Evaluating arguments based on Toulmin's scheme

Bart Verheij

Department of Metajuridica, Universiteit Maastricht
P.O. Box 616, 6200 MD Maastricht, The Netherlands
bart.verheij@metajur.unimaas.nl, http://www.metajur.unimaas.nl/~bart/

Abstract

Toulmin's argument scheme (1958) represents an influential tool for the analysis of arguments. The scheme enriches the traditional premises-conclusion model of arguments by distinguishing additional elements, like warrant, backing and rebuttal. The present paper contains a formal elaboration of Toulmin's scheme, and extends it with a treatment of the formal *evaluation* of Toulmin-style arguments, which Toulmin did not discuss at all. The present work builds on recent research on defeasible arguments (cf. e.g. the work of Pollock, Loui, Vreeswijk and Dung). More specifically, the author's work on the dialectical logic DEFLOG and the argumentation tool ARGUMED serve as starting points.
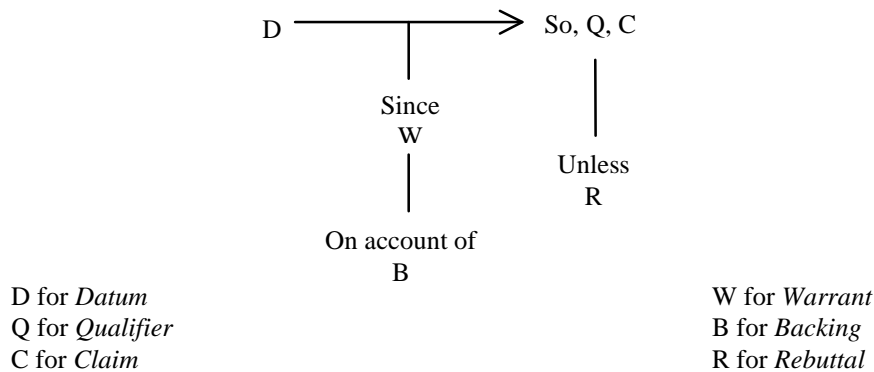
1   Introduction

In his book 'The Uses of Argument', Stephen Toulmin (1958) has argued that arguments need to be analyzed using a richer format than the traditional one of formal logic in which only premises and conclusions are distinguished. He has proposed a scheme that next to *data* and *claim* distinguishes between *warrant*, *backing*, *rebuttal* and *qualifier*.

As an illustration, Toulmin discusses the claim that Harry is a British subject. The claim can be supported by the datum that Harry was born in Bermuda. That there is a connection at all between datum and claim is expressed by the warrant that a man born in Bermuda will generally be a British subject. On its turn, the warrant can be supported by the backing that there are certain statutes and other legal provisions to that effect. The warrant does not have total justifying force, so the claim that Harry is a British subject must be qualified: it follows presumably. Moreover there are possible rebuttals, for instance when both his parents were aliens or he has become a naturalized American.

Schematically, the result is as follows (Toulmin 1958, p. 105):
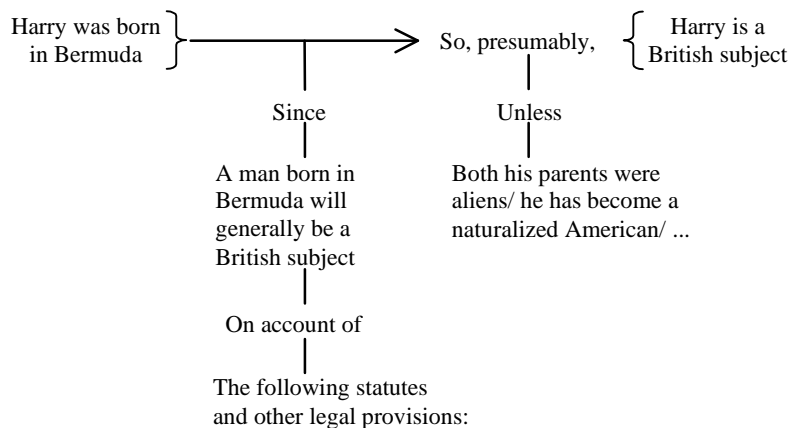
Toulmin's argument scheme has had a continuing influence on argumentation researchers (cf., e.g., Van Eemeren *et al.* 1996, pp. 129-160; Bench-Capon 1997). Its general form is thus (Toulmin 1958, p. 104):

```
D  ─────────────┬──────────▶  So, Q, C
                │                  │
              Since                │
                W                  │
                │               Unless
                │                  R
          On account of
                B
```

D for *Datum*	W for *Warrant*
Q for *Qualifier*	B for *Backing*
C for *Claim*	R for *Rebuttal*

The Datum consists of certain facts that support the Claim. The Warrant is an inference license according to which the Datum supports the Claim, while the Backing provides on its turn support for the Warrant. A Rebuttal provides conditions of exception for the argument, and the Qualifier can express a degree of force that the Datum gives to the Claim by the Warrant.

Good points of Toulmin's work were his emphasis on the following:

- In argumentation, the warrants of arguments (in the sense of inference licenses) can be at issue and their backings can differ from domain to domain.
- Arguments can be subject to rebuttal in the sense that there can be conditions of exception.
- Arguments can have qualified conclusions.
- Other kinds of arguments than just that based on the standard logical quantifiers and connectives (for all *x*, for some *x*, not, and, or, etc.) need to be analyzed.

Since the appearance of Toulmin's book, all of these points have found increasing support in different research communities (under the direct influence of Toulmin or independently).

Notwithstanding Toulmin's critical stance towards formal logic, in the present paper, a *formal* elaboration of Toulmin's central ideas will be given. It is shown that Toulmin's central ideas can well be set out in a formal way by the use of techniques of today's formal logic.

It goes without saying that the game of formalization is here *not* played for mathematical pleasure. The main reason for formalization is that it allows the repair of an omission in Toulmin's work. Toulmin's omission is that he has only discussed the *structure* of arguments (in terms of the roles of the different kinds of elements of arguments), but has not paid attention to the *evaluation* of arguments. In other words, he does not provide an analogue of logical *validity* as an evaluation criterion for arguments - perhaps because he does not believe there to be one.

This is a major omission, especially in view of the role of rebuttals in his argument scheme: clearly an argument for a particular claim should be evaluated differently when there (actually) is a rebuttal and when there is none. Using Toulmin's example above, it depends on the fact whether Harry's parents were aliens, whether the datum that Harry is born in Bermuda justifies the claim that he is a British subject. Assuming that Harry's parents were aliens, the datum does *not* justify the claim, while assuming that they weren't aliens, the datum will normally justify the claim (unless there is another

rebuttal, such as that Harry has become a naturalized American). As said, perhaps Toulmin did not believe there to be an evaluation criterion analogous to logical validity. Here it is shown that there is such a criterion by providing one: it is based on the idea of *dialectical interpretation* (cf. Verheij 2000a, b). Briefly, the statements in an argument are evaluated with respect to the argument's assumptions. Statements can be justified (e.g., when there is a justifying reason for them), defeated (e.g., when there is a defeating reason against them) or neither.

Of course it is impossible to retain all of Toulmin's central ideas in their exact original form - some ideas will be adapted and extended. The task set here is one of reconstruction with a contemporary eye, and thus by necessity involves an *interpretation* of Toulmin's ideas. Still the goal is that all elements of Toulmin's argument scheme will find a recognizable place in the present elaboration.

The starting point for the reconstruction of Toulmin's scheme is a particular theory of dialectical argumentation, called DEFLOG. It is related to my work on automated argument assistance (see, e.g., Verheij 1999a on the ARGUMED system). In the present paper, the relevant parts of DEFLOG are introduced alongside the reconstruction of Toulmin's argument scheme. For a more extensive account of DEFLOG, the reader may want to consult Verheij (2000a, b). The present work builds on recent research on defeasible and dialectical argumentation (cf. e.g. the work of Pollock 1987, Loui 1998, Vreeswijk 1997 and Dung 1995).

2   Elaborating on Toulmin's argument scheme

In the following, the elements of Toulmin's scheme will be reconstructed, in a way that is formally explicit. A step-by-step strategy will be followed. Simple arguments will be analyzed in terms of primitive notions like statements, justifying reasons etc.

2.1   Datum and claim

The first step in the present reconstruction is to consider Datum and Claim. In Toulmin's (1958) own words, we have

> 'one distinction to start with: between the *claim* or conclusion whose merits we are seeking to establish (C) and the facts we appeal to as a foundation for the claim - what I shall refer to as our *data* (D)' (p. 97).

Let's look at Toulmin's example: if we seek to establish the claim that Harry is a British subject, we could for instance appeal to the datum that Harry was born in Bermuda. In an ordinary language argument this could be expressed thus:

> Harry was born in Bermuda. So he is a British subject.

Already at this early stage, it is convenient to adopt some conventions that elaborate on Toulmin's. The first is the convention that arguments consist of one or more *statements* that are expressed by *sentences*. In the example we find the two statements that Harry is a British subject and that Harry was born in Bermuda, expressed by the sentences 'Harry is a British subject' and 'Harry was born in Bermuda', respectively.

The second convention is to distinguish between two different *roles* of statements in arguments: some are *assumptions*, others *issues*. In the example, it is assumed that Harry was born in Bermuda, and at issue whether Harry is a British subject. If it were assumed that Harry is a British subject, it would not be needed to provide a reason for it. If the datum were itself not assumed, i.e., if it were at issue whether Harry was born in Bermuda, that issue would have to be established by further argument.

The third convention concerns the *evaluation* of the statements in an argument. The idea is that the statements in an argument can be evaluated, depending on the

information expressed in the argument. For instance, the assumption that Harry was born in Bermuda is - since it is an assumption - taken to be justified. Also the issue that Harry is a British subject is justified, but only since there is a reason justifying it, viz. that Harry was born in Bermuda. At present, two evaluation statuses are distinguished: statements can be justified or they can be unevaluated. If for instance the statement that Harry was born in Bermuda were itself at issue, that statement and the statement that Harry was born in Bermuda, would not be justified. (Later we will also distinguish statements that are defeated. See section 2.4.)

This brings us to the fourth convention, that concerns *support by reasons*. One point of an argument like 'Harry was born in Bermuda. So he is a British subject' is that it expresses that the statement that Harry was born in Bermuda *supports* the statement that he is a British subject. It not only expresses the assumption that Harry was born in Bermuda and the issue that Harry is a British subject, but also that there is a support relation between the two statements. In other words, the argument implies that there is a conditional relation between the two statements, that can - for instance - be expressed by the compound sentence 'If Harry was born in Bermuda, he is a British subject'. (As yet no further interpretation of the conditional 'If ..., (then) ...' is presumed, and certainly not that of the notorious material conditional of standard logic. Cf. Hage 1997, Prakken 1997, Verheij 1996. In order to prevent confusion with the latter conditional, a dedicated arrow ~> is used here.) Statements expressed by sentences of the form 'If ..., (then) ...' can themselves be assumed or at issue, just like all other statements. For instance, an argument like 'Harry was born in Bermuda. So he is a British subject' can be considered to imply the assumption that if Harry was born in Bermuda, he is a British subject. A conditional assumption of this form can be referred to as the *implicit assumption* of the argument.[1]

According to the present conventions, the argument 'Harry was born in Bermuda. So he is a British subject' implies the following:

It is assumed that Harry was born in Bermuda.
It is at issue whether Harry is a British subject.
It is assumed that if Harry was born in Bermuda, he is a British subject.
It is justified that Harry was born in Bermuda.
It is justified that Harry is a British subject.
It is justified that if Harry was born in Bermuda, he is a British subject.

The issue whether Harry is a British subject, is justified since the statement that Harry was born in Bermuda is a justifying reason for it. In general, a statement D is said to be a *justifying reason* for another statement C if the statements that D and that if D, then C are both justified.

The above can be summarized in the following graphical representation of the argument (cf. Verheij 1999a, 2000b):
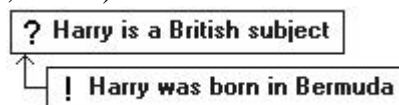
The exclamation mark indicates an assumed statement, the question mark a statement that is at issue. The implicit assumption (that if Harry was born in Bermuda, he is a British subject) is depicted by the arrow. That all statements are justified is indicated by the dark bold font. If it were at issue whether Harry was born in Bermuda we would have the following:
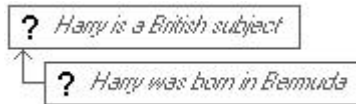
Now both statements are unevaluated (as is indicated by the light italic font) since for neither there is a justifying reason. The light color of the arrow indicates that it is not justifying.

By the present conventions an argument like 'Harry was born in Bermuda. So he is a British subject' is closely linked to the following formal derivation based on *Modus ponens* (From φ and φ ~> ψ, conclude ψ):

Harry was born in Bermuda. Harry was born in Bermuda ~> Harry is a British subject.

Harry is a British subject.

The two premises of the derivation correspond to the two assumptions of the datum-claim argument (one of them implicit), the conclusion of the derivation to the issue. The *Modus ponens* form of the derivation makes explicit how the evaluation status is transferred from the assumptions to the issue: when the statements that φ and that φ ~> ψ are both justified, then the statement that ψ is also justified.

It is tempting to read the *Modus ponens*-derivation itself as the following informal argument:

> Harry was born in Bermuda. If Harry was born in Bermuda, then he is a British subject. So he is a British subject.[2]

However, this informal argument is clearly *another* argument than the example 'Harry was born in Bermuda. So he is a British subject' (one difference being that while the shorter example could occur in an ordinary argumentative text, the latter can only occur in a logic-oriented text).

Until now, we have for convenience omitted the qualifier in Toulmin's example. In the example, the datum 'Harry was born in Bermuda' does not support the claim 'Harry is a British subject' in an unqualified manner, but only *presumably*. Concerning such *qualifiers* as Toulmin distinguishes them in his scheme, we will here be very brief. A qualifier is simply concerned as a modal operator on statements. As a result, Toulmin's qualifier will be considered as being a *part* of the sentence that expresses the claim supported by the datum. In Toulmin's example, we will take it that the datum supports the claim 'Presumably, Harry was born in Bermuda'.[3] As a result, the relevant conditional is 'If Harry was born in Bermuda, then presumably he is a British subject' instead of the one used above. No semantic assumptions about the qualifiers are assumed here, in style with Toulmin's loose treatment. For instance, he seems to use the qualifiers 'presumably' and 'generally' more or less interchangeably (even though the former suggests an epistemical and the letter an ontological point of view).

2.2 Warrant

The next step is to consider the *warrants* of arguments. Toulmin (1958) describes warrants as 'general, hypothetical statements, which can act as bridges [between datum and claim, BV], and authorise the sort of step to which our particular argument commits us' (p. 98). In other words, warrants are inference licenses that express that certain claims follow from certain data. The warrant of Toulmin's example on British citizenship is the statement 'A man born in Bermuda will generally be a British subject'.

Toulmin thinks of warrants as *general* inference licenses: not only if *Harry* was born in Bermuda, he is a British subject, but also if *John* was born in Bermuda, he is a British subject, etc. It seems that for Toulmin the statement 'A man born in Bermuda will generally be a British subject' is equivalent to a scheme like the following, were *Person* stands for any person:

If *Person* was born in Bermuda, then generally *Person* is a British subject.

Here '*Person* was born in Bermuda' and '*Person* is a British subject' can be taken as a *generic* datum and claim, respectively. The apparent equivalence for Toulmin between rule statements like 'A man born in Bermuda will generally be a British subject' and conditional schemes like the one above is for instance suggested by his emphasis on the bridge-like character of warrants (e.g., p. 98 as cited above, and p. 105, where he discusses the difference between warrants and backings). The discussion of the formal validity of arguments of the type 'D. W. So C' (p. 118f., especially p. 119), where D is the datum, W the warrant and C the claim, strengthens this suggestion. The formal structure of such an argument could be made explicit in the following way:

$D(t)$. $D(x) \rightsquigarrow C(x)$. So $C(t)$.

Here x is a variable instantiated by t. Toulmin mentions a variant of this formal structure in the style of classical syllogisms:

X is an A. All A's are B's. So X is a B.

Somewhat confusingly, Toulmin is not fully consistent in his discussion of warrants: occasionally, he seems to refer to an *instance* of a conditional scheme (like the one above) as a warrant, for instance when he mentions the short form 'If D, then C', where D and C are a particular (and not a generic) datum and claim (p. 98). However, as said, at other places Toulmin unambiguously emphasizes the generality of warrants (e.g., in the description of warrants cited above, p. 98, but also on p. 100, where he states that warrants certify the soundness of all arguments of the appropriate type).

Summarizing, it is convenient to distinguish the following three:

A man born in Bermuda will generally be a British subject.
If *Person* was born in Bermuda, then generally *Person* is a British subject.
If Harry was born in Bermuda, then generally he is a British subject.

The first is the ordinary language expression of a warrant. It is a rule statement that connects certain kinds of data to certain kinds of claims. It is these that in the present paper will be referred to as warrants (which is in agreement with Toulmin's conception of warrants and with most of Toulmin's examples).
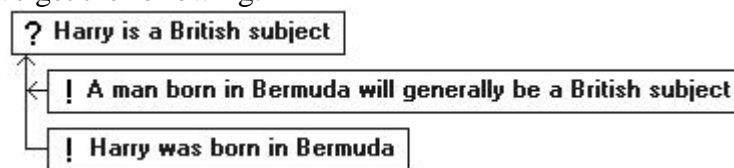
The second is the conditional scheme that is the formal explication of the bridge-like connection resulting from warrants. In a sense (and as said apparently according to Toulmin) an ordinary language warrant is equivalent to such a conditional scheme: it may be taken that each implies the other. Moreover, it seems straightforward to write algorithms that translate the one to the other (at least for significant subsets of ordinary language examples of warrants, e.g., for simple generic sentences like 'Bermudans are British subjects' and 'Thieves are punishable'). It should be noted however that in ordinary language argument, the conditional schemes will themselves never occur. They

are merely convenient constructs in order to summarize the range of argument steps as they are licensed by the warrant that corresponds to the scheme.

The third, 'If Harry was born in Bermuda, then generally he is a British subject', is the normally implicitly assumed conditional that expresses that a particular datum implies a particular claim. Conditionals like these are the instances of the conditional schemes that correspond to ordinary language warrants.[4]

The only formal connection between the three that will be used below is that between the warrant (in the sense of an ordinary language rule statement) and the implicit conditional that forms the bridge from datum to claim. It can be expressed as the nested conditional 'If W, then if D, then C', or formally $W \sim> (D \sim> C)$. The nested conditional expresses that it follows from the warrant that the claim follows from the datum. This nested conditional is normally left implicit in an argument, just like the conditional 'If D, then C'.

Graphically we get the following:

| ? Harry is a British subject |
| ! A man born in Bermuda will generally be a British subject |
| ! Harry was born in Bermuda |

That the statement that a man born in Bermuda will generally be a British subject, is a warrant is visualized by an arrow pointing to an arrow: here the arrow from the warrant statement points to the arrow representing the implicit conditional that if Harry was born in Bermuda, he is a British subject.

The transfer of evaluation status from assumptions to issues can be made explicit as the following two-step *Modus ponens* derivation:

$$
\frac{D. \quad \dfrac{W. \quad W \sim> (D \sim> C).}{D \sim> C.}}{C.}
$$

D and W are the explicit assumptions of the datum-claim-warrant argument, and C its explicit issue. $W \sim> (D \sim> C)$ is its implicit assumption, and $D \sim> C$ (so to speak) its implicit issue.
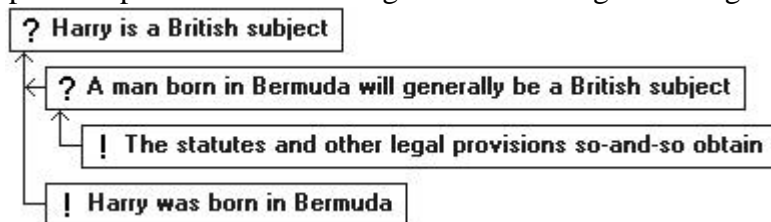
2.3 Backing

The next important element of arguments distinguished by Toulmin is that of the *backing* of warrants. Backings provide support for warrants. They become relevant when the legitimacy of the range of arguments as licensed by a warrant is challenged. In Toulmin's example, the warrant that a man born in Bermuda will generally be a British subject, is supported by appeal to particular statutes and legal provisions (which are not explicitly mentioned by Toulmin). Formally, the relation between backing and warrant is the same as the relation between datum and claim,[5] and requires nothing new. An argument 'B. So W' (where B is a backing of some warrant B) can be analyzed just like an argument 'D. So C'. In the former case, the implicit conditional has the form 'If B, then W'. In the example, it becomes 'If the statutes and other legal provisions so-and-so obtain, then a man born in Bermuda will generally be a British subject'.

Even though datum and backing formally play a related role, Toulmin notes a relevant difference between them: there are arguments containing datum but without explicit backing (like in 'D. So C'), while there are no arguments containing backing

but lacking datum. The occurrence of a backing *presupposes* the occurrence of a datum (and claim). Toulmin also emphasizes the difference between backing and warrant: backings can be categorical statements of fact just like data, while warrants always are general bridge-like statements (cf. p. 105). A central point in Toulmin's book is that the kinds of backings as they occur in different fields of argument vary. Among Toulmin's examples of backings are statutes and acts of Parliament, statistical reports, appeals to the results of experiments and references to taxonomical systems. All can provide the backing that warrant the arguments as they are acceptable in particular fields.

Here is a graphical representation of an argument involving a backing:

If the backing were itself at issue, we would get the following:
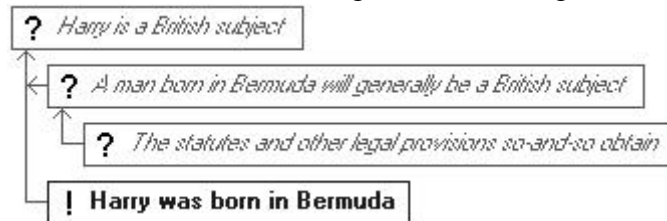
Neither the statement that Harry is a British subject nor the warrant statement that a man born in Bermuda will generally be a British subject, is now justified since there is no justifying reason for them.

It should be noted that the reconstruction of Toulmin's analysis of arguments with datum, (qualified) claim, warrant and backing as above is hardly a deviation of standard logical notions. This is in contrast with Toulmin's claims. As has been noted by others, the distinction of data, claim, qualifier, warrant and backing can be dealt with in a way that is not too far from standard logic, while retaining the differences between them. In the present reconstruction, the technical tools used are essentially a Modus ponens-validating conditional, variables and modal qualifiers. The main differences between Toumin's treatment and standard logic arise from differences in emphasis with respect to philosophical and argumentation-theoretical starting points (such as the mentioned field-dependency of the backings of warrants and the distinction of different roles of statements in an argument).

2.4 Rebuttal

A genuine and radical deviation from standard logic is required by Toulmin's notion of *rebuttals*. However, Toulmin hardly elaborates on the nature of rebuttals. As Toulmin puts it, rebuttals involve conditions of exception for the argument (p. 101). Apparently, for Toulmin, rebuttals can have several functions. For instance, rebuttals can 'indicate circumstances in which the general authority of the warrant would have to be set aside' (p. 101), but can also be (and for Toulmin apparently equivalently) 'exceptional circumstances which might be capable of defeating or rebutting the warranted conclusion' (p. 101). On p. 102, he also speaks about the applicability of a warrant in connection with rebuttals. In other words, Toulmin speaks of the defeat (or rebutting) of the conclusion, of the applicability of the warrant and of the authority of the warrant, in a rather loose manner, without further distinction. Toulmin is unclear about the relation

of these seemingly different situations. Here the three will be distinguished, in a way that naturally fits the reconstruction of the other elements of Toulmin's scheme above, as follows. If we look at the warrant-datum-claim part of Toulmin's scheme (which in our reconstruction includes the qualifier), there are five statements that can be argued against:

1. The datum D
2. The claim C
3. The warrant W
4. The implicit conditional 'If D, then C' that expresses the bridge from datum to claim.
5. The implicit conditional 'If W, then if D, then C' that expresses the bridge between warrant and the previous implicit conditional.

Reasons against any of these statements can be seen as a *kind of rebuttal* of an argument that consists of warrant, datum and claim. The first three speak for themselves, and are clearly all different. An argument against the datum that Harry was born in Bermuda (for instance by claiming that Harry was born in London) differs from an argument against the claim that Harry is a British citizen (for instance by claiming that Harry has become a naturalized American) and from an argument against the warrant that a man born in Bermuda will generally be a British subject (for instance by claiming that those born in Bermuda are normally French).

An argument against the fourth kind of statement (the first implicit conditional), can be regarded is an attack of the connection between datum and claim. Such attacks have been dubbed *undercutting defeaters* by Pollock (1987). Harry having become a naturalized American could be an argument against the connection between Harry being born in Bermuda and Harry being a British citizen.

An argument against the fifth kind of statement can be regarded as attack against the warrant's applicability: normally the warrant can justify the conditional that connects datum and claim, but since there is a rebuttal, the warrant does not apply. In other words, when the implicit conditional if W, then if D, then C, is not justified, the warrant that normally gives rise to a bridge between datum of type D and claim of type C, does for the actual datum D and C at hand *not* give rise to such a bridge. For instance, Harry's parents both being aliens could well be an argument against the applicability of the warrant that a man born in Bermuda will generally be a British subject.[6]

The three situations to which Toulmin attaches the term rebuttal (defeat of the conclusion, of the applicability of the warrant and of the authority of the warrant) are among these five kinds of rebuttals, viz. the second, fifth and third, respectively. The other two kinds of rebuttals of a warrant-datum-claim argument, viz. that of the first and fourth kind, are apparently not mentioned by Toulmin.

We now turn to a major omission of Toulmin's discussion: a topic he does not address is the *effect* of rebuttal on the evaluation status of the statements in an argument. Interestingly, this has turned out a notorious, but rewarding topic for logical study (cf. the work on formalizing defeasible and dialectical argument; see e.g. Chesnevar et al. *to appear*, Prakken en Vreeswijk *to appear*).

In the present approach, a third evaluation status is introduced in order to analyze the effect of rebuttal on the evaluation status: statements can not only be justified or unevaluated, but also *defeated*.[7] The latter status applies when there is a defeating reason against the statement. For instance, in the example, it can be defeated that Harry

is a British subject in light of the reason against it that he has become a naturalized American, even though Harry was born in Bermuda.

One difficulty is the fundamental *nonmonotonicity* of the resulting logic: when assumptions are added it can occur that previously justified statements are no longer justified (cf. in a more general logical setting Gabbay et al. 1994). In standard logic, when sentences are added to an initial set of sentences the truth of which is assumed, the set of implied truths never gets smaller. Additional assumptions normally allow more conclusions to be drawn, and never less. In contrast, in a dialectical context, as for instance required by Toulmin's analysis of argument, where there is not only support but also attack, statements that are initially justified with respect to a set of assumptions can become defeated (or unevaluated) in light of additional assumptions.

Toulmin's example can be used as an illustration (though he does not discuss effects on evaluation): when it is not assumed or otherwise justified that Harry's parents were aliens or that he has become a naturalized American, i.e., when the possible rebuttals are not effective, it is justified that Harry is a British subject on the assumption that he was born in Bermuda and that a man born in Bermuda will generally be a British subject. However, when one of the rebuttals is effective, e.g., when it is assumed that Harry's parents were aliens, Harry being a British subject no longer follows as a justified statement, but is unevaluated (given that there are no *other* reasons for or against it).

Just as justifying reasons for a statement can make an issue justified, defeating reasons against a statement can make it defeated. Formally, it is convenient to distinguish statements of the form 'It is defeated that ...', where the dots indicate a sentence expressing another statement. In this way, *attack by reasons* can be dealt with in analogy with support by reasons (cf. section 2.1). Whereas the support relation between datum D and claim C is expressed by the (usually implicit) conditional 'If D, then C', the attack relation between a rebuttal R and a claim C is expressed by the statement 'If R, then it is defeated that C' (for a rebuttal of the second kind above). The latter will formally be denoted as R ~> xC (where xC expresses that it is defeated that C and ~> is the conditional used to express support)[8] or more briefly as R ~x C. In general, a statement R is said to be a *defeating reason* against another statement C if the statement that R, and the statement that if R, then it is defeated that C, are both justified (cf. the notion of a justifying reason in section 2.1).

It can now be defined which statements are supported by a set of assumptions (using the simple logical language with connectives ~> and x as used above) and which attacked. A statement S is *supported* by the assumptions if the statement S is itself an assumption or follows from the assumptions by the repeated application of Modus ponens (From φ and φ ~> ψ, conclude ψ). A statement S is *attacked* by the assumptions if the statement that the statement is defeated (i.e. the statement that xS) is an assumption or follows from the assumptions by the repeated application of Modus ponens. When a set of assumptions is *conflict-free*, i.e., when there are no statements that are both supported and attacked by the assumptions, it is now easy to define the *evaluation status* of statements with respect to the assumptions, as follows. A statement is *justified* (with respect to the conflict free set of assumptions) when it is supported by the assumptions. A statement is *defeated* (with respect to the conflict free set of assumptions) when it is attacked by the assumptions. Any other statement is *unevaluated* (with respect to the conflict free set of assumptions).

Evaluation in the more important general case of possibly conflicting sets of assumptions is subtle (and those who are not formally inclined may want to pass over this). The definition for conflict-free sets does not work since when there is a conflict there are statements that are both supported and attacked by the assumptions. Therefore

in the general case, the assumptions are considered to be *defeasible* in the sense that not all statements assumed to be justified need also turn out to be evaluated as justified when the whole set of assumptions is evaluated. The idea is that some of the assumptions are themselves defeated since they are attacked by other assumptions. The formal definition is as follows:

> Let Δ be a set of sentences defeasibly assumed to be justified, and let J be a subset of Δ (possibly equal to Δ). Then J *dialectically interprets* Δ when the following hold:
> 1. J is conflict-free.
> 2. Any statement in Δ that is not in J is attacked by J.
>
> A *dialectical interpretation* of Δ is a pair of sets (Supp(J), Att(J)), where J dialectically interprets Δ and Supp(J) consists of the statements supported by J, and Att(J) of the sentences attacked by J. The statements supported by J are said to be *justified* in the dialectical interpretation of Δ corresponding to J, those attacked by J *defeated*.

There is a lot to say about this definition of dialectical interpretation. For instance, while some theories have a unique dialectical interpretation, others have none or several. The reader is referred to Verheij (2000a, b). Here we will confine ourselves to examples related to Toulmin's argument scheme.

Let's again consider Toulmin's example concerning Harry. Assume as datum that Harry was born in Bermuda (D) and as rebuttal that Harry has become a naturalized American (R). It is at issue whether Harry is a British subject (the claim C). For simplicity we first forget about warrant and backing. The relation between datum and warrant is implicitly assumed, and can formally be expressed as D ~> C. The relation of the rebuttal with datum and claim is assumed to be of the *fourth* kind above, i.e., as an attack against the implicit conditional that If D, then C. Therefore the relation is expressible as R ~> x(D ~> C). Formally, the set of assumptions Δ consists of four sentences, viz. D, R, D ~> C and R ~> x(D ~> C). The set contains a conflict: D ~> C is both supported (since it is an assumption) and attacked (by applying Modus ponens on R and R ~> x(D ~> C)). The set of assumptions has a dialectical interpretation however: in it, the assumption D ~> C is defeated, while the other three are justified. The claim C is unevaluated since it is neither justified nor defeated in the dialectical interpretation. This is in accordance with the intuition that there is neither a justifying reason for C, nor a defeating reason against it. It is not hard to check that this is the *only* dialectical interpretation. If the rebuttal R is *not* assumed, i.e., if we start with the three assumptions D, R, D ~> C and R ~> x(D ~> C), the situation is simpler since there is no conflict. In the only dialectical interpretation, all three assumptions are justified. Also C is justified since D is a justifying reason for it.

Graphically, we get the following for the case that the rebuttal is assumed:
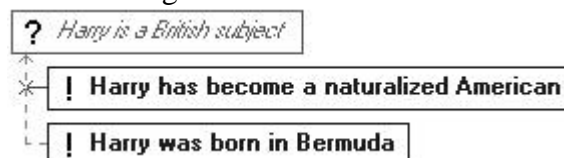
The arrow ending in a cross indicates that the rebuttal attacks the implicit conditional that if Harry was born in Bermuda, he is a British subject. It is now not justified that Harry is a British subject since it is an issue for which there is no justifying reason. (It is neither defeated.) The implicit conditional is defeated since there is a defeating reason against it, viz. the rebuttal.

If it were at issue whether Harry has become a naturalized American, the following would result:

> ? Harry is a British subject
>
> ─ * ─ ? *Harry has become a naturalized American*
>
> ! Harry was born in Bermuda
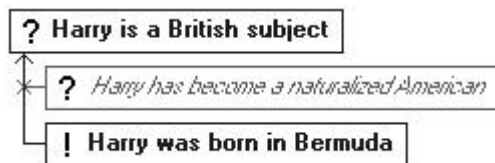
Now it is justified that Harry is a British subject, since the rebuttal does not have effect and the implicit conditional is not defeated. The graphical representation clearly demonstrates that rebuttals can have a decisive effect on the evaluation of arguments and the statements in them: when the rebuttal is at issue, the claim is justified, but when the rebuttal is assumed, the claim is not justified.

Note that Toulmin did not distinguish between kinds of rebuttals. In the previous paragraphs, the rebuttal R that Harry has become a naturalized American was considered to be a rebuttal of the fourth kind. Let's now analyze the example considering R as a rebuttal of the *fifth* kind, i.e., it attacks the applicability of the warrant in Harry's case. Formally, this is denoted thus: R ~> x(W ~> (D ~>C))). In words, if the rebuttal R is justified, then it is defeated that the conditional that if D then C, follows from the warrant W. As a result, in case of R, W does not imply that D can support C. The following are assumed:
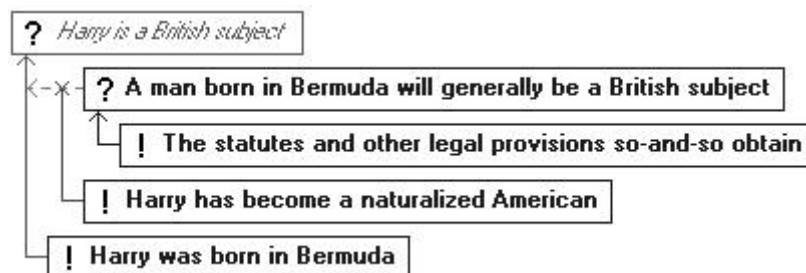
D: Harry was born in Bermuda
W: A man born in Bermuda will generally be a British subject
R: Harry has become a naturalized American
B: The statutes and other legal provisions so-and-so obtain

It is at issue whether Harry is a British subject (C). The assumed logical connections (that are normally left implicit) are as follows:

B ~> W
W ~> (D ~> C)
R ~> x(W ~> (D ~>C)))

The set of assumptions contains a conflict: the nested conditional W ~> (D ~> C) that expresses that W implies that D can support C, is both supported (as an assumption) and attacked (since it follows from R and R ~> x(W ~> (D ~>C)))). In the unique dialectical interpretation, only W ~> (D ~> C) is defeated (by the defeating reason R against it), while the other six assumptions are justified. As a result, the statements that D ~> C and that C are unevaluated since there is no justifying reason for them.

Assuming that the rebuttal is of the fifth kind, we get the following graphical representation:

> ? *Harry is a British subject*
>
> ←─×─ ? A man born in Bermuda will generally be a British subject
>
> ! The statutes and other legal provisions so-and-so obtain
>
> ! Harry has become a naturalized American
>
> ! Harry was born in Bermuda

The arrow ending in a cross indicates that the rebuttal that Harry has become a naturalized American attacks that the conditional that if Harry was born in Bermuda, he

is a British subject, follows from the warrant statement. It is now neither justified nor defeated whether Harry is a British subject since there is no justifying reason for it nor a defeating reason against it.

If the rebuttal were itself at issue, the evaluation would be different, and it would be justified that Harry is a British subject:
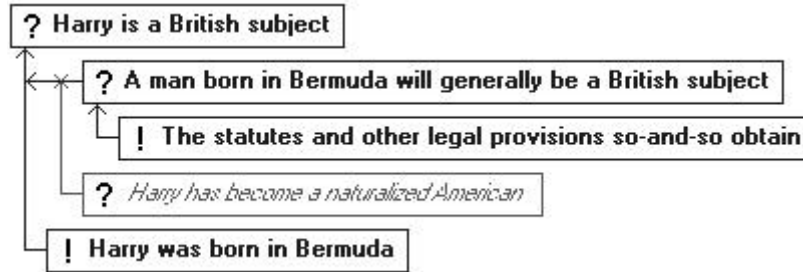
An important phenomenon related to the evaluation of arguments as discussed above is the *reinstatement* of a claim. This occurs when a claim is at first justified, for instance in the light of an argument without rebuttal, subsequently unjustified, for instance when a rebuttal is added, and finally again justified, when the rebuttal is on its turn shown to be unjustified. The latter might occur when the rebuttal is itself the claim of a scheme that is extended by a rebuttal.

Toulmin does not discuss this phenomenon of reinstatement (simply since he does not discuss the evaluation of arguments at all). An example based on Dutch tort law is the following. It may be claimed that someone has committed an unlawful act on the basis of the datum that he has violated a property right. An example of the violation of a property right is the breaking of someone else's window. In Dutch tort law, a warrant could be that violations of property rights are unlawful, which can be backed by article 6:162 of the Civil Code. If the argument is limited to the present information, the claim can be thought of as justified (on the basis of the argument). It can subsequently be argued that there is a rebuttal: there is a ground of justification for the act since the person obeyed an official's command. (Note that the rebuttal is here the claim of a second scheme with the obeying of the official's command as datum.) At this stage, the claim that an unlawful act has been committed is no longer justified on the basis of the argument involved. The argument can further be extended by adding a rebuttal to the second scheme: the official giving the command was unauthorized to do so. As a result the claim of the second scheme (viz. the existence of a ground of justification) is no longer justified. Since it is at the same time the rebuttal of the first scheme, it loses its rebutting effect in that scheme, and the claim that an unlawful act has been committed, is again justified.

Verheij (1999a, 2000b) gives more examples of the evaluation of arguments and the reinstatement of statements.

3   Conclusion

In the present paper, Toulmin's argument scheme has been formally reconstructed. An important omission of Toulmin's treatment has been repaired by providing an account of the evaluation of Toulmin-styled arguments. It has been discussed how the evaluation status of the statements in an argument is determined by a dialectical interpretation of the assumptions of the argument. In such an interpretation, assumptions are taken to be defeasible, so that they need not all be evaluated as justified in an interpretation: some can be defeated, viz. when they are attacked by a defeating reason against them.

It can be concluded from the present work that the main deviation from standard logic of Toulmin's argument scheme is the element of rebuttal. Since Toulmin, the logic

of rebuttal has deservedly received widespread attention (under the name of defeasible argumentation). The present paper is a continuation of that line of research.

Another important point of Toulmin's work is the field-dependency of the backings of the warrants underlying arguments. Toulmin has stressed that the rules of argument can vary from domain to domain. This raises a fundamental and a practical question. The fundamental question is what remains of logic when the rules of argument are variable. The practical question is to determine the rules of argument in particular domains. Both questions still require substantial research. Beginnings of answers to these important questions can be found in the works of Hage (1997), McBurney and Parsons (2000), Verheij (1999b, 2001) and Walton (1996).

References

Bench-Capon, T.J.M. 1997. "Argument in Artificial Intelligence and Law", *Artificial Intelligence and Law* 5: 249-261.

Chesnevar, C.I., A.G. Maguitman and R.P. Loui. To appear. "Logical Models of Argument', *ACM Computing Surveys*.

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

Eemeren, F.H. van, R. Grootendorst, and F. Snoeck Henkemans. 1996. *Fundamentals of Argumentation Theory. A Handbook of Historical Backgrounds and Contemporary Developments.* Mahwah (New Jersey): Lawrence Erlbaum Associates.

Gabbay, D.M., C.J. Hogger, and J.A. Robinson. (eds.) 1994. *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3. Nonmonotonic Reasoning and Uncertain Reasoning.* Oxford: Clarendon Press.

Hage, J.C. 1997. *Reasoning with Rules. An Essay on Legal Reasoning and Its Underlying Logic.* Dordrecht: Kluwer Academic Publishers.

Hage, J.C. 2000. "Dialectical Models in Artificial Intelligence and Law", *Artificial Intelligence and Law* 8: 137-172.

Loui, R.P. 1998. "Process and Policy: Resource-Bounded Non-Demonstrative Reasoning", *Computational Intelligence* 14: 1-38.

McBurney, P., and S. Parsons 2000. "Tenacious Tortoises: A formalism for argument over rules of inference", *Computational Dialectics (ECAI 2000 Workshop).* Berlin.

Pollock, J.L. 1987. "Defeasible reasoning", *Cognitive Science* 11: 481-518.

Prakken, H. 1997. *Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law.* Dordrecht: Kluwer Academic Publishers.

Prakken, H., and G. Vreeswijk To appear. "Logics for Defeasible Argumentation". In D. Gabbay (ed.), *Handbook of Philosophical Logic*. Dordrecht: Kluwer Academic Publishers.

Toulmin, S.E. 1958. *The Uses of Argument.* Cambridge: University Press.

Verheij, B. 1996. *Rules, Reasons, Arguments. Formal studies of argumentation and defeat*. Dissertation Universiteit Maastricht.

Verheij, B. 1999a. "Automated Argument Assistance for Lawyers", *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, 43-52. New York: ACM.

Verheij, B. 1999b. "Logic, context and valid inference. Or: Can there be a logic of law?" In H.J. van den Herik, M.-F. Moens, J. Bing, B. van Buggenhout, J. Zeleznikow, and C.A.F.M. Grütters (eds.), *Legal Knowledge Based Systems. JURIX 1999: The Twelfth Conference*, 109-121. Nijmegen: Gerard Noodt Instituut.

Verheij, B. 2000a. "DefLog - a logic of dialectical justification and defeat". See http://www.metajur.unimaas.nl/~bart/publications.htm.

Verheij, B. 2000b. "Dialectical Argumentation as a Heuristic for Courtroom Decision-Making". In P.J. van Koppen and N.H.M. Roos (eds.), *Rationality, Information and Progress in Law and Psychology. Liber Amicorum Hans F. Crombag*, 203-226. Maastricht: Metajuridica Publications.

Verheij, B. 2001. "Legal decision making as dialectical theory construction with argumentation schemes". See http://www.metajur.unimaas.nl/~bart/publications.htm.

Vreeswijk, G.A.W. 1997. "Abstract argumentation systems", *Artificial Intelligence* 90: 225-279.

Walton, D.N. 1996. *Argumentation Schemes for Presumptive Reasoning.* Mahwah: Lawrence Erlbaum Associates.

---

[1]   By the present conventions, the implicit assumption of a datum-claim argument is a 'logically minimal' commitment implicit in the argument, viz. simply the conditional that connects datum and claim. In argumentation theory, the generalization of such a conditional to a rule statement (like in our example the rule that those born in Bermuda are British subjects) is also often referred as the implicit assumption of an argument. Cf. argumentation theory (Van Eemeren *et al.* 1996, p. 14 and elsewhere). Those generalizations are here referred to as the warrants of an argument. See section 2.2 below.

[2]   Such a reading of logical derivations as informal arguments is the basis of many textbooks that connect logic with informal argument. Here logical derivations play a related, but more modest role, viz. as explications of the transfer of evaluation status from assumptions to issues.

[3]   This style of analysis has shown its fruitfulness in the abundance of work on modal logics. Cf. e.g. the entry on modal logic in the Stanford Encyclopedia of Philosophy (available online at http://plato.stanford.edu/entries/logic-modal/).

[4]   It may be noted that our treatment has the effect that warrants, conditional schemes and their instances all use the same qualifier (unlike in Toulmin's example where the qualifier in the warrant is 'generally' and the qualifier of the claim is 'presumably'). Cf. the remarks on qualifiers as modal operators in section 2.1.

[5]   As a result, the warrant of an argument 'B. So W' (where B is a backing of a warrant W) can again be challenged and require further backing. When the warrant W* of the argument 'B. So W' is backed by B*, the warrant of the corresponding argument 'B*. So W*' can on its turn be challenged, and so on, *ad infinitum*. As Toulmin remarks, *some* warrants must be left unchallenged (p. 106), just as some data for that matter. See e.g. Hage (2000) for a recent discussion of (amongst other topics) the resulting problem of justification (with an emphasis on the context of legal reasoning).

[6]   Especially in the law, the applicability of rules (that play a role similar to Toulmin's warrants) is often at issue. Cf. Hage 1997, Prakken 1997, Verheij 1996. In Hage's and Verheij's work, it is also discussed how the *weighing* of reasons for and against a conclusion can be formally treated.

[7]   It depends on ones taste whether this approach is regarded as a three-valued interpretation of the whole of a logical language (in terms of the values j, d and u for justified, defeated and unevaluated) or as a two-valued interpretation of a part of the language (in terms of the values j and d, with possibly some sentences of the language left uninterpreted).

[8] It can be said that the sentence 'It is defeated that C' expresses the *dialectical negation* of the statement that C. Cf. Verheij 2000a, b.