and assume that there is a set $\mathcal{T}$ whose elements represent specific **treatments**. So, for example a course of tamoxifen for 5 years would be a different element of $\mathcal{T}$ than a course of tamoxifen for 2 years. In the excerpt above, the phrases "*methotrexate, fluorouracil, and tamoxifen*" and "*tamoxifen*", denoted as MFT and TAM respectively, would be elements of $\mathcal{T}$. Note that even though the excerpt does not make this explicit, these are specific treatments with specific dosage, regime, duration, etc.

We also need to represent the criteria of comparison used in a clinical trial (see, e.g., [8] for details on the statistical methods employed in clinical trials). Results from clinical trials compare treatments in terms of **clinical outcome indicators**. These are statistical measures aiming to encompass an important aspect of the effect of the treatment. In the example, the outcome indicator used is "*disease-free survival*" which is measured as the percentage of patients that are alive and disease-free after a fixed period of time, usually 5 years. We will assume that there is a set COI of clinical outcome indicators, and in the example $DFS \in COI$ would denote disease-free survival.

Finally, we need a way to capture the result of the comparison. Such comparisons can be very information-rich, as is the case in the excerpt above: "*90% for methotrexate, fluorouracil, and tamoxifen (MFT) versus 85% for tamoxifen [P = .01]*". Here, the percentages of patients for disease-free survival are presented with respect to the treatment administered, accompanied by a *P* value, which is a statistical indicator of how reliable the result of the comparison is. We would like to focus on qualitative comparisons, thereby simplifying the available information, but also aiming to capture a common-sense way of reasoning with it. To this end, we allow for two possibilities for the comparison of two treatments $T_1, T_2 \in \mathcal{T}$ in terms of the clinical outcome indicator $I \in \mathcal{COI}$ used:

- One treatment, say T_1, is unambiguously (i.e., statistically significantly) better than T_2 with respect to I. We will represent this as $T_1 >_I T_2$.

- The trial was unable to show statistically significant differences between the two treatments with respect to I. We denote this by $T_1 \sim_I T_2$ (which will be treated as equivalent to $T_2 \sim_I T_1$).

Therefore, we define these formulae as follows.

Definition 2. *Let $T_1, T_2 \in \mathcal{T}$ be distinct treatments ($T_1 \neq T_2$) and $I \in \mathcal{COI}$. Then, any formula of the form $T_1 >_I T_2$ or of the form $T_1 \sim_I T_2$ is a **treatment comparison formula**. The set of all treatment comparison formulae will be denoted by $\mathcal{TF}$.*

We should note that the failure of a study to demonstrate superiority of one treatment over another, when interpreted in a strict statistical sense means that *there is no evidence to suggest superiority* as opposed to the common-sense interpretation that *there is evidence of equivalence*. A corollary of the strict statistical interpretation is that we should not be encoding non-statistically significant results at all. However, many clinicians are interested in inferences that could be drawn from the common-sense interpretation of such results, even though strictly speaking incorrect, e.g., when a clinical trial employing a very large number of patients yields a statistically insignificant result. For these reasons we adopt an agnostic approach where encoding and reasoning with such knowledge is allowed but not enforced.

Putting all these components together, we define the language $\mathcal{R}$ as follows.

Definition 3. *Let ϕ be a patient class and t a treatment comparison formula. Then, the language $\mathcal{R}$ consists of any formula of the form $\phi \rightarrow t$. A **rule** is any formula $\alpha \in \mathcal{R}$.*

Note that we call such formulae *rules*, reflecting the common-sense interpretation that if a patient belongs to the patient class ϕ and a trial has concluded $\phi \rightarrow t$, then this result is relevant to this patient. From now on we will use the lower case Greek letters $\alpha, \beta, \gamma, \ldots$ to denote rules, and upper case roman letters $A, B, C, \ldots$ to denote finite subsets of $\mathcal{R}$. Using this language, the result reported in the excerpt from [10] would be encoded as follows.

$$LN^- \wedge ER^+ \wedge BC \rightarrow MFT >_{DFS} TAM$$

For reasons that will become apparent in the next sections, we will visualise a set of rules using a labelled graph where the nodes are the treatments and the edges

correspond to rules and are labelled by the patient class of the associated rule. Below is the graph for the rule above.

$$MFT \xrightarrow{\;LN^+ \wedge ER^+ \wedge BC\;} TAM$$

We will also define the length of a rule $\phi \to t$ as the length of its patient class, i.e., $|\alpha| = |\phi|$. We do this since the length of any treatment comparison formula is constant. The number of elements of a set S is denoted by $|S|$ and the total length of a set of rules A is defined to be the sum of the lengths of all formulae in A, and denoted by $\|A\|$.

This language allows for expressing judgements of the type "for patient class ϕ, treatment T_1 is better than treatment T_2 with respect to clinical outcome indicator I". However, judgements involving different clinical outcome indicators are orthogonal and cannot interact in any way.[1] Therefore, for simplicity, we will assume that any given set of rules contains rules referring to the same clinical outcome indicator. If this is not the case, it is trivial to split the set in question into subsets where this condition holds. Then, each subset can be treated separately. For these reasons we will also omit the notation for clinical outcome indicators entirely.

A final observation is that there is no negation in $\mathcal{R}$. The main reason behind this is that clinical studies, generally, report positive results or their absence, none of which is equivalent to the negation of a positive result. In other words, normally a study cannot conclude that for the patient class ϕ, treatment T_1 is *not better* than treatment T_2, without at the same time concluding something stronger, i.e., that treatment T_2 is better than T_1, or that there is no significant evidence for either comparison. For these reasons and in order not to introduce something that would invalidate our desideratum about a simple language we decided against including negation. We will return to the issue of negation in Section 6, when we consider conflicts.

3. A Logic for Clinical Knowledge

Having defined a language for encoding the results of clinical trials, we are now able to approach the second desideratum which is the ability to reason with such results. Given a set of rules that represents a set of clinical trial results, the intention is to provide a way to construct the intuitive consequences of that set. We emphasise that we aim to define a simple, credulous logic, since we do not want to restrict ourselves only to the statistically valid inferences, which would be impossible to generate at this level of abstraction. Instead, we aim to capture some of the common-sense reasoning that clinicians employ, especially in the absence of directly relevant, statistically valid, results. To this end, we define an entailment relation $\Vdash$ on the language $\mathcal{R}$. We will do this by setting out inference rules that capture an appropriate notion of consequence.

The first inference rule captures equivalence of rewriting treatment comparison formulas involving $\sim$, and is straightforward.

---

[1]There can be interactions due to integrity constraints inherent in the meaning of the clinical outcome indicators. However, for simplicity, we will ignore such cases.

$$\frac{\phi \to T_1 \sim T_2}{\phi \to T_2 \sim T_1} \text{ EQ}$$

The first non-trivial inference rule we propose is one of **specialisation**. The motivation behind this inference rule comes from the process by which a doctor decides which of the available results are relevant to a patient to be treated. The patient could be potentially described as belonging to a patient class ϕ which includes all the observed characteristics. If there is no result that directly references the patient class ϕ then it is usual to consider as potential candidates all results whose patient class is a *superclass* of ϕ. We will, therefore, formalise this principle in the following inference rule.

$$\frac{\psi \to t}{\phi \to t} \text{ SP} \quad (\text{given } \phi \vdash \psi)$$

It can be easily seen that this inference rule will easily lead to contradictory results given a large enough rule set. To see this, suppose that a patient belongs to the patient class ϕ and that we have obtained the following results:

$$\psi \to T_1 > T_2 \qquad\qquad \chi \to T_2 > T_1$$

where $\phi \vdash \psi$ and $\phi \vdash \chi$. Using SP we can infer both $\phi \to T_1 > T_2$ and $\phi \to T_2 > T_1$, an intuitively contradictory statement. We will indeed define such situations as contradictory in Section 6, but we will refrain from trying to guess which rule is the 'right' one, a task which has been studied in the literature (e.g., [3]) and is related to the *specificity principle*, i.e., that if we know that $\chi \vdash \psi$ then we should conclude $\phi \to T_1 > T_2$ and *not* $\phi \to T_2 > T_1$. The reason we do not follow this line of enquiry is that (a) it would yield a non-monotonic logic which we have ruled out in our list of desiderata in Section 1 and (b) there are cases where a clinician would decide in favour of the more 'general' rule, i.e., $\chi \to T_2 > T_1$, e.g., if they believe that the clinical study resulting in this rule is somehow more credible or reliable than the one providing the other rule. One such occurrence that is common has to do with the statistical power of the clinical study, which is related to the number of participants.

Another important inference rule is **disjunction introduction** which can be thought of as the dual of specialisation. The motivation behind it is that if the same result has been obtained by comparing two treatments on two different patient classes, then it is reasonable to expect that the same holds for the union of the two patient classes.

$$\frac{\phi \to t \qquad \psi \to t}{\phi \vee \psi \to t} \text{ DI}$$

Finally, it is reasonable to expect that given a set of results on the same patient class, we can use **transitivity** to infer relationships between treatments. For example, if we know that $\phi \to T_1 > T_2$ and $\phi \to T_2 > T_3$ then we could infer that $\phi \to T_1 > T_3$. This deductive process is an abstraction of the reasoning used in the following example. Suppose that for a patient class ϕ suffering from a particular disease, there exists a treatment that is accepted as the standard of care, say T_2. Treatment T_2 has been tested in the past against placebo, T_3, and has been established as better according to some outcome, i.e., $\phi \to T_2 > T_3$. A new treatment T_1 is being trialled against T_2 and it turns out that it performs better, i.e., $\phi \to T_1 > T_2$. Now, in the case where the disease

in question is very serious, it is not ethical to perform a trial comparing T_1 against placebo, since then an inferior treatment (i.e., no treatment) would be knowingly given to patients of the disease. Therefore, we cannot possibly know directly whether $\phi \to T_1 > T_3$, and yet we infer from the two separate trials that this is indeed the case.

We formalise the transitivity inference rule below. Since we use two relations, $>$ and $\sim$, we obtain a set of four inference rules.

$$\frac{\phi \to T_1 > T_2 \qquad \phi \to T_2 > T_3}{\phi \to T_1 > T_3} \qquad \frac{\phi \to T_1 > T_2 \qquad \phi \to T_2 \sim T_3}{\phi \to T_1 > T_3}$$

$$\frac{\phi \to T_1 \sim T_2 \qquad \phi \to T_2 > T_3}{\phi \to T_1 > T_3} \qquad \frac{\phi \to T_1 \sim T_2 \qquad \phi \to T_2 \sim T_3}{\phi \to T_1 \sim T_3}$$

We will use symbols such as $\circ$, $\bullet$, $\diamond$ as meta-variables, for the symbols $>$ and $\sim$ in treatment comparison formulae. Using these, we will refer to the following inference rule as the transitivity rule, encompassing all four of the above possibilities.

$$\frac{\phi \to T_1 \circ T_2 \qquad \phi \to T_2 \bullet T_3}{\phi \to T_1 \diamond T_3} \; \text{TR}$$

Of course, $\diamond$ must represent a symbol in accordance with $\circ$ and $\bullet$, as above.

We should point out that specialisation, disjunction introduction and transitivity are inference rules that *do not necessarily draw statistically valid inferences*. The reason is that the presence of a confounding variable in the patient samples involved can invalidate the conclusion. By way of example, consider a clinical trial result of the form $\phi \to T_1 > T_2$ and a patient class ψ such that $\psi \vdash \phi$, thereby allowing the application of the specialisation rule. However, it could be that the relative sizes of the populations of ϕ and ψ are so different such that almost no patient of ψ took part in the trial on the patient class ϕ. This means that it is actually possible that $\psi \to T_2 > T_1$, or that $\psi \to T_1 \sim T_2$ is the case. Similar arguments exist for disjunction introduction and transitivity (see, e.g., [4]). As explained earlier, consequences derived using these inference rules are defeasible conclusions, intended to allow users to find potential ramifications of clinical study results.

For simplicity, whenever we apply an inference rule we will allow in-place rewriting of the patient class in the conclusion to any $\vdash$-equivalent one, as illustrated in Example 1. This is equivalent to applying SP after a given inference rule.

Having set out these inference rules, we can define entailment.

Definition 4. *A **proof** of a rule α from a set of rules A is a finite sequence of rules $\alpha_1, \ldots, \alpha_n$ such that $\alpha_n = \alpha$ and for any $i \le n$ either $\alpha_i \in A$, or there exist an inference rule and an index $j < i$ (in the case of SP and EQ), or indices $j, k < i$ (in the case of DI and TR) such that applying the rule in question to α_j, or to α_j, α_k respectively, yields α_i.*

Using the notion of proof we define entailment.

Definition 5. **Rule entailment** *is a relation $\Vdash \subseteq 2^{\mathcal{R}} \times \mathcal{R}$, defined as follows.*

$$A \Vdash \alpha \quad \textit{iff} \quad \textit{there exists a proof of } \alpha \textit{ from } A$$

Let A, B, C be sets of rules. We generalise $\Vdash$ to a relation between sets of rules in the obvious way: $A \Vdash B$ iff for all $\beta \in B$ it is the case that $A \Vdash \beta$. It is trivial to show that this relation is transitive, i.e., if $A \Vdash B$ and $B \Vdash C$ then $A \Vdash C$. It is also reflexive, i.e., $A \Vdash A$, and monotonic in the sense that if $A \Vdash B$ then it is also the case that $C \Vdash B$ for any $C \supseteq A$.

Two rules $\alpha, \beta \in \mathcal{R}$ are **equivalent** whenever it is the case that $\alpha \Vdash \beta$ and $\beta \Vdash \alpha$. Two sets of rules A, B are *equivalent* if $A \Vdash B$ and $B \Vdash A$. It is evident that two rules are equivalent precisely when each one can be derived from the other by application of EQ and SP only.

In what follows, a structure that functions as a witness for rule entailment that is different to that of a proof, will be useful. This structure is a proof tree.

Definition 6. *Let A be a set of rules and α a rule. A **proof tree** for α on A is a tree structure whose leaves are members of A, and its internal nodes represent applications of the inference rules, yielding a new rule as a consequence. The consequence of the root node is α.*

It should be clear that $A \Vdash \alpha$ iff there is a proof tree for α on A.

Example 1. *Let A be the following set of rules (left) and its corresponding graph (right). P and Q are patient characteristics, i.e., propositional letters in $\mathcal{L}_P$.*

$$A = \left\{ \begin{array}{l} P \to T_1 > T_2, \\ P \wedge Q \to T_2 > T_3, \\ \neg Q \to T_2 > T_3 \end{array} \right\}$$

We demonstrate a proof for $A \Vdash P \to T_1 > T_3$.

$$\cfrac{\cfrac{\cfrac{\neg Q \to T_2 > T_3}{P \wedge \neg Q \to T_2 > T_3}\ SP \qquad P \wedge Q \to T_2 > T_3}{P \to T_2 > T_3}\ DI \qquad P \to T_1 > T_2}{P \to T_1 > T_3}\ TR$$

Note the use of in-place rewriting of the patient class in the application of DI, i.e., the use of the fact that $\vdash (P \wedge \neg Q) \vee (P \wedge Q) \iff P$ in rewriting the conclusion of the application of DI accordingly.

4. Rule Entailment as a Decision Problem

We have effectively defined a simple, monotonic logic $\langle \mathcal{R}, \Vdash \rangle$ that, by applying inference rules that reflect important ways of common-sense reasoning employed by clinicians, can be used for working out the consequences of a given set of rules. The obvious next step is to look at this logic from an algorithmic viewpoint. We will call RULE-ENTAILMENT the decision problem of ascertaining whether it is the case that $A \Vdash \alpha$. In other words if A is a finite set of rules and α is a rule, a yes-instance for RULE-ENTAILMENT is a pair A, α if $A \Vdash \alpha$ and a no-instance otherwise.

Proposition 7. *RULE-ENTAILMENT is decidable.*

Proof. Assuming that there are p propositional letters in $\mathcal{L}_P$, we obtain that there are 2^p $\vdash$-equivalence classes of $\mathcal{L}_P$ formulae. By fixing an arbitrary representative formula for each $\vdash$-equivalence class (effectively a canonical normal form) any patient class can be rewritten in this normal form. Let the function that rewrites a patient class into its normal form be $\cdot^N$.

Given a finite set of treatments $\mathcal{T}$, there are at most $n = 2^p \cdot |\mathcal{T}|^2 \cdot 2$ many $\Vdash$-equivalence classes of rules. A normal form for rules clearly exists, extending the propositional one. We will denote the set of $\Vdash$-equivalence classes by $\mathcal{R}^N$ and extend the function $\cdot^N$ to rules and sets of rules. We can modify the inference rules given previously so that the conclusion is always rewritten in normal form, obtaining $\Vdash^N$, a new entailment relation. Clearly, $A \Vdash \alpha$ iff $A^N \Vdash^N \alpha^N$.

Given a finite set of rules A, there is only a finite number of non-repeating sequences of rules from A^N since the maximum length of such a sequence cannot be greater than $|\mathcal{R}^N| = n$. Any such sequence can be checked as to whether it constitutes a proof of α^N for some rule α. Therefore, checking whether $A^N \Vdash^N \alpha^N$ is decidable, and thus, so is checking whether $A \Vdash \alpha$. □

Since RULE-ENTAILMENT is decidable, it is reasonable to enquire about what algorithms can be constructed for it, as well as about its computational complexity. To address these questions we will focus on proof trees as witnesses of entailment rather than proofs, and on their properties. To this end, we will define special forms of proof trees and show that if a proof tree exists, then a proof tree of a particular form also exists. We start with the notion of an ordered proof tree.

Definition 8. *A proof tree is **ordered** iff in every branch from the root to the leaves, the sequence of applications of inference rules consists of:*

1. *Zero or more DI nodes, followed by,*
2. *zero or more TR nodes, followed by,*
3. *zero or one SP node, followed by,*
4. *zero or one EQ node.*

Example 2. *Returning to Example 1, we can see that there is an ordered proof tree for $P \to T_1 > T_3$ on A. We recall that:*

$$A = \left\{ \begin{array}{l} P \to T_1 > T_2, \\ P \wedge Q \to T_2 > T_3, \\ \neg Q \to T_2 > T_3 \end{array} \right\} \qquad T_1 \xrightarrow{\quad P \quad} T_2 \underset{\neg Q}{\overset{P \wedge Q}{\rightleftarrows}} T_3$$

Then, the following ordered proof tree for $P \to T_1 > T_3$ on A exists.

$$\cfrac{\cfrac{\cfrac{P \to T_1 > T_2}{P \wedge Q \to T_1 > T_2} \, SP \quad P \wedge Q \to T_2 > T_3}{P \wedge Q \to T_1 > T_3} \, TR \quad \cfrac{\cfrac{P \to T_1 > T_2}{P \wedge \neg Q \to T_1 > T_2} \, SP \quad \cfrac{\neg Q \to T_2 > T_3}{P \wedge \neg Q \to T_2 > T_3} \, SP}{P \wedge \neg Q \to T_1 > T_3} \, TR}{P \to T_1 > T_3} \, DI$$

In this case, an ordered proof tree exists. The natural question to ask is whether this is always the case. The following result addresses this question.

Proposition 9. *An ordered proof tree for α on A exists iff $A \Vdash \alpha$.*

Proof. The left-to-right direction is obvious. For the other direction, we prove the desired result by showing that applications of inference rules commute in certain directions. Thus, an arbitrary proof tree for α on A can be rewritten into an ordered one. We start by showing that EQ can be "pushed down" towards the leaves.

Suppose $A = \{\psi \to T_2 \sim T_1\}$, and that $\phi \vdash \psi$. Then if the following proof tree on the left can be rewritten as the one on the right.

$$\cfrac{\cfrac{\psi \to T_2 \sim T_1}{\phi \to T_2 \sim T_1}\ \text{SP}}{\phi \to T_1 \sim T_2}\ \text{EQ} \qquad \cfrac{\cfrac{\psi \to T_2 \sim T_1}{\psi \to T_1 \sim T_2}\ \text{EQ}}{\phi \to T_1 \sim T_2}\ \text{SP}$$

Suppose $A = \{\phi \to T_2 \sim T_1, \psi \to T_2 \sim T_1\}$. Again, the proof tree on the left can be replaced by the one on the right.

$$\cfrac{\cfrac{\phi \to T_2 \sim T_1 \quad \psi \to T_2 \sim T_1}{\phi \lor \psi \to T_2 \sim T_1}\ \text{DI}}{\phi \lor \psi \to T_1 \sim T_2}\ \text{EQ} \qquad \cfrac{\cfrac{\phi \to T_2 \sim T_1}{\phi \to T_1 \sim T_2}\ \text{EQ} \quad \cfrac{\psi \to T_2 \sim T_1}{\psi \to T_1 \sim T_2}\ \text{EQ}}{\phi \lor \psi \to T_1 \sim T_2}\ \text{DI}$$

Finally, consider $A = \{\phi \to T_3 \sim T_2, \phi \to T_2 \sim T_1\}$. The proof tree on the left can be rewritten as the one on the right.

$$\cfrac{\cfrac{\phi \to T_3 \sim T_2 \quad \phi \to T_2 \sim T_1}{\phi \to T_3 \sim T_1}\ \text{TR}}{\phi \to T_1 \sim T_3}\ \text{EQ} \qquad \cfrac{\cfrac{\phi \to T_3 \sim T_2}{\phi \to T_2 \sim T_3}\ \text{EQ} \quad \cfrac{\phi \to T_2 \sim T_1}{\phi \to T_1 \sim T_2}\ \text{EQ}}{\phi \to T_1 \sim T_3}\ \text{TR}$$

Also, it is easy to see that any branch consisting exclusively of applications of EQ can be simplified to one of length at most one. At this stage, we have shown that a proof tree can be transformed into one where at most one application of EQ appears only at the end of each root-to-leaves branch.

Next, we show that SP can be "pushed down" over TR and DI. Suppose $A = \{\psi \to T_1 > T_2, \psi \to T_2 > T_3\}$ and that $\phi \vdash \psi$. Then the following proof trees are equivalent.

$$\cfrac{\cfrac{\psi \to T_1 > T_2 \quad \psi \to T_2 > T_3}{\psi \to T_1 > T_3}\ \text{TR}}{\phi \to T_1 > T_3}\ \text{SP} \qquad \cfrac{\cfrac{\psi \to T_1 > T_2}{\phi \to T_1 > T_2}\ \text{SP} \quad \cfrac{\psi \to T_2 > T_3}{\phi \to T_2 > T_3}\ \text{SP}}{\phi \to T_1 > T_3}\ \text{TR}$$

The remaining combinations of $>$ and $\sim$ allowed by TR are handled in an identical manner.

Similarly, suppose that $A = \{\psi \to t, \chi \to t\}$ for some treatment comparison formula t and that $\phi \vdash \psi \lor \chi$. Then, the following proof trees for $\phi \to t$ on A are equivalent, given that from $\phi \vdash \psi \lor \chi$ and propositional logic we can show that $\vdash \phi \Leftrightarrow ((\phi \land \psi) \lor (\phi \land \chi))$.

$$\cfrac{\cfrac{\psi \to t \quad \chi \to t}{\psi \lor \chi \to t}\ \text{DI}}{\phi \to t}\ \text{SP} \qquad \cfrac{\cfrac{\psi \to t}{\phi \land \psi \to t}\ \text{SP} \quad \cfrac{\chi \to t}{\phi \land \chi \to t}\ \text{SP}}{\phi \to t}\ \text{DI}$$

Note the use of re-writing of a patient class in a logically equivalent one.

These cases demonstrate that applications of SP can be "pushed down" below applications of DI and TR. Also, any branch consisting of SP nodes can be transformed into one with only a single application of SP, due to the transitivity of $\vdash$. At this stage we have shown that an arbitrary proof tree can be transformed into one where on any root-to-leaves branch one encounters applications of TR and DI, followed by at most one application of SP, followed by at most one application of EQ.

Finally, we show that TR can be "pushed down" over DI. Let A be as follows.

$$A = \{\phi \rightarrow T_1 > T_2, \quad \psi \rightarrow T_1 > T_2, \quad \phi \vee \psi \rightarrow T_2 > T_3\}$$

Consider the following proof tree for $\phi \vee \psi \rightarrow T_1 > T_3$ on A.

$$\cfrac{\cfrac{\phi \rightarrow T_1 > T_2 \qquad \psi \rightarrow T_1 > T_2}{\phi \vee \psi \rightarrow T_1 > T_2}\text{ DI} \qquad \phi \vee \psi \rightarrow T_2 > T_3}{\phi \vee \psi \rightarrow T_1 > T_3}\text{ TR}$$

We can construct the following proof tree for $\phi \vee \psi \rightarrow T_1 > T_3$ on A.

$$\cfrac{\cfrac{\phi \rightarrow T_1 > T_2 \quad \cfrac{\phi \vee \psi \rightarrow T_2 > T_3}{\phi \rightarrow T_2 > T_3}\text{ SP}}{\phi \rightarrow T_1 > T_3}\text{ TR} \qquad \cfrac{\psi \rightarrow T_1 > T_2 \quad \cfrac{\phi \vee \psi \rightarrow T_2 > T_3}{\psi \rightarrow T_2 > T_3}\text{ SP}}{\psi \rightarrow T_1 > T_3}\text{ TR}}{\phi \vee \psi \rightarrow T_1 > T_3}\text{ DI}$$

Again, other combinations of $>$ and $\sim$ are identical. Note that the new occurrences of SP can be dealt with as previously noted. This completes the proof. $\qquad\square$

Now that we have proved that $A \Vdash \alpha$ iff an ordered proof tree for α on A exists, we can make several observations on such proof trees. First we note that, since disjunction introduction applications are found only at the top of the proof tree, their effect is equivalent to constructing a single, long disjunction, which is also the consequence of the proof tree. This observation leads us to define a new inference rule that will replace all such applications of DI. We call this new rule DI-n, since it is an n-ary version of DI.

$$\cfrac{\phi_1 \rightarrow t \qquad \ldots \qquad \phi_n \rightarrow t}{\bigvee_{i=1}^n \phi_i \rightarrow t}\text{ DI-}n$$

It should be clear that DI-n is sound in the sense that any application of DI-n can be replaced by several applications of DI.

The second observation on ordered proof trees is that applications of TR are also grouped together at the top of each sub-tree hanging off a disjunction introduction node. Clearly, each such sub-tree corresponds to a kind of path on the set of treatments. We formalise this intuitive notion here, which will be useful in what follows. Before doing that, we will define a function Tr that returns the set of treatments employed in a finite sequence of rules (below, $\circ$ stands for $>$ or $\sim$).

$$\text{Tr}(\langle \phi \rightarrow T_1 \circ T_2 \rangle) = \{T_1, T_2\} \text{ for any } \phi \in \mathcal{L}_P \text{ and distinct } T_1, T_2 \in \mathcal{T}$$
$$\text{Tr}(\langle \alpha_1, \ldots, \alpha_n, \alpha_{n+1} \rangle) = \text{Tr}(\langle \alpha_1, \ldots, \alpha_n \rangle) \cup \text{Tr}(\langle \alpha_{n+1} \rangle) \text{ for } n \geq 1$$

Definition 10. *Let S be a finite sequence of rules and T_s, T_e two treatments. We will say that S **is a simple path from** T_s **to** T_e (and write $SP(S, T_s, T_e)$) iff S is a simple $\sim$-path from T_s to T_e, or S is a simple $>$-path from T_s to T_e.*

*We define inductively whether a sequence S is a **simple $\sim$-path from** T_s **to** T_e (written $SP_\sim(S, T_s, T_e)$).*

$$SP_\sim(\langle\alpha\rangle, T_s, T_e) \quad iff \quad \begin{cases} \alpha = \phi \to T_s \sim T_e \quad or \\ \alpha = \phi \to T_e \sim T_s \end{cases}$$

$$SP_\sim(\langle\alpha_1, \ldots, \alpha_n, \alpha_{n+1}\rangle, T_s, T_e) \quad iff \quad \begin{cases} T_e \notin \mathrm{Tr}(\langle\alpha_1, \ldots, \alpha_n\rangle) \text{ and } \exists T_m \text{ s.t.} \\ SP_\sim(\langle\alpha_{n+1}\rangle, T_m, T_e) \text{ and,} \\ SP_\sim(\langle\alpha_1, \ldots, \alpha_n\rangle, T_s, T_m) \end{cases}$$

*We also define inductively whether a sequence S is a **simple $>$-path from** T_s **to** T_e, written $SP_>(S, T_s, T_e)$.*

$$SP_>(\langle\alpha\rangle, T_s, T_e) \quad iff \quad \alpha = \phi \to T_s > T_e$$

$$SP_>(\langle\alpha_1, \ldots, \alpha_n, \alpha_{n+1}\rangle, T_s, T_e) \quad iff \quad \begin{cases} T_e \notin \mathrm{Tr}(\langle\alpha_1, \ldots, \alpha_n\rangle) \text{ and } \exists T_m \text{ s.t.} \\ SP_>(\langle\alpha_{n+1}\rangle, T_m, T_e) \text{ and,} \\ SP(\langle\alpha_1, \ldots, \alpha_n\rangle, T_s, T_m) \\ \quad or, \\ SP_\sim(\langle\alpha_{n+1}\rangle, T_m, T_e) \text{ and,} \\ SP_>(\langle\alpha_1, \ldots, \alpha_n\rangle, T_s, T_m) \end{cases}$$

*Finally, we will say that a sequence of rules S is a **simple path** (respectively **simple $\sim$-path** or **simple $>$-path**) whenever there exist treatments T_s, T_e such that S is a simple path from T_s to T_e (respectively simple $\sim$-path from T_s to T_e or simple $>$-path from T_s to T_e).*

Note that this definition allows the starting and ending treatments to be the same, but forbids repetition of treatments within the path. Since no cycling is allowed inside a path there is only one way to order a set of rules into a path when the start and end treatments are designated. We will make use of this fact and somewhat abuse notation by treating simple paths as sequences but also as sets of rules.

Example 3. *Consider the following set of rules A'.*

$$A' = \left\{ \begin{array}{c} P \to T_1 > T_2, \\ P \wedge Q \to T_2 > T_3, \\ \neg Q \to T_2 \sim T_3 \end{array} \right\} \qquad T_1 \xrightarrow{\quad P \quad} T_2 \begin{array}{c} \overset{P \wedge Q}{\curvearrowright} \\ \underset{\neg Q}{\dashrightarrow} \end{array} T_3$$

Note that we indicate the use of $\sim$ with a dashed, non-directed edge, in agreement with the interpretation of $\sim$.

It is easy to see that $\langle\neg Q \to T_2 \sim T_3\rangle$ is a simple $\sim$-path from T_2 to T_3 and that $\langle P \to T_1 > T_2, P \wedge Q \to T_2 > T_3\rangle$ and $\langle P \to T_1 > T_2, \neg Q \to T_2 \sim T_3\rangle$ are both simple $>$-paths from T_1 to T_3.

Now we can return to our earlier observation and refine it. We can see that the use of applications of transitivity in ordered proof trees effectively corresponds to transitively inferring a rule on the basis of a simple path. Then, intuitively, the application of SP and EQ serves the purpose of transforming the original rules so that TR can be applied consecutively, building up a simple path. We define the following inference rule, which we call *compound transitivity* and denote it by CTR-n, and show how this rule captures exactly this process,

$$\frac{\phi_1 \to T_1 \circ_1 T_1' \qquad \ldots \qquad \phi_n \to T_n \circ_n T_n'}{\bigwedge_{i=1}^{n} \phi_i \to T_1 \bullet T_n'} \text{ CTR-}n$$

where the sequence $S = \left\langle \phi_1 \to T_1 \circ_1 T_1', \ldots, \phi_n \to T_n \circ_n T_n' \right\rangle$ forms a simple $\bullet$-path from T_1 to T_n', i.e., if S forms a simple $>$-path then $\bullet$ will stand for $>$, otherwise $\bullet$ will stand for $\sim$.

Proposition 11. *The compound transitivity rule is sound.*

Proof. Suppose $S = \left\langle \phi_1 \to T_1 \circ_1 T_1', \ldots, \phi_n \to T_n \circ_n T_n' \right\rangle$ forms a simple $\bullet$-path from T_1 to T_n'. For simplicity, we will assume that $T_i' = T_{i+1}$ for $i < n$ so that applying EQ is not necessary. Then, we can construct the following proof tree.

$$\frac{\dfrac{\dfrac{\phi_1 \to T_1 \circ_1 T_1'}{\bigwedge_{i=1}^{k} \phi_i \to T_1 \circ_1 T_1'} \text{ SP} \qquad \dfrac{\phi_2 \to T_2 \circ_2 T_2'}{\bigwedge_{i=1}^{k} \phi_i \to T_2 \circ_2 T_2'} \text{ SP}}{\bigwedge_{i=1}^{k} \phi_i \to T_1 \bullet_1 T_2'} \text{ TR} \qquad \ldots}{\begin{array}{c} \vdots \\ \hline \bigwedge_{i=1}^{k} \phi_i \to T_1 \bullet_{n-1} T_n' \end{array}} \text{ TR} \;\; \ldots \text{ TR}$$

It should be clear by Definition 10 and the conditions of the transitivity rule that $\bullet_{n-1}$ should match $\bullet$, therefore establishing that the application of CTR-n is sound in this case.

It should also be easy to see that relaxing the assumption that $T_i' = T_{i+1}$ for $i < n$ does not affect the soundness of CTR-n: indeed, it would at most necessitate the introduction of applications of EQ in the proof tree at the appropriate points just before the application of SP. $\qquad\square$

By using DI-n and CTR-n instead of DI and TR we can define a new kind of proof tree that will lead us to a characterisation result as well as a complexity result for rule entailment. This form of proof tree we call n-ordered and is as follows.

Definition 12. *A proof tree is called n-**ordered** iff on any branch from the root to the leaves, there is at most one application of inference rules in the following order: SP, DI-n, CTR-n.*

Example 4. *Continuing our running example, we show that there is an n-ordered proof tree for $P \to T_1 > T_3$ on A. We recall that*

$$A = \{P \to T_1 > T_2, \quad P \land Q \to T_2 > T_3, \quad \neg Q \to T_2 > T_3\}.$$

Then, the following n-ordered proof tree for $P \to T_1 > T_3$ on A exists.

$$\dfrac{\dfrac{P \to T_1 > T_2 \qquad P \wedge Q \to T_2 > T_3}{P \wedge Q \to T_1 > T_3} \; CTR\text{-}n \qquad \dfrac{P \to T_1 > T_2 \qquad \neg Q \to T_2 > T_3}{P \wedge \neg Q \to T_1 > T_3} \; CTR\text{-}n}{P \to T_1 > T_3} \; DI\text{-}n$$

Similarly to the case of ordered proof trees, we can ask the question of whether an n-ordered proof tree exists whenever an arbitrary proof tree exists. The following proposition settles this question.

Proposition 13. *There exists an n-ordered proof tree for α on A iff $A \Vdash \alpha$.*

Proof. The left-to-right direction is obvious. For the other direction, we first obtain the corresponding ordered tree by Proposition 9.

Let A be a set of rules such that a proof tree for a rule $\psi \to T_s \bullet T_e$ on A exists in which any branch from the root to the leaves contains one or more TR nodes followed by at most one EQ node, and then followed by at most one SP node. It should then be clear that a permutation of a subset of A forms a simple $\bullet$-path from T_s to T_e. Let this sequence of rules be $\langle \phi_1 \to T_1 \circ_1 T_1', \ldots, \phi_n \to T_n \circ_n T_n' \rangle$, with $T_1 = T_s$ and $T_n' = T_e$ (if the first or last rule employs $\sim$ then it could be that $T_1' = T_s$ or $T_n = T_e$, but for simplicity and without loss of generality we will assume the former pair of equations). Note also that applications of SP must specialise all ϕ_i to the same patient class ψ for TR to be applicable, and therefore it must be that $\psi \vdash \phi_i$ for all $i \leq n$. This means we can construct the following proof tree on A.

$$\dfrac{\dfrac{\phi_1 \to T_1 \circ_1 T_1' \qquad \ldots \qquad \phi_n \to T_n \circ_n T_n'}{\bigwedge_{i \leq n} \phi_i \to T_1 \bullet T_n'} \; CTR\text{-}n}{\psi \to T_s \bullet T_e} \; SP$$

This is sound because $\psi \vdash \phi_i$ for all $i \leq n$ entails that $\psi \vdash \bigwedge_{i \leq n} \phi_i$. By applying this result to an ordered proof tree, we can transform it to one that on any of its branches, one encounters from the root to the leaves, one or more DI nodes, at most one SP node and at most one CTR-n node.

We then prove that specialisation can be "pushed up" this transformed proof tree over DI. Suppose that we have the proof tree on the left and that $\phi \vdash \chi$. Then we can construct the proof tree on the right, given that $\phi \vdash \chi$ entails $\phi \vee \psi \vdash \chi \vee \psi$.

$$\dfrac{\dfrac{\chi \to t}{\phi \to t} \; SP \qquad \psi \to t}{\phi \vee \psi \to t} \; DI \qquad\qquad \dfrac{\dfrac{\chi \to t \qquad \psi \to t}{\chi \vee \psi \to t} \; DI}{\phi \vee \psi \to t} \; SP$$

Finally, we can merge successive applications of SP into a single application, as well as replace successive applications of DI with a single DI-n node. This completes the proof. $\qquad\square$

This result establishes the completeness of SP, DI-n and CTR-n with respect to the original inference rules. It also addresses an algorithmic issue related to the application of SP: in effect, SP is an inference rule that can yield multiple, different consequences if applied repeatedly to the same premise. This amounts to a non-deterministic choice of a consequence in algorithmic terms, and obscures the true complexity of rule entailment. By showing that we only ever need one application of SP, the one at the top of the

proof tree, we have fixed both the premises as well as the required consequence of the application of SP, avoiding in this way the non-deterministic choice.

A notational convenience we will use relates to the patient class appearing in the consequence of the application of CTR-n: let $S = \langle \phi_1 \to t_1, \ldots, \phi_n \to t_n \rangle$ be a simple path; we call $\bigwedge_{i \leq n} \phi_i$ the **patient class of** S and denote it by PC_S.

The following result characterises rule entailment on the basis of the existence of simple paths.

Theorem 14. *Let A be a set of rules and $\phi \to T_s \circ T_e$ a rule. Then,*

$$A \Vdash \phi \to T_s \circ T_e \quad \textit{iff} \quad \phi \vdash \bigvee_{S \in P} PC_S \ \textit{where}$$

$$P = \{\, S \subseteq A \mid S \ \textit{is a simple} \circ\textit{-path from } T_s \textit{ to } T_e \,\}$$

Proof. From Proposition 13 we know that $A \Vdash \phi \to T_s \circ T_e$ iff there is a n-ordered proof tree for $\phi \to T_s \circ T_e$ on A. By Definition 12, we know that there is a set P of simple $\circ$-paths from T_s to T_e, namely the premises of applications of CTR-n. The patient class of each such simple $\circ$-path is PC_S for $S \in P$, again by the definition of CTR-n. The application of DI-n collects all these patient classes into $\bigvee_{S \in P} PC_S \to T_s \circ T_e$, and finally, the application of SP requires that $\phi \vdash \bigvee_{S \in P} PC_S$. $\square$

Having obtained Theorem 14, we are in a position to develop an algorithm for $\Vdash$-entailment, by searching through all possible simple paths between the treatments in the query formula within a set of premises.

Algorithm 1 $A \Vdash_? \phi \to T_s \circ T_e$
___
$\psi \leftarrow \bot$
for all simple $\circ$-paths S from T_s to T_e in A **do**
$\quad \psi \leftarrow \psi \vee PC_S$
$\quad$ **if** $\phi \vdash \psi$ **then**
$\quad\quad\quad$ **return** yes
$\quad$ **end if**
end for
return no
___

Theorem 15. *Algorithm 1 is sound and complete.*

Proof. With respect to soundness, it should be clear that Algorithm 1, whenever it returns a positive answer, it does so having checked conditions witnessing the existence of a proof tree of the form delineated in Proposition 13. Conversely, with respect to completeness, Proposition 13 establishes that $A \Vdash \alpha$ iff there is a proof tree of the appropriate form for α on A. Algorithm 1 clearly searches all such proof trees. $\square$

5. The Complexity of Rule Entailment

We can now turn to the complexity of rule entailment. We recall that the class Π_2^p of the polynomial hierarchy [21, 14] can be defined in terms of oracle Turing Machines as

follows. Given a complexity class C, a C-*oracle* can be thought of as a sub-routine that can decide any decision problem in C in constant time. A C-oracle Turing Machine is a Turing Machine that makes queries to a C-oracle during its execution. The class of all decision problems decidable in the complexity class D by a C-oracle Turing Machine is denoted by D^{C}. Therefore, if, for example, a Turing Machine exists that decides a problem A in polynomial time given access to a NP-oracle, then $A \in \mathsf{P}^{\mathsf{NP}}$. Then, the class Π_2^p can be defined as the complement of the class of problems decidable in non-deterministic polynomial time given access to a NP-oracle, or in other words, $\Pi_2^p = \mathsf{coNP}^{\mathsf{NP}}$.

Proposition 16. *RULE-ENTAILMENT is in* Π_2^p.

Proof. Let $A = \left\{ \beta_i = \phi_i \to T_i \circ_i T_i' \mid i \leq n \right\}$ be a set of rules and let also $\alpha = \psi \to T_s \bullet T_e$. We want to decide whether $A \Vdash \alpha$. The size of the input is linear in $\|A\| + |\psi|$.

From Theorem 14, we know that:

$$A \Vdash \psi \to T_s \bullet T_e \quad \text{iff} \quad \psi \vdash \bigvee_{S \in P} PC_S$$

where P stands for the set of all simple $\bullet$-paths from T_s to T_e in A. We will use $p_1, \ldots, p_k$ to denote the propositional letters of the language of patient classes $\mathcal{L}_P$. Note that for a given set of rules A we need up to $\|A\|$ propositional letters, thus $k \leq \|A\|$. We can then rewrite the right-hand part as follows, denoting by M a valuation on $p_1, \ldots, p_k$.

$$\forall M, \text{ if } M \models \psi, \text{ then } \exists S \in P, \text{ such that } M \models PC_S$$

This is equivalent to

$$\forall M, \text{ if } M \models \psi, \text{ then } \exists B \subseteq A \text{ s.t. } B \text{ is a simple } \bullet\text{-path from } T_s \text{ to } T_e \text{ and } M \models PC_B.$$

First we show that given a valuation M and a set of rules $B \subseteq A$,

$$B \text{ is a simple } \bullet\text{-path from } T_s \text{ to } T_e \text{ and } M \models PC_B. \tag{*}$$

can be checked in time polynomial in $\|A\|$. It is well known that given a valuation M and a formula χ, checking whether $M \models \chi$ takes at most polynomial time in the sizes of M and χ. Also, note that for any $B \subseteq A$ it is the case that $|PC_B| \leq \|A\|$, and recall that $k \leq \|A\|$. Thus, checking $M \models PC_B$ takes at most polynomial time in $\|A\|$.

Moreover, given a set of rules $B \subseteq A$, we can check whether B constitutes a simple $\bullet$-path from T_s to T_e in time polynomial in $|B|$, akin to checking whether a set of edges in a graph constitutes a path between two given vertices. Since $|B| \leq |A|$, this establishes that (*) can be checked in polynomial time in $\|A\|$.

Therefore, given a valuation M, we can check whether there exists $B \subseteq A$ that satisfies (*) in non-deterministic polynomial time (in $\|A\|$). Recall once again that checking $M \models \chi$ can be checking in time polynomial in $\|A\| + |\psi|$. Thus, checking whether for any valuation M that satisfies ψ there exists a set of rules $B \subseteq A$ as above, is in $\mathsf{coNP}^{\mathsf{NP}}$ (in $\|A\| + |\psi|$), completing the proof. $\square$

6. Argumentation with Clinical Knowledge

We have seen how the logic $\langle \mathcal{R}, \Vdash \rangle$ allows the drawing of inferences from a set of rules that, while not necessarily statistically valid, are intended to reflect the ways experts might reason with clinical knowledge. Such inferences may help us to understand incomplete sets of rules, but may not help when such sets are inconsistent. In this context, argumentation can help in identifying the pros and cons of such inferences, in the form of arguments and counter-arguments. Therefore, in this section we will (a) define an argumentation system, (b) investigate notions of attacks between arguments, (c) examine the complexity of some relevant decision problems, and (d) show how notions of defeasible inference from the argumentation literature such as warrant, can be defined within this framework.

We have already noted that there is no negation in $\mathcal{R}$, and that one of our aims is to employ $\mathcal{R}$ within an argumentation approach, which obviously rests upon a notion of conflict. Without negation, this notion needs to be explicitly defined and we do this now. We define contradiction, a symmetric relation between subsets of $\mathcal{R}$, i.e., $\bowtie \subseteq 2^{\mathcal{R}} \times 2^{\mathcal{R}}$, as follows.

Definition 17. *Two sets $A, B \subseteq \mathcal{R}$ are **contradictory**, denoted as $A \bowtie B$, whenever there exist a consistent patient class ϕ ($\phi \nvdash \bot$) and two treatment comparison formulae f, g such that $A \Vdash \phi \rightarrow f$, $B \Vdash \phi \rightarrow g$ and for some $T_1, T_2 \in \mathcal{T}$, it is the case that:*

- *$f = T_1 > T_2$ and $g = T_2 > T_1$, or*

- *$f = T_1 > T_2$ and $g = T_1 \sim T_2$, or $g = T_2 \sim T_1$.*

Whenever the sets involved are singletons we will omit the set delimiters, e.g., if $\{\alpha\} \bowtie \{\beta\}$ we will write $\alpha \bowtie \beta$. A single set of rules A is called contradictory iff $A \bowtie A$.

Example 5. *Consider the following set of rules B, extending the set A we have seen in previous examples.*

$$B = \begin{cases} P \rightarrow T_1 > T_2, \\ P \wedge Q \rightarrow T_2 > T_3, \\ \neg Q \rightarrow T_2 > T_3 \\ P \rightarrow T_3 > T_1 \end{cases}$$

It should be clear that $B \Vdash P \rightarrow T_1 > T_3$ and, by the last rule in B, $B \Vdash P \rightarrow T_3 > T_1$. Therefore, B is contradictory.

Proposition 18. *Checking whether $\alpha_1 \bowtie \alpha_2$ is NP-complete.*

Proof. NP-hardness is immediate by reducing propositional satisfiability of ψ to checking whether $(\psi \rightarrow T_1 > T_2) \bowtie (\top \rightarrow T_2 > T_1)$.

Let $\alpha_1 = \phi_1 \rightarrow f_1, \alpha_2 = \phi_2 \rightarrow f_2$ be two rules. According to the definition they are contradictory iff there is a patient class ψ and treatment comparison formulae g_1, g_2 such that $\alpha_1 \Vdash \psi \rightarrow g_1$ and $\alpha_2 \Vdash \psi \rightarrow g_2$ and g_1, g_2 are as in Definition 17. Clearly, only EQ and SP can be used for proving these entailments. Only the application of

EQ affects the treatment comparison formulae, and it should be clear that a constant-time check is enough to ascertain whether from f_1 and f_2 one can derive appropriate formulae g_1, g_2 that satisfy Definition 17. In order to apply SP we need to check that $\psi \vdash \phi_1$ and $\psi \vdash \phi_2$, so it suffices to check whether $\phi_1 \wedge \phi_2 \nvdash \bot$, as the most general specialisation of ϕ_1 and ϕ_2. This establishes that checking whether $\alpha_1 \bowtie \alpha_2$ is within NP. $\qquad\square$

We have shown that checking whether two rules are contradictory can be done by using an algorithm for propositional satisfiability. By extension, we pose the question of how we can check whether a set of rules is contradictory. In a sense, the following definition in relation to contradiction mirrors that of Definition 10 in relation to entailment, and it will provide us with a means to the desired algorithm.

Definition 19. *A set of rules A is a **simple >-cycle** iff there exists a treatment T_0 such that all the members of A can be arranged into a simple >-path S from T_0 to T_0 with $PC_S \nvdash \bot$.*

Example 6. *The set of rules B is as follows.*

$$B = \left\{ \begin{array}{r} P \to T_1 > T_2, \\ P \wedge Q \to T_2 > T_3, \\ \neg Q \to T_2 > T_3 \\ P \to T_3 > T_1 \end{array} \right\}$$

It can be seen that $\{P \to T_1 > T_2, \neg Q \to T_2 > T_3, P \to T_3 > T_1\}$ is a simple >-cycle on account of the fact that $P \wedge \neg Q$ is consistent.

As before, it is reasonable to enquire whether the existence of a cycle within a set is equivalent to that set being contradictory. The following proposition shows that this is the case.

Proposition 20. *A set of rules A is contradictory iff there is a subset $B \subseteq A$ such that B is a simple >-cycle.*

Proof. Right-to-left: by assumption there is a treatment $T_0 \in \mathcal{T}$ such that B can be arranged into a sequence $\langle \phi_1 \to f_1, \ldots, \phi_n \to f_n \rangle$ that forms a simple >-path from T_0 to T_0. Let $S = \langle \phi_1 \to f_1, \ldots, \phi_{n-1} \to f_{n-1} \rangle$ (it is clear that for the premise to be true n must be greater than 1). By Definition 19 it follows that $\bigwedge_{i \leq n} \phi_i \nvdash \bot$. By Definition 10, there are two cases:

1. either S is a simple path from T_0 to some treatment T_n and $\langle \phi_n \to f_n \rangle$ is a simple >-path from T_n to T_0,
2. or, S is a simple >-path from T_0 to some treatment T_n and $\langle \phi_n \to f_n \rangle$ is a simple $\sim$-path from T_n to T_0,

In all cases, by applying SP to $\phi_n \to f_n$ we obtain $\bigwedge_{i \leq n} \phi_i \to f_n$.

In case (1), the application of CTR-n on S yields either $\bigwedge_{i<n} \phi_i \to T_0 > T_n$ or $\bigwedge_{i<n} \phi_i \to T_0 \sim T_n$, and both rules can be specialised to $\bigwedge_{i\leq n} \phi_i \to T_0 > T_n$ or

$\bigwedge_{i \leq n} \phi_i \rightarrow T_0 \sim T_n$ respectively. Clearly, f_n must be equal to $T_n > T_0$ and this completes the proof for this case.

In case (2), applying CTR-n on S yields $\bigwedge_{i<n} \phi_i \rightarrow T_0 > T_n$ which can be specialised to $\bigwedge_{i \leq n} \phi_i \rightarrow T_0 > T_n$. But then f_n must be equal to $T_n \sim T_0$ or to $T_0 \sim T_n$ and in either cases we have a contradiction.

Left-to-right: by assumption, $A \bowtie A$, which in turn means that there exist a patient class χ and distinct treatments $T_1, T_2 \in \mathcal{T}$ such that $A \Vdash \chi \rightarrow T_1 \circ T_2$, $A \Vdash \chi \rightarrow T_2 \bullet T_1$ and $\circ, \bullet$ stand for a combination that makes these two rules contradictory. Consider the two ordered proof trees obtained as in Proposition 9. Let $\phi_1, \ldots, \phi_n$ and $\psi_1, \ldots, \psi_m$ be the patient classes that the specialisation rules result in, in the two proof trees. By the structure of the proof trees and the assumptions above, it follows that $\vdash \chi \Leftrightarrow \bigvee_{i \leq n} \phi_i$ and $\vdash \chi \Leftrightarrow \bigvee_{j \leq m} \psi_j$, hence $\vdash \bigvee_{i \leq n} \phi_i \Leftrightarrow \bigvee_{j \leq m} \psi_j$. Therefore, it must be that there exist $k \leq n$ and $l \leq m$ such that $\phi_k \wedge \psi_l \nvdash \bot$, since $\chi \nvdash \bot$. The indices k and l correspond to two sub-trees of the two proof trees that only contain applications of TR, SP and EQ, with leaves $B, C \subseteq A$ respectively, such that $B \Vdash \phi_k \rightarrow T_1 \circ T_2$ and $C \Vdash \psi_l \rightarrow T_2 \bullet T_1$. It should then be clear that $B \cup C$ can be arranged into a simple $>$-path from T_1 to T_1, thus completing the proof. $\qquad \square$

Proposition 20 implicitly points at an algorithm for checking whether a set of rules is contradictory. By examining this link we can conclude that the complexity of this decision problem turns out to be potentially lower than that of entailment, as the next result shows.
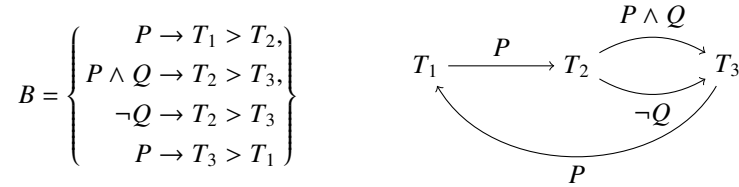
Proposition 21. *Let A be a set of rules. Checking whether A is contradictory is* NP-*complete.*

Proof. Proof of NP-hardness is implied by Proposition 18. We show membership of the problem in NP by supplying a non-deterministic polynomial time algorithm. Given Proposition 20, in order to check whether A is contradictory we can check whether A contains a simple $>$-cycle. To do that, one needs to non-deterministically guess a subset of A, check that it is a path from a treatment to itself, and check the consistency of its patient classes. This is possible by:
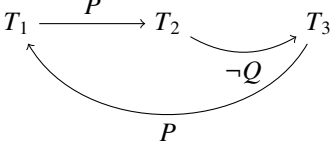
1. non-deterministically guessing a valuation M for $\mathcal{L}_P$, a subset $B \subseteq A$ and a treatment T_0,
2. checking whether B is a simple $>$-path from T_0 to T_0,
3. and checking whether $M \models PC_B$.

As we have argued earlier, steps 2 and 3 can be performed in polynomial time in $\|A\|$, and therefore the problem is in NP. $\qquad \square$
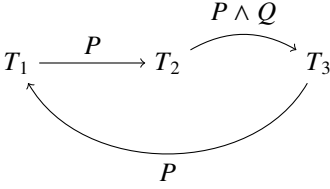
Example 7. *We recall the set of rules B as in previous examples.*

$$B = \begin{Bmatrix} P \rightarrow T_1 > T_2, \\ P \wedge Q \rightarrow T_2 > T_3, \\ \neg Q \rightarrow T_2 > T_3 \\ P \rightarrow T_3 > T_1 \end{Bmatrix}$$

We have seen that

$$C = \left\{ \begin{array}{c} P \rightarrow T_1 > T_2, \\ \neg Q \rightarrow T_2 > T_3 \\ P \rightarrow T_3 > T_1 \end{array} \right\}$$

is a simple >-cycle. The following set is also a simple >-cycle

$$D = \left\{ \begin{array}{c} P \rightarrow T_1 > T_2, \\ P \wedge Q \rightarrow T_2 > T_3, \\ P \rightarrow T_3 > T_1 \end{array} \right\}$$

on account of the fact that $P \wedge Q \nvdash \bot$. Moreover, both C and D are minimal contradictory sets in the sense that the removal of any of their members would make them non-contradictory.

Therefore, and in anticipation of constructing counter-arguments, we may ask if it is the case that a *minimal* contradictory set not only contains a cycle, but actually is one. Again, the answer is positive as shown in the next proposition.

Proposition 22. *Let A be a contradictory set of rules. It is the case that for all $\beta \in A$, $A \setminus \{\beta\}$ is non-contradictory iff A is a simple >-cycle.*

Proof. Right-to-left: if A is a simple >-cycle, then it is obvious that the removal of any of its members will cause A to cease being a simple >-cycle. By Proposition 20 we obtain that $A \setminus \{\beta\}$ for any $\beta \in A$, is non-contradictory.

Left-to-right: assuming that A is a contradictory set of rules, the application of Proposition 20 yields a subset $B \subseteq A$ such that B is a simple >-cycle. If $B = A$ then we are done, so we assume that $B \subset A$. By applying Proposition 20 again we obtain that B is contradictory. The assumption that for all $\beta \in A$, $A \setminus \{\beta\}$ is non-contradictory leads us to a contradiction, completing the proof. $\square$

We are now ready to define the notion of an argument within our framework. Given that several different conclusions are possible from a given set of rules A, as is usual in some deduction-based argumentation systems, we will include the designated conclusion as a component of an argument. In this way, we are able to distinguish between arguments that may have the same premises but different conclusions. Also, from now on we will be concerned with arguments constructed on the basis of a designated set of rules. We will denote this finite set of rules as K.
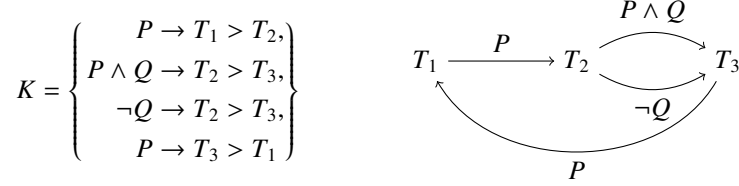
Definition 23. *An **argument** is a pair $\langle A, \alpha \rangle$, where A is a finite set of rules (the **support** of the argument) and α is a rule (the **claim** of the argument), such that:*

1. *$A \Vdash \alpha$ (entailment).*

2. *A is not contradictory (non-contradiction).*
3. *There is no $\beta \in A$ such that $A \setminus \{\beta\} \Vdash \alpha$ (minimality).*

The set of all arguments is denoted by $\mathcal{A}$. The set of all arguments whose support is a subset of K is denoted by $\mathcal{A}_K$.

Example 8. *Recall the set of rules B as in previous examples. We will use this set as the designated set of rules K out of which all arguments are to be constructed.*

$$K = \begin{cases} P \to T_1 > T_2, \\ P \wedge Q \to T_2 > T_3, \\ \neg Q \to T_2 > T_3, \\ P \to T_3 > T_1 \end{cases}$$

It should be clear that the following pairs are arguments.

$$\langle \{P \to T_3 > T_1, \quad P \to T_1 > T_2\}, \quad P \to T_3 > T_2 \rangle$$
$$\langle \{P \wedge Q \to T_2 > T_3, \quad \neg Q \to T_2 > T_3\}, \quad P \to T_2 > T_3 \rangle$$
$$\langle \{P \to T_1 > T_2, \quad \neg Q \to T_2 > T_3\}, \quad P \wedge \neg Q \to T_1 > T_3 \rangle$$

Proposition 24. *Let $\langle A, \phi \to T_1 \circ T_2 \rangle$ be an argument. Then, there is a non-empty set P of subsets of A such that each $S \in P$ is a simple $\circ$-path from T_1 to T_2, and $A = \bigcup_{S \in P} S$.*
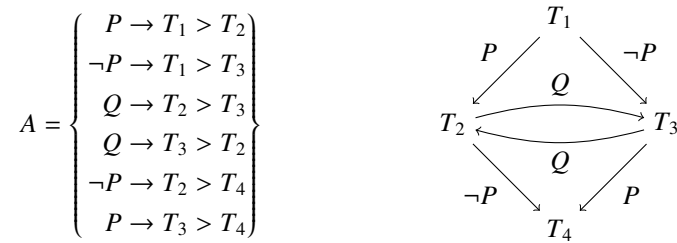
Proof. By assumption, $A \Vdash \phi \to T_1 \circ T_2$ and therefore, using Theorem 14 we obtain that there is a set P of subsets of A with the following property:

$$P = \{ S \subseteq A \mid S \text{ is a simple } \circ\text{-path from } T_1 \text{ to } T_2 \}$$

This set P satisfies by construction the property that any of each members is a simple $\circ$-path from T_1 to T_2. It remains to show that $A = \bigcup_{S \in P} S$. Let $A' = \bigcup_{S \in P} S$, which is obviously a subset of A. By applying Theorem 14 again on P we obtain that $A' \Vdash \phi \to T_1 \circ T_2$. But by assumption, A minimally entails $\phi \to T_1 \circ T_2$ thus it cannot be that $A' \subset A$. Therefore $A = A'$, completing the proof. $\square$

A valid question is whether our inclusion of the minimality condition in Definition 23 obviates the requirement for a non-contradictory support. In other words, if $A \Vdash \alpha$ and for all $\beta \in A$ it is the case that $A \setminus \beta \nVdash \alpha$, then can we conclude that A is non-contradictory? The following example refutes this conjecture.

Example 9. *Consider the following set of rules.*

$$A = \begin{cases} P \to T_1 > T_2 \\ \neg P \to T_1 > T_3 \\ Q \to T_2 > T_3 \\ Q \to T_3 > T_2 \\ \neg P \to T_2 > T_4 \\ P \to T_3 > T_4 \end{cases}$$

It should be clear that A is contradictory, since it includes a simple $>$-cycle ($Q \to T_2 > T_3$, $Q \to T_3 > T_2$). Also, it is easy to see that $A \Vdash Q \to T_1 > T_4$, and that there is no strict subset $B \subset A$ such that $B \Vdash Q \to T_1 > T_4$.

Having defined what an argument is, we naturally come to the issue of what is the complexity of the decision problem of checking whether a pair $\langle A, \alpha \rangle$ is an argument. The following result establishes an upper bound, given the upper bound for rule entailment. We recall that the complexity class Δ_3^p includes decision problems that can be solved using a polynomial number of queries to a Π_2^p (or, equivalently to a Σ_2^p) oracle.

Corollary 25. *Checking whether $\langle A, \alpha \rangle$ constitutes an argument is in Δ_3^p.*

Proof. 1. We need to check that $A \Vdash \alpha$, which by Proposition 16 is in Π_2^p.
  2. We need to check that A is non-contradictory, a problem which by Proposition 21 is coNP-complete.
  3. Finally, we need to check that for any $\beta \in A$, $A \setminus \{\beta\} \not\Vdash \alpha$. This can be done by issuing $|A|$ queries to a Π_2^p oracle, therefore this problem is in Δ_3^p.

$\square$

We now turn to studying notions of attack between arguments. There are several possible definitions of attack; we could define attack as two arguments having contradictory claims; or we would define it by expecting one argument to have a claim that contradicts one of the premises of another argument. We will opt for the latter option which is a more general notion, and allows for demonstrating a specific kind of attack commonly called undercutting.

Definition 26. *Let $\mathbf{A} = \langle A, \alpha \rangle$ and $\mathbf{B} = \langle B, \beta \rangle$ be two arguments. We say that $\mathbf{A}$ is an **undercut** of $\mathbf{B}$ iff there exists $\gamma \in B$ such that $\alpha \bowtie \gamma$. In this case we will also say that $\mathbf{A}$ **undercuts** $\mathbf{B}$ and that $\mathbf{A}$ **undercuts** $\mathbf{B}$ **at** γ.*

When $\langle A, \alpha \rangle$ is an argument and $\alpha \bowtie \gamma$ for some rule γ we will also say that $\mathbf{A}$ undercuts γ. An example of arguments and undercuts is given next.

Example 10. *Recall the set of rules K.*

$$K = \{P \to T_1 > T_2, \quad P \land Q \to T_2 > T_3, \quad \neg Q \to T_2 > T_3, \quad P \to T_3 > T_1\}$$

We set out below arguments and list the cases where one undercuts another below.

$$\mathbf{A} = \langle \{P \to T_3 > T_1, \qquad P \to T_1 > T_2\}, \qquad\qquad P \to T_3 > T_2 \rangle$$
$$\mathbf{B} = \langle \{P \land Q \to T_2 > T_3, \quad \neg Q \to T_2 > T_3\}, \qquad\qquad P \to T_2 > T_3 \rangle$$
$$\mathbf{C} = \langle \{P \to T_1 > T_2, \qquad \neg Q \to T_2 > T_3\}, \qquad P \land \neg Q \to T_1 > T_3 \rangle$$

We can see that $\mathbf{A}$ undercuts $\mathbf{B}$ at $P \land Q \to T_2 > T_3$; $\mathbf{A}$ undercuts $\mathbf{B}$ at $\neg Q \to T_2 > T_3$; $\mathbf{A}$ undercuts $\mathbf{C}$ at $\neg Q \to T_2 > T_3$; and that $\mathbf{B}$ undercuts neither $\mathbf{A}$ nor $\mathbf{B}$.

The complexity of checking whether an argument undercuts another is potentially lower than that of deciding entailment, as the next result demonstrates.

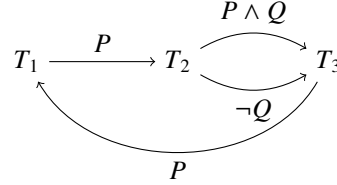**Proposition 27.** *Let* **A***,* **B** *be two arguments. Checking whether* **A** *undercuts* **B** *is* NP-*complete.*

Proof. Let ϕ be a propositional formula. Let $\alpha = \phi \to T_1 > T_2$ and $\beta = \top \to T_2 > T_1$. It is easy to see that $\mathbf{A} = \langle \{\alpha\}, \alpha \rangle$ and $\mathbf{B} = \langle \{\beta\}, \beta \rangle$ are arguments. Moreover, **A** undercuts **B** iff $\phi \nvdash \bot$, establishing NP-hardness.

Let $\alpha = \phi \to f$ be the claim of **A**, and B the support of **B**. We give a non-deterministic polynomial-time algorithm for checking that there exists $\gamma \in B$ such that $\alpha \bowtie \gamma$. First, a valuation v of all propositional letters $p_1, \ldots, p_n$ appearing in α and B is chosen. A rule $\gamma \in B$ with $\gamma = \psi \to g$ is also chosen non-deterministically.[2] We check that the valuation v satisfies $\phi \wedge \psi$, something that can be done in time polynomial in $|\phi| + |\psi|$ (note that $|\psi| \leq \|B\|$). We finally check that f, g are as in Definition 17, requiring constant time and return the result of the comparison as the answer to the decision problem. This completes the proof of membership in NP. □

Proposition 20 establishes a relationship between a structural property of a set of rules, i.e., whether it contains a simple $>$-cycle, and whether that set is contradictory. Extending this line of enquiry leads us to the question of whether there is a similar result relating structural properties of an argument **A** and a claim γ, to the fact that **A** undercuts γ. The following example illustrates such a situation.

Example 11. *Recall the set of rules K.*

$$K = \begin{cases} P \to T_1 > T_2, \\ P \wedge Q \to T_2 > T_3, \\ \neg Q \to T_2 > T_3, \\ P \to T_3 > T_1 \end{cases}$$

Consider the rule $\alpha = P \to T_1 > T_2$ and the following argument.

$$\mathbf{D} = \langle \{P \wedge Q \to T_2 > T_3, \ \neg Q \to T_2 > T_3, \ P \to T_3 > T_1\}, \ P \to T_2 > T_1 \rangle$$

It should be clear that **D** *undercuts α and that its support contains two distinct simple $>$-paths from T_2 to T_1, and that, therefore, the union of the support of* **D** *with $\{\alpha\}$ will contain two distinct simple $>$-cycles.*

So, given an argument $\langle A, \alpha \rangle$ that undercuts a rule γ, the conjecture illustrated in the last example is that a particular relationship holds between the support A and γ, namely that $A \cup \{\gamma\}$ will consist entirely of simple $>$-cycles. The following result verifies this conjecture.

Proposition 28. *Let* $\mathbf{A} = \langle A, \alpha \rangle$ *be an argument and β be a rule such that* **A** *undercuts β. Then, there exist a number $n \geq 1$ and a collection of sets of rules $\{ C_i \subseteq \mathcal{R} \mid i \leq n \}$, such that the following hold:*

---

[2] In fact, the domain of this choice is not exponential in the size of the problem so this can also be achieved by using disjunction instead of non-determinism.

1. *Every C_i is a simple >-cycle, for $i \leq n$.*
2. *For any $i \leq n$, it is the case that $\beta \in C_i$.*
3. *$A \cup \{\beta\} = \bigcup_{i \leq n} C_i$.*

Proof. We define $\{C_i \mid i \leq n\}$ to be the set of all simple >-cycles contained in $A \cup \{\beta\}$. We need to show that this collection is non-empty and that it satisfies the second and third conditions set above.

To show that the set $\{C_i \mid i \leq n\}$, as defined, is non-empty it suffices to show that $A \cup \{\beta\}$ is contradictory, since then Proposition 20 would assert the existence of at least one simple >-cycle. By assumption, since $\alpha \bowtie \beta$, there exist $\phi \in \mathcal{L}_P$ and $f, g \in \mathcal{TF}$ as in Definition 17 such that $\alpha \Vdash \phi \rightarrow f$ and $\beta \Vdash \phi \rightarrow g$. Since **A** is an argument, it follows that $A \Vdash \alpha$ and therefore $A \Vdash \phi \rightarrow f$, by the transitivity of $\Vdash$. Therefore, $A \cup \{\beta\}$ is contradictory.

By the assumption that **A** is an argument, A is non-contradictory and, therefore, by applying Proposition 20, we obtain that there is no simple >-cycle contained in A. Thus, any simple >-cycle in $A \cup \{\beta\}$ will have to involve β, and therefore, the second condition holds.

Suppose $\alpha = \psi \rightarrow T_1 > T_2$. Then, β must be of the form $\chi \rightarrow T_2 > T_1$ or $\chi \rightarrow T_1 \sim T_2$ or $\chi \rightarrow T_2 \sim T_1$. Consider the n-ordered proof tree for α on A obtained through Proposition 13. Every set of rules S used in each application of CTR-n must be a simple >-path from T_1 to T_2, and therefore, $S \cup \{\beta\}$ is a simple >-path from T_1 to T_1. In addition, $\psi \wedge \chi$ is consistent by the assumption that $\alpha \bowtie \beta$, and therefore $S \cup \{\beta\}$ is a simple >-cycle. When α is of the form $\psi \rightarrow T_1 \sim T_2$, we can prove in an identical manner that $S \cup \{\beta\}$ is a simple >-cycle.

Let us suppose then that the third condition does not hold, i.e., that there is $\gamma \in A \cup \{\beta\}$ such that for all $i \leq n$, $\gamma \notin C_i$. By the second condition we can conclude that $\gamma \neq \beta$ and therefore it must be that $\gamma \in A$. But then, by the minimality condition, γ must belong to a simple >-path $S \subseteq A$ as above, and therefore $\gamma \in S \cup \{\beta\}$, a contradiction. $\qquad\square$

We now turn to the generation of undercuts to a given argument. As usual in argumentation, we may have to consider different undercuts as somehow equivalent, in that they may be distinct arguments, but differing only in ways that we consider unimportant. The example below demonstrates the redundancy due to varying the claim of an undercut.

Example 12. *Again, consider the set of rules K.*

$$K = \{P \rightarrow T_1 > T_2, \quad P \wedge Q \rightarrow T_2 > T_3, \quad \neg Q \rightarrow T_2 > T_3, \quad P \rightarrow T_3 > T_1\}$$

The following is obviously an argument in $\mathcal{A}_K$, as we have previously seen.

$$\mathbf{B} = \langle \{P \wedge Q \rightarrow T_2 > T_3, \neg Q \rightarrow T_2 > T_3\}, P \rightarrow T_2 > T_3 \rangle$$

We have also seen how **A** *(shown below) is an undercut of* **B**. *The arguments* **A'** *and* **A''** *have the same support as* **A** *and also undercut* **B** *at $P \wedge Q \rightarrow T_2 > T_3$. Note that we*

assume that the language $\mathcal{L}_P$ *contains the propositional letter R in addition to those we have used before.*

$$\mathbf{A} = \langle \{P \rightarrow T_3 > T_1, P \rightarrow T_1 > T_2\}, \quad P \rightarrow T_3 > T_2 \rangle$$
$$\mathbf{A}' = \langle \{P \rightarrow T_3 > T_1, P \rightarrow T_1 > T_2\}, \quad P \wedge Q \rightarrow T_3 > T_2 \rangle$$
$$\mathbf{A}'' = \langle \{P \rightarrow T_3 > T_1, P \rightarrow T_1 > T_2\}, \quad P \wedge R \rightarrow T_3 > T_2 \rangle$$

As the above example shows, when we ask the question which are the possible undercuts to a given argument, we can generate a large number of seemingly redundant arguments which share the same support but have increasingly narrow patient classes in their claims. Indeed, the number of such arguments will in general be exponential to the number of propositional letters in $\mathcal{L}_P$. Therefore, we would like to single out the most general representative undercuts for a given argument. We do this with the following definition and by noting that in the entailment $A \Vdash \alpha$ in an argument $\langle A, \alpha \rangle$, it is the specialisation step in the corresponding n-ordered tree that allows for such redundancy.

Definition 29. *Let* $\mathbf{A} = \langle A, \alpha \rangle$ *be an argument in* $\mathcal{A}_K$. *We will call* $\mathbf{A}$ *maximally liberal iff there is no argument* $\langle A, \beta \rangle$ *such that* $\beta \Vdash \alpha$ *and* $\alpha \nVdash \beta$. *The set of all maximally liberal arguments within* A_K *will be denoted by* $\mathcal{A}_K^L$.

Note that the only inference rule that can be applied to a single rule is SP. Therefore, if $\mathbf{A} = \langle A, \phi \rightarrow t \rangle$ is an argument and $\mathbf{A}' = \langle A, \psi \rightarrow t \rangle$ is a maximally liberal argument, then it must be that $\phi \vdash \psi$.

Corollary 30. *Let* $\mathbf{A} = \langle A, \phi \rightarrow T_1 \circ T_2 \rangle$ *be an argument, and let*

$$P = \{ S \subseteq A \mid S \text{ is a simple } \circ\text{-path from } T_1 \text{ to } T_2 \}.$$

Then $\mathbf{A}$ *is maximally liberal iff* $\vdash \phi \Leftrightarrow \bigvee_{S \in P} PC_S$.

Proof. It should be clear by the construction of the relevant n-ordered proof tree on A that the most general patient class for the entailed rule is $\bigvee_{S \in P} PC_S$. Thus, it follows from Definition 29 that if $\vdash \phi \Leftrightarrow \bigvee_{S \in P} PC_S$ then $\mathbf{A}$ must be maximally liberal. Conversely, if $\mathbf{A}$ is maximally liberal then ϕ is the most general patient class, thus $\vdash \phi \Leftrightarrow \bigvee_{S \in P} PC_S$. $\qquad\square$

Putting together Proposition 24 and Corollary 30, it is easy to see that given a set P of simple $\circ$-paths from T_1 to T_2 there is only one patient class modulo propositional equivalence such that $\langle \bigcup_{S \in P} S, \alpha \rangle$ is a maximally liberal argument, namely $\alpha = \bigvee_{S \in P} PC_S \rightarrow T_1 \circ T_2$ (to ensure the minimality of the support we assume that all paths $S \in P$ have non-equivalent patient classes). This entails that we can omit the claim of a maximally liberal argument as it can be inferred from the support. For notational convenience we will at times replace the claim with $*$.

Our intention, then, is to concentrate our attention only on the maximally liberal undercuts of a given argument, thereby reducing the redundancy of such a set. But have we addressed all forms of redundancy that exist within the set of undercuts of an argument? We present a counter-example below.

Example 13. *Recall the set of rules K.*

$$K = \{P \rightarrow T_1 > T_2, \quad P \wedge Q \rightarrow T_2 > T_3, \quad \neg Q \rightarrow T_2 > T_3, \quad P \rightarrow T_3 > T_1\}$$

Consider again the rule $P \rightarrow T_1 > T_2$ and the argument $\mathbf{D}$.

$$\mathbf{D} = \langle \{P \wedge Q \rightarrow T_2 > T_3, \neg Q \rightarrow T_2 > T_3, P \rightarrow T_3 > T_1\}, P \rightarrow T_2 > T_1 \rangle$$
$$\mathbf{D_1} = \langle \{P \wedge Q \rightarrow T_2 > T_3, P \rightarrow T_3 > T_1\}, P \wedge Q \rightarrow T_2 > T_1 \rangle$$
$$\mathbf{D_2} = \langle \{\neg Q \rightarrow T_2 > T_3, P \rightarrow T_3 > T_1\}, P \wedge \neg Q \rightarrow T_2 > T_1 \rangle$$

We can see that $\mathbf{D}$, $\mathbf{D_1}$ and $\mathbf{D_2}$ all undercut the rule $P \rightarrow T_1 > T_2$, and that they are all maximally liberal. Also, observe that the supports of $\mathbf{D_1}$ and $\mathbf{D_2}$ are strict subsets of the support of $\mathbf{D}$. In some sense, $\mathbf{D}$ undercuts $P \rightarrow T_1 > T_2$ in a way that can be broken down to the ways $\mathbf{D_1}$ and $\mathbf{D_2}$ undercut $P \rightarrow T_1 > T_2$. This is because the claim of $\mathbf{D}$ is too general and therefore requires more rules from K.

Following on from the example, the next definition aims to capture those maximally liberal arguments that have a minimal support in order to undercut a particular rule.

Definition 31. *Let $\mathbf{A} = \langle A, \alpha \rangle$ be a maximally liberal argument in $\mathcal{A}_K^L$ and let γ be a rule. We say that $\mathbf{A}$ **canonically undercuts** γ iff (a) $\mathbf{A}$ undercuts γ and (b) for any argument $\langle B, \beta \rangle \in \mathcal{A}_K^L$ that undercuts γ, it is the case that $B \not\subset A$. As usual, we say that $\mathbf{A}$ **canonically undercuts** $\mathbf{B} = \langle B, \beta \rangle$, or that $\mathbf{A}$ is a **canonically undercut** of $\mathbf{B}$ iff there exists $\gamma \in B$ such that $\mathbf{A}$ canonically undercuts γ.*

The relation $\mathcal{U}_K$ is defined as follows:

$$\mathcal{U}_K = \Big\{ (\mathbf{A}, \mathbf{B}) \,\Big|\, \mathbf{A}, \mathbf{B} \in \mathcal{A}_K^L \text{ and } \mathbf{A} \text{ is a canonical undercut of } \mathbf{B} \Big\}$$

It is easy to see that $\mathcal{U}_K$ in general is not symmetric nor transitive.

Example 14. *Recall the set of rules K.*

$$K = \begin{cases} P \rightarrow T_1 > T_2, \\ P \wedge Q \rightarrow T_2 > T_3, \\ \neg Q \rightarrow T_2 > T_3, \\ P \rightarrow T_3 > T_1 \end{cases}$$

It is easy to see that the canonical undercuts of the rule $P \rightarrow T_1 > T_2$ are the following.

$$\langle \{P \wedge Q \rightarrow T_2 > T_3, P \rightarrow T_3 > T_1\}, \qquad P \wedge Q \rightarrow T_2 > T_1 \rangle$$
$$\langle \{\neg Q \rightarrow T_2 > T_3, P \rightarrow T_3 > T_1\}, \qquad P \wedge \neg Q \rightarrow T_2 > T_1 \rangle$$

Once again, we see that there is an intuitive link between the structure of canonical undercuts and the rule they canonically undercut. The following result confirms this intuition.

Proposition 32. *Let $\mathbf{A} = \langle A, \alpha \rangle$ be an argument in $\mathcal{H}_K^L$ and γ a rule. Then, $\mathbf{A}$ canonically undercuts γ iff $A \cup \{\gamma\}$ is a simple $>$-cycle.*

Proof. Right-to-left: on the basis that $A \cup \{\gamma\}$ is a simple $>$-cycle and the fact that $\mathbf{A}$ is an argument, it is easy to show that $\alpha \bowtie \gamma$ and that, therefore, $\mathbf{A}$ undercuts γ. Now assume that there is $B \subset A$ and $\beta \in \mathcal{R}$ such that $\langle B, \beta \rangle$ is in $\mathcal{H}_K^L$ and undercuts γ. By applying Proposition 28 we obtain that there must be $C \subseteq B$ such that $C \cup \{\gamma\}$ forms a simple $>$-cycle, a contradiction. Therefore, $\mathbf{A}$ canonically undercuts γ.

Left-to-right: let $\alpha = \phi \to T_1 \circ T_2$, $\gamma = \psi \to t$ and let P be the set of simple $\circ$-paths from T_1 to T_2 in A. Clearly, $|P| \geq 1$. From Corollary 30 we know that $\vdash \phi \Leftrightarrow \bigvee_{S \in P} PC_S$, and by assumption $\phi \wedge \psi \nvdash \bot$, thus $\psi \wedge \bigvee_{S \in P} PC_S \nvdash \bot$. This means that there exists $S_\psi \in P$ such that $PC_{S_\psi} \wedge \psi \nvdash \bot$.

Now, assume that $P > 1$. For some $S' \in P$, suppose that $PC_{S'} \wedge \psi \vdash \bot$ is true, and let

$$A' = \bigcup_{S \in P \setminus \{S'\}} S$$

$$P' = \{\, S \subseteq A' \mid S \text{ is a simple } \circ\text{-path from } T_1 \text{ to } T_2 \,\}$$

$$\mathbf{A}' = \left\langle A', \bigvee_{S \in P'} PC_S \to T_1 \circ T_2 \right\rangle.$$

Effectively, A' is A without the rules that exist only in S' and, as $P > 1$, $A' \neq \emptyset$ and $P' \neq \emptyset$. From Corollary 30, it should be clear that $\mathbf{A}'$ is a maximally liberal argument. By assumption it must be that $S_\psi \neq S'$, thus $S_\psi \in P'$, therefore $\bigvee_{S \in P'} PC_S \wedge \psi \nvdash \bot$. Thus $\mathbf{A}'$ is also an undercut of γ, contradicting the assumption that $\mathbf{A}$ is a canonical undercut of γ.

Hence, it must be that $PC_S \wedge \psi \nvdash \bot$ for all $S \in P$. Note that, as $|P| > 1$ it is the case that for any $S \in P$, $S \subset A$. Therefore, for any $S \in P$, using Corollary 30 we obtain that $\langle S, PC_S \rangle$ is a maximally liberal argument that undercuts γ, again contradicting the assumption that $\mathbf{A}$ is a canonical undercut of γ.

Therefore it must be that $|P| = 1$. From Proposition 28 we know that there is at least one simple $>$-cycle in $A \cup \{\gamma\}$ thus $A \cup \{\gamma\}$ is a simple $>$-cycle, completing the proof. $\qquad\square$

We can now turn our attention to structures that aggregate arguments and canonical undercuts together, expressing aspects of the conflict intrinsic within a set of rules. The main such structure we will examine is the *argument tree*, a structure that summarises a supporting argument for a particular claim, its canonical undercuts and the canonical undercuts of those, recursively. To define this structure we need some auxiliary definitions first. The following one concerns effectively a branch of an argument tree, and precludes infinite recursion by forcing arguments to use rules in their support that no previous argument in the branch has used.

Definition 33. *Let $\langle A_1, \alpha_1 \rangle, \ldots, \langle A_n, \alpha_n \rangle$ be a sequence of arguments. This sequence is **non-repeating** iff for every $i \leq n$ it is the case that $A_i \not\subseteq \bigcup_{j < i} A_j$.*

Before defining argument trees, we quickly recall definitions relevant to trees in general. Let N be a set, the set of *nodes*, and T a binary relation over N. The tuple $\langle N, T \rangle$ is called a tree iff T forms a directed acyclic graph such that for all nodes n there exists at most one node m such that $(m, n) \in T$, and there exists exactly one node n_r, the *root*, such that there is no node m with $(m, n_r) \in T$. A sequence of nodes $n_1, \ldots, n_k$ such that $n_1 = n_r$ and for any $i < k$, $(n_i, n_{i+1}) \in T$ is called a *branch*.
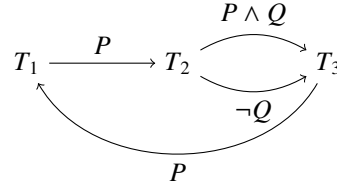
Definition 34. *An **argument tree** is a tuple $\langle N, T, f \rangle$ where $\langle N, T \rangle$ is a tree and f is a function from N to arguments in A_K^L such that:*

- *For all $n_1, n_2 \in N$, if $(n_1, n_2) \in T$ then $(f(n_2), f(n_1)) \in \mathcal{U}_K$.*

- *For any branch $n_1, \ldots, n_k$, the sequence $f(n_1), \ldots, f(n_k)$ is non-repeating.*

We will say that an argument tree $\langle N, T, f \rangle$ **supports** a rule γ iff $\alpha \Vdash \gamma$ where $f(n_r) = \langle A, \alpha \rangle$ and n_r is the root of the tree.

Example 15. *Given the set of rules K,*

$$K = \begin{cases} P \to T_1 > T_2, \\ P \wedge Q \to T_2 > T_3, \\ \neg Q \to T_2 > T_3, \\ P \to T_3 > T_1 \end{cases}$$

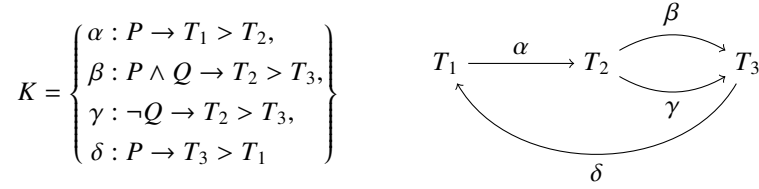*there is the following argument tree supporting $P \wedge Q \to T_1 > T_3$.*

$$\langle \{P \to T_1 > T_2, P \wedge Q \to T_2 > T_3\}, P \wedge Q \to T_1 > T_3 \rangle$$

$$\uparrow$$

$$\langle \{P \wedge Q \to T_2 > T_3, P \to T_3 > T_1\}, P \wedge Q \to T_2 > T_1 \rangle$$

$$\uparrow$$

$$\langle \{P \to T_1 > T_2, \neg Q \to T_2 > T_3\}, P \wedge \neg Q \to T_1 > T_3 \rangle$$

The arrows denote that the argument below the arrow is a canonical undercut of the argument above the arrow.
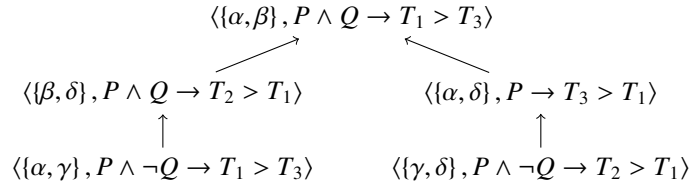
We can see in the example that by its definition, an argument tree does not need to list all the possible interactions between the arguments contained in it, nor does it have to contain all arguments that are relevant. The next definition allows us to express such a condition.

Definition 35. *An argument tree $\langle N, T, f \rangle$ is **full** iff there exists no argument tree $\langle N', T', f' \rangle$ such that $N \subset N'$, $T \subset T'$ and for all $n \in N$, $f(n) = f'(n)$.*

Example 16. *Recall the set of rules K. For convenience we add labels to the rules, and use the labels in the graph on the right, instead of the patient classes.*

$$K = \left\{ \begin{array}{l} \alpha : P \rightarrow T_1 > T_2, \\ \beta : P \wedge Q \rightarrow T_2 > T_3, \\ \gamma : \neg Q \rightarrow T_2 > T_3, \\ \delta : P \rightarrow T_3 > T_1 \end{array} \right\}$$

Continuing the previous example, the following argument tree supporting $P \wedge Q \rightarrow T_1 > T_3$ is full.

$$\langle \{\alpha, \beta\}, P \wedge Q \rightarrow T_1 > T_3 \rangle$$

$$\langle \{\beta, \delta\}, P \wedge Q \rightarrow T_2 > T_1 \rangle \qquad\qquad \langle \{\alpha, \delta\}, P \rightarrow T_3 > T_1 \rangle$$

$$\langle \{\alpha, \gamma\}, P \wedge \neg Q \rightarrow T_1 > T_3 \rangle \qquad \langle \{\gamma, \delta\}, P \wedge \neg Q \rightarrow T_2 > T_1 \rangle$$
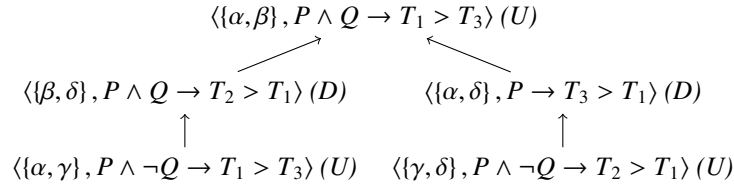
In line with the literature on argumentation [19, 13], we will define a notion of *defeat* which works alongside the notion of *warrant*. Such notions allow us to reason about the collective defeat status of a statement that is supported by some argument from a set of rules.

Definition 36. *A node n in an argument tree $\langle N, T, f \rangle$ is **undefeated** (i.e., not defeated) whenever every node m such that $(n, m) \in T$ is **defeated**.*

Note that if a node n is a leaf-node, i.e., there is no m such that $(n, m) \in T$ then n is trivially undefeated.

Example 17. *Returning to the last example, the argument tree with markings indicating that a node is defeated (D) or undefeated (U) is shown below.*

$$\langle \{\alpha, \beta\}, P \wedge Q \rightarrow T_1 > T_3 \rangle \ (U)$$

$$\langle \{\beta, \delta\}, P \wedge Q \rightarrow T_2 > T_1 \rangle \ (D) \qquad\qquad \langle \{\alpha, \delta\}, P \rightarrow T_3 > T_1 \rangle \ (D)$$

$$\langle \{\alpha, \gamma\}, P \wedge \neg Q \rightarrow T_1 > T_3 \rangle \ (U) \qquad \langle \{\gamma, \delta\}, P \wedge \neg Q \rightarrow T_2 > T_1 \rangle \ (U)$$
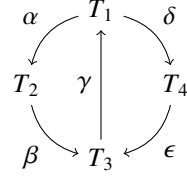
If the root node of a full argument tree is undefeated means that there exists an argument supporting the claim of the root node (and its consequences) that survives all attacks from its canonical undercuts, and that therefore, is *warranted* with respect to the set of rules K. We capture this common definition of warrant below.
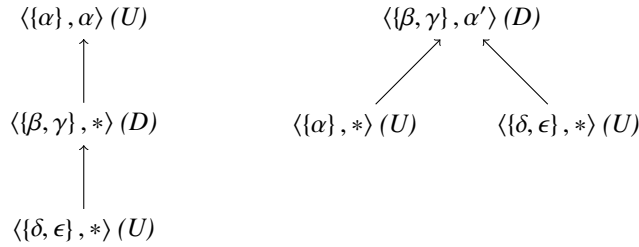
Definition 37. *A rule γ is **warranted** (w.r.t. K) iff there exists a full argument tree $\langle N, T, f \rangle$ that supports γ and whose root node is undefeated.*

In the last example, the root node is undefeated, therefore, the rule supported by the argument tree is warranted. Below we look at another example.

Example 18. *Consider the following graph. For simplicity we omit the exact form of the rules and we simply assume that all combinations of patient classes are consistent.*

Let α' be the rule employing > that is contradictory to α, e.g. $\alpha' = \phi \to T_2 > T_1$ if $\alpha = \phi \to T_1 > T_2$. There is exactly one full argument tree supporting α, shown below on the left, and one full argument tree supporting α', shown below on the right. Note that here we use the symbol $*$ as a shorthand for the claim of a canonical undercut, since as explained earlier, this can be deduced on the basis of the start and end treatments, and the patient classes of the support.

As can be seen, α is warranted and α' is not warranted.

Whilst warrant is a useful argumentation-theoretic notion, it is obviously necessary for the medical expert to decide on its validity in this context. We therefore only propose it as a guide to the user.

7. Discussion and Conclusions

In summary, we have first defined a language that is capable of capturing some of the most important aspects of clinical study results, and representing those as rules. We then defined the logic $\langle \mathcal{R}, \Vdash \rangle$, attempting to model some of the important ways by which practitioners reason with clinical knowledge. This logic allows the drawing of inferences from sets of rules that are incomplete and/or inconsistent. We looked at the decision problem of entailment and provided upper bounds for its worst-case complexity, as well as a characterisation result that can be used as the basis for the construction of theorem-proving algorithms. To aid in analysing conflicting sets of rules, we proceeded to define an argumentation system on top of this logic. We looked at structural properties as well as the computational complexity of several relevant decision problems and examined how concepts from the argumentation literature such as argumentation trees and the notion of warrant can be used with our approach. We feel that systems built around this approach will be of significant value to practitioners.

We motivated our choice to define a logic such as $\langle R, \Vdash \rangle$ and to use an argumentation approach, in the introduction. Here, we look at other approaches to similar problems.

There are similarities between the way we formalise the argumentation system here and the approach of [9]. An *assumption-based framework* (ABF) is a structure consisting of a logic along with a set of formulae called *assumptions* and a mapping of formulae to formulae that effectively produces something akin to the negation of its input (its *contrary*). While there is a similarity between our approach and ABFs in that we define a logic that could plausibly be used in an ABF, there are important differences. First, as is, the language $\mathcal{R}$ does not allow the definition of a mapping of a formula to its contrary, since the negation would potentially require a disjunction of rules (e.g., the negation of $\phi \rightarrow T_1 > T_2$ would be $\phi \rightarrow T_2 > T_1$ *or* $\phi \rightarrow T_1 \sim T_2$). Moreover, the definition of attack in [9] sets the claim of the attacking argument to be the contrary of the assumption being attacked. This is not appropriate in our approach for the reasons cited earlier regarding the contrary mapping, but also due to the way the patient classes interact with each other. For example, given a set of clinical trial results, we may only be able to prove $P \wedge Q \rightarrow T_2 > T_1$ when attacking $P \rightarrow T_1 > T_2$, though the former should still be considered a potentially adequate attack for the latter.

The requirement for transparent reasoning coupled with that for a simple language naturally leads to a monotonic, credulous logic as we argued in the introduction. A different approach that has been studied in the past, is *direct inference*, the aim of which is to somehow assign degrees of belief to statements, given a set of statistical statements such as a set of clinical trial results. So, for example, the statement "if patient a takes treatment T_1, then a's disease-free survival will improve" would be assigned a belief degree based on an interpretation of a set of clinical trial results such as the one we presented in the introduction. Within this context, [3] represents perhaps one of the most important contributions in this line of research. Although potentially applicable in our context, direct inference fails to satisfy our requirement for transparent reasoning. In addition, practitioners employ clinical knowledge in a very different, non-probabilistic way, thereby making user acceptance of such a system more difficult.

Other argumentation-based approaches to common-sense reasoning with applications in the medical domain exist, such as [22] on drug prescription, and [1, 11, 18, 12] on the execution of clinical guidelines. The focus of these approaches is different to the one presented here, since they are aimed at capturing medical guidelines directly in their rules while in this paper we capture clinical trial results as rules. Another obvious difference is that these systems generally rely on first-order logic for representing knowledge, failing therefore to satisfy the requirement for a simple language. Another difference is that in these approaches there is an implied requirement to manually create the arguments that reflect new pieces of information to be handled by the system, while in our approach this is simplified to the addition of the new results as additional rules.

Previous work presented in [23] utilised an argumentation system similarly to this paper, and although its adoption of Defeasible Logic Programming does not allow the use of new inference rules as presented here, we did draw upon its proposed integration of ontologies with argumentation in our definition of the language $\mathcal{R}$. Moreover, $\mathcal{R}$ allows expressing directly a superiority result comparing two treatments, i.e., with a

treatment comparison formula of the form $T_1 > T_2$, whereas in [23] it was generally assumed that the clinical studies involved would compare a treatment against placebo, or no treatment.

There are several possible avenues for further research. We have used propositional logic to express patient classes for simplicity and we have already hinted at the possibility for integrating a description logic that is in use with medical ontologies, an obvious candidate being the underlying logic used in [17]. In addition, it would be interesting to explore extensions of the language where the treatments are themselves parts of an ontology as well.

Further research is also required with respect to the complexity of some of the decision problems we examined in this paper. We have provided upper bounds for the worst-case complexity of entailment, one of the most central components of our framework, placing it within Π_2^p, a class generally considered intractable. Clearly, a more accurate result regarding completeness would be desirable, or, otherwise, a proof that the actual complexity is lower. Based on such results, algorithms and heuristics can be sought, as well as an investigation of average-case complexity.

Another important line of research is the investigation of how a practitioner or a patient can express their values. For example, a doctor may use preferences for expressing that a particular trial is not of adequate statistical power and, therefore, its conclusion may not defeat another, more powerful study's conclusion. Alternatively, a patient at high risk of a terminal cancer may accept strong side-effects that come with a treatment that significantly improves their survival probability. Expressing these values or preferences in our system would greatly improve the usefulness of argumentation.

Finally, from a practical point of view, the utility and usability of our proposal need to be evaluated in a user context. This will necessitate an implementation of some of the key aspects of the framework, the creation of test cases for use in an evaluation study and the recruitment of clinicians other than the one involved in the writing of this paper for participating in the study. We are now at the planning stages of this work.

References

[1] K. Atkinson, T.J.M. Bench-Capon, and S. Modgil. Argumentation for Decision Support. In *Database and Expert Systems Applications, 17th International Conference, DEXA 2006*, volume 4080 of *Lecture Notes in Computer Science*, pages 822–831. Springer, 2006.

[2] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider, editors. *The description logic handbook: theory, implementation, and applications*. Cambridge University Press, New York, NY, USA, 2003.

[3] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. From statistical knowledge bases to degrees of belief. *Artificial Intelligence*, 87(1-2):75–143, 1996.

[4] S.G. Baker and B.S. Kramer. The transitive fallacy for randomized trials: If A bests B and B bests C in separate trials, is A better than C? *BMC Medical Research Methodology*, 2(13), 2002.

[5] T.J.M. Bench-Capon and P.E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.

[6] Ph. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.

[7] C.I. Chesñevar, A. Maguitman, and R. Loui. Logical models of argument. *ACM Computing Surveys*, 32(4):337–383, 2000.

[8] T.J. Cleophas, A.H. Zwinderman, and T.F. Cleophas. *Statistics Applied to Clinical Trials*. Springer, 3rd edition, 2006.

[9] P.M. Dung, R.A. Kowalski, and F. Toni. Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence*, 170(2):114–159, 2006.

[10] B. Fisher, J. Dignam, N Wolmark, A. DeCillis, B. Emir, D.L. Wickerham, J. Bryant, N.V. Dimitrov, N. Abramson, J.N. Atkins, H. Shibata, L. Deschenes, and R.G. Margolese. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *Journal of the National Cancer Institute*, 89(22):1673–1682, 1997.

[11] J. Fox and S. Das. *Safe and Sound: Artificial Intelligence in Hazardous Applications*. MIT Press, 2000.

[12] J. Fox, D. Glasspool, D. Grecu, S. Modgil, M. South, and V. Patkar. Argumentation-Based Inference and Decision Making–A Medical Perspective. *IEEE Intelligent Systems*, 22(6):34–41, 2007.

[13] A.J. García and G.R. Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(1-2):95–138, 2004.

[14] C. Umans M. Schaefer. Completeness in the Polynomial-Time Hierarchy: A Compendium. *SIGACT News*, September 2002.

[15] U.S. National Library of Medicine. `http://www.pubmed.gov/`.

[16] OpenClinical. `http://www.openclinical.org/aisinpractice.html`.

[17] International Health Terminology Standards Development Organisation. Systematized Nomenclature of Medicine-Clinical Terms. `http://www.ihtsdo.org/snomed-ct/`.

[18] V. Patkar, C. Hurt, R. Steele, S. Love, A. Purushotham, M. Williams, R. Thomson, and J. Fox. Evidence-based guidelines and decision support services: A discussion and evaluation in triple assessment of suspected breast cancer. *British Journal of Cancer*, 95(11):1490–1496, 2006.

[19] H. Prakken and G. Sartor. Argument-Based Extended Logic Programming with Defeasible Priorities. *Journal of Applied Non-Classical Logics*, 7(1):25–75, 1997.

[20] H. Prakken and G. Vreeswijk. Logical Systems for Defeasible Argumentation. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer, 2002.

[21] L.J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, 1976.

[22] R. Walton, C. Gierl, P. Yudkin, H. Mistry, M. Vessey, and J. Fox. Evaluation of Computer Support for Prescribing (CAPSULE). *British Medical Journal*, 315:791–795, 1997.

[23] M. Williams and A. Hunter. Harnessing Ontologies for Argument-Based Decision-Making in Breast Cancer. In *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, volume 2, pages 254–261, 2007.