Legal knowledge based systems
JURIX 94
The Relation with Legal Theory

The Foundation for Legal Knowledge Systems
Editors:
H. Prakken
A.J. Muntjewerff
A. Soeteman

J. Hage, *Two Metaphors for Reasoning,* in: A. Soeteman (eds.), Legal knowledge based systems JURIX 94: The Foundation for Legal Knowledge Systems, Lelystad: Koninklijke Vermande, 1994, pp. 127-138, ISBN 90 5458 190 5.

More information about the JURIX foundation and its activities can be obtained by contacting the JURIX secretariat:

Mr. C.N.J. de Vey Mestdagh
University of Groningen, Faculty of Law
Oude Kijk in 't Jatstraat 26
P.O. Box 716
9700 AS Groningen
Tel: +31 50 3635790/5433
Fax: +31 50 3635603
Email: sesam@rechten.rug.nl

 http://jurix.bsk.utwente.nl/

Two Metaphors for Reasoning

Jaap Hage
University of Limburg, Dept. of Metajuridica
P.O. Box 616, 6200 MD Maastricht, The Netherlands
email: jaap.hage@metajur.rulimburg.nl

Abstract

This article discusses two metaphors that can be applied to reasoning. One metaphor is based on the notion of a container filled with information, the other on the idea of colliding forces. The metaphors are connected to different notions of logical validity, namely the semantic and the pragmatic notion. It is argued that the first metaphor is primarily suited to illuminate reasoning with factual statements, while the latter metaphor is more suitable to deal with reasoning with rules. The difference between the two metaphors is shown to be reflected in different approaches to defeasible reasoning.
Keywords: Reasoning with rules, Logical validity, Defeasibility

1 Introduction

Metaphors are very important for our view of the world [Lakoff & Johnson, 1980]. They strongly influence how we look at phenomena, and as a consequence how we try to solve problems in certain areas. Metaphors can for that reason be seen as central elements in scientific paradigms. Just as the quality of a paradigm strongly influences the success of a research program based on this paradigm, the quality of a metaphor is to a large degree determinative for the success of a direction of research. In fact, the metaphor underlying a paradigm may be seen as part of the hard core of a research program, where the various theories built on the metaphor are the protective belt of the program (Cf. [Lakatos, 1970, pp. 132f.]).

The *container metaphor*, especially its elaboration in terms of possible worlds, plays a central role in much of our thinking about valid reasoning and logic. A classical definition of a valid argument is that an argument is valid if its conclusion is true in all possible worlds in which all of its premises are true, that is, in all of the models of the premises. Fruitful as this metaphor may have been for much logical theorizing, it hinders an adequate analysis of reasoning with rules. At least, that is what I will argue in this article.

Based on a short characterization of the role of argumentation in everyday life, I propose another metaphor for arguments, the metaphor of colliding forces. In this metaphor, the notion of a reason for a conclusion plays a central role. I will show how the proposed metaphor makes sense of various forms of defeasibility. Moreover, I will give examples how the container metaphor has influenced the development of nonmonotonic logics, and how the colliding forces metaphor provides an alternative that may be more fruitful.

I have noticed that the ideas that are developed in this paper meet with considerable resistance. At the end of the paper, I try to answer some of the obvious criticisms of my approach. However, an argument for a different metaphor for reasoning can hardly be cogent. Therefore, the reader is asked to adopt the colliding forces metaphor as an experiment, and to look at classical logical problems from this new perspective. Just as with new glasses, it is only after a while possible to decide whether the new view is better than the old one. The single observation that the old glasses seem good enough will not do: carriages seemed good enough before there were cars.

2 The container metaphor for arguments

A common view of valid reasoning is that valid arguments make in their conclusions information explicit that was implicitly available in the premises. The theory that consists of the premises of an argument is considered as a *container filled with information.* Valid arguments show us parts of the contained information.

This container metaphor is often elaborated in terms of possible worlds. A possible world is a collection of states of affairs. A sentence divides the set of all possible worlds into the set of worlds in which this sentence is true, and the set of worlds in which this sentence is false. Analogously, every set of sentences (theory) divides the set of possible worlds into two sets, the models and the non-models of the theory.

Given this relation between a theory and the set of its models of this theory, it is possible to give an attractive definition of a valid argument. Each possible world can be considered as a container of information. Those possible worlds that are models of a particular theory, have a certain amount of information in common. In fact, the common information of all these possible worlds is precisely the information contained in the theory whose models they are. As a consequence, any sentence that contains information that is also contained in the theory, will be true in all models of this theory. That the conclusion of a valid argument can only contain information that was also contained in its premises becomes in possible world terminology: *An argument is valid if and only if its conclusion is true in all possible worlds in which all of its premises are true.*

Possible worlds play an important role in the model theoretic semantics that are given for many systems of logic. Not all of these uses of possible worlds are based on the container metaphor that is discussed here. A possible worlds semantics is only based on the container metaphor, if the truth of a sentence can be evaluated by means of one possible world. At the end of this paper, I will say a little about logics that make use of models that are not based on the container metaphor.

The notion of logical validity as expressed by means of possible worlds is essentially *semantic.* It makes use of the truth values of sentences in possible worlds to determine whether an argument is valid or not. In a calculus that is interpreted as a logic, the rules of inference define the *syntactic* notion of validity. An argument is valid if its conclusion can be produced from the premises by means of the rules of inference that belong to the calculus. Logical calculi are influenced by the possible worlds metaphor insofar as they make their rules of inference reflect the semantic notion of validity.

3 The container metaphor and rules

Rules have no truth value, at least not in the sense in which ordinary statements such as 'Amsterdam is the capital of the Netherlands' have a truth value. It is not simple to give a definition of a rule, if only because there are many different kinds of them. Rules can be distinguished on the basis of their conclusions, which can, amongst others, be classifications, and deontic and epistemic modalities. The latter rules can also function as (material) rules of inference. Moreover, rules exist in different ways. Some exist because they are valid within a social institution, others because they are employed by a sufficient number of members of a social group.

Yet, in spite of all these differences between kinds of rules, all rules have in common that they do not state the presence of facts in the world, but that they create a connection between these facts,[1] that is, all rules are *constitutive*. Still, for the following discussions, all rules may be considered as rules of inference, because constitutive rules justify the adoption of a similar rule of inference. For instance, the validity of the rule that thieves are punishable justifies the adoption of the rule of inference that if a person is a thief, it may be assumed that he is punishable.

That rules have no truth value has been mentioned as a reason why the traditional notion of logical validity cannot be applied to arguments in which rules are involved, because this classical notion makes essentially use of the notion of truth. This objection can be dealt with by extending the notion of truth, so that rules can be true or false.[2] If the notion of truth is extended, so that rules that are considered valid or acceptable, are treated as true for the purposes of a formal logic, it becomes possible to deal with normal rule application in for instance (modal) sentential and predicate logic. Rules can (but need not) be represented as material or strict implications, in which case their application is modeled as arguments of the form Modus Ponens.

This works as long as only normal rule application is involved, that is application of a rule whose conditions are satisfied. However, things go wrong if an argument based on a rule turns out to be defeasible.[3] A simple example can make this clear.

```
Thieves are punishable.
John is a thief.
Therefore: John is punishable.
```

The first premise of this argument expresses a valid rule; the second contains a true sentence; the form of the argument is valid; in short nothing seems to be wrong with this argument. Still the conclusion is false, because John's crime is prescribed because of lapse of time. This is a simple example of defeasible reasoning; the addition of new information makes a previously derivable conclusion underivable. Still, simple as the example may be, it poses serious problems for the Modus Ponens model of rule application, and also for the semantic notion of validity as used in connection with rule

[1] A more extensive discussion of rules, under the heading of nexus, can be found in [Hage, 1987].

[2] This approach has been taken by [Brouwer, 1982; Soeteman, 1989, pp. 62f], and Prakken (conversation with the author). Extending the notions of true and false may be accompanied by a change in terminology for the truth values.

[3] Things go also wrong in case of analogous rule application, but I do not want to discuss this here. Cf. [Verheij & Hage, 1994].

application, because this semantic notion, based on the container metaphor, validates Modus Ponens.

The cause of the problems is the interpretation of the first premise as the statement that *all* thieves are punishable. If the first premise is a statement, it must be false, because John, who is a thief, is not punishable. Still, the rule on which the premise is based is valid. It is possible to reformulate the first premise, so that it contains the additional condition that the theft is not prescribed, but this would be rather ad hoc. To make such a maneuver work, it is necessary to have a list with all possible exceptions, and exceptions to exceptions, etc. And even if it would be possible to think of such a list, it is not the case that the rules of law contain such complex conditions. This does not mean that legal rules can only be based on single statutory regulations. It is possible that more than one regulation is to be taken into account in formulating a rule of law. It is, however, not seriously possible to add all the exceptions based on the particular circumstances *of individual cases* into the conditions of a rule.

In short, it will not do to have rules in an argument if they are treated as if they were statements. This conclusion returns us to the issue of semantic validity. Does it make sense to work with the semantic notion of validity in case of arguments based on rule application? My answer to this question would be negative, because the whole issue of the information content, around which the semantic notion of validity is built, is meaningless in connection with rules. It is possible to have formal theories of reasoning with rules, but they cannot be based on the container metaphor and its corresponding notion of semantic validity. As an alternative for the semantic notion, I will therefore introduce the pragmatic notion of validity.

4 The pragmatic notion of validity

The pragmatic notion of validity is defined in terms of *effectiveness* of argumentation. Argumentation is the speech act in which a line of reasoning is brought to the fore. Only argumentation as an actual event can be effective; a line of reasoning can metaphorically be said to be effective if its use in argumentation tends to be effective. In a similar fashion, argumentation can be said to be valid if it employs a line of reasoning that is valid.

There is a connection between logical validity and the effectiveness of means to persuade an audience. Means of persuasion that are never effective, under whatever circumstances, are not valid means. On the contrary, valid means of persuasion will be normally effective. A deductively valid argument would never have counted as valid if it would most of the times be ineffective.

Notice that I almost automatically slipped into the discussion of argument types, rather than individual arguments. In the case of individual arguments, the relation between effectiveness and validity is not very interesting. We are interested in a possible general connection between the validity and the effectiveness of arguments, and such a general connection can only exist on the level of argument types, however a type may precisely be defined.[4]

[4]The notions of effectiveness and validity are connected to types themselves. Effectiveness is a causal concept, and as such connected to causal laws that operate on types of events [Quinton, 1974]. Validity means that there is conformance with criteria, which are also intrinsically connected to types [Hage,

We found that valid arguments will normally be effective. Is it possible to turn the connection between validity and effectiveness around, and say that argument types that are usually effective, are also valid? I think that the answer to this question should *in general* be affirmative, but that there are three conditions that must be fulfilled. The first condition is that the *connection* between effectiveness and validity is confined to argument *types*. The second condition is that arguments of an effective type should also be considered as *reasons* for the conclusion. The third condition, finally, is that an effective argument type should usually not loose its effectiveness if the rule of inference that defines this argument type is *criticized* on the basis of other standards.

4.1 Effectiveness and validity in connection with argument types

Not every argument that turns out to be effective, is ipso facto valid. People tend to make mistakes, also in their judgment of arguments, and this may lead to the effectiveness of invalid argumentation. However, people do not always make mistakes. If a type of argument is most of the times effective, this very fact means that this type of argument is considered to be a suitable type of argument for the conclusion at stake, or – in other words – that it is (informally) valid. Of course, I assume here that the other two conditions for validity are also satisfied.

If validity is a characteristic of argument types, it is important to know how arguments can be divided into the relevant types. Here the notion of a *rule of inference* comes in. Rules of inference must abstract from some of the details of the arguments they govern. If only for practical reasons, there cannot be a rule of inference for every individual argument; a rule must state the validity of classes of inferences, which are identified by a number of characteristics identified in the rule. A rule of inference specifies which characteristics of an argument determine whether the argument is valid. An argument type is valid if arguments of this type are licensed by a rule of inference.

If an audience uses a rule of inference, it will both recognize arguments based on this rule as valid, and be normally convinced by these arguments. It is not possible to use a rule of inference, and still be normally unconvinced by arguments based on these rules. A rule with such dubious effects can hardly be called a rule of *inference* anymore. That is why valid arguments will usually be effective arguments too.

4.2 Causes and reasons

Not every argument type that is usually effective, is also valid. Suppose that it would be an effective strategy for an attorney, who wants a suspect to be convicted, to point out that the suspect is black. Does this make the argument that somebody is black a valid one for the thesis that this person should be convicted?[5]

Pointing out that the suspect is black may trigger unconscious mechanisms that would not be sanctioned by the persons subject to them, if they would be aware of

1984].

[5]The word 'argument' is ambiguous. It denotes both a line or reasoning and a particular statement made in a line of reasoning. I use the word in both senses and trust that, where the difference is important, the context makes it clear which sense is used. Logicians sometimes use the word 'argument' in still another sense, namely in the sense of the set of sentences that is employed (as the premises) in a line of reasoning.

them. Effectiveness of arguments only implies validity if the causal connection between the arguments and their effects is approved by the persons for whom this connection exists. Such approval can take place in the form of adducing the arguments as the *reason* (not only the cause) why the conclusion was adopted. In other words, the facts adduced in the argument should not only *explain*, but also *justify* the conviction in the eyes of the audience.

Of course, explicit approval is only possible in those cases in which the convinced person acknowledges the arguments as the cause of his conviction. This need not be the case in all cases of conviction. However, for an argument type to be valid it is necessary that the persons for whom the type is effective approve of this effectiveness *if they are aware of it.*

4.3 Rules of inference should withstand criticism

Until now, I discussed rules of inference as if they were rules adopted by individual persons. In the end, this seems to be correct. Rule following behavior ultimately comes down to behavior of individuals. Still, the validity of arguments is usually evaluated by means of standards that are in practice in a social group. Rules of inference are normally social rules.

What counts as valid in one society or period of time, is based on the rules of inference used in that society at that time; it may be completely invalid in another society or period. The standards of validity are relative to a particular audience. Still, this does not mean that it is impossible to argue about standards. Even in a racist society it is possible to argue against discrimination on the basis of race. Such arguments will only be effective (and valid) if this society adopts additional standards that make it possible to evaluate the first-mentioned standards. The result of this evaluation may be that the racist standards are wrong and consequently that it is invalid to argue for a person's conviction by pointing out that he is black.[6] [7]

4.4 The fact-value distinction

It may seem that the pragmatic notion of logical validity is based on a confusion between the evaluative notion of validity and the factual notion of effectiveness. This appearance is deceptive. The pragmatic notion of validity is based on rules of inference (standards for the validity of arguments) and is as such evaluative. An argument is valid if it is in accordance with a valid rule of inference.

Effectiveness of arguments plays a role in the analysis of the existence (validity) of rules of inference. Whether a rule of inference exists (whether it is valid) is a purely factual matter, and whether an argument is in accordance with such a rule is also a

[6] Notice that the rule of inference involved here, namely that if somebody is black, he should be convicted, is a material one, comparable to a warrant I the sense of [Toulmin, 1958]. In general, the definition of pragmatic validity leaves room for both formal and material rules of inference.

[7] There is a problem here with arguments based on standards which should be rejected if they were criticized, but which are actually not criticized. Such arguments would be considered as valid ones by the audience, but they would not be valid according to the definition of pragmatic validity given here. I will not discuss this case, but refer the reader to [Hage *et al.*, 1994], where a similar problem is discussed in connection with hard cases.

purely factual matter. Therefore, the validity of an argument is a purely factual matter, and the crucial facts have, in the pragmatic view, to do with the effectiveness of types of argumentation.

Still, it may be objected that I confuse the internal and the external point of view here. The reply to this objection is that this distinction is irrelevant for logical purposes. I may use a standard in making an evaluative statement. In that case I take the internal point of view, and do not relativize the evaluation to the validity of the standard. I seem to accept the standard. However, the truth value of the evaluative statement still depends on the validity of the standard. It does not matter whether the internal or the external point of view is taken, a standard must be valid in order to support evaluative judgments.

Since a logical theory is involved with the truth (or the acceptability) of conclusions, the internal/external distinction does not play a role in it. An external theory that specifies the truth conditions for evaluative judgments, for instance about the validity of arguments, also specifies under which circumstances the evaluations are justified.

It will be clear that the pragmatic notion of validity that is defined here is completely different from the semantic notion. It is much more connected to social practices and human behavior, and less to the meaning and the truth conditions of sentences. Although there seems to be room for both the semantic and the pragmatic notion of logical validity, I think that the pragmatic notion is more useful if we deal with rules. In the next section, I develop the forces metaphor of reasoning, to make this claim more clear.

5 The forces metaphor of reasoning

A rational audience that is subjected to arguments that try to convince it of a particular conclusion, or of its denial, can be compared to a physical body that is subject to a number of colliding forces. Just as the combination of the forces determines in which direction the body will actually be moved (accelerated), the combination of the adduced reasons determines whether a rational audience will accept the conclusion or its negation, or refrain from judgment.

The forces that act upon a body are determined by the facts of the situation at hand, and by the applicable laws of nature. In fact, the forces can be considered as instantiations of the physical laws. None of the forces by itself determines the behavior of the physical body; only in their interaction they cause the body to move. Clearly, if there is only one force, that force by itself determines what will happen. In this case, the one force should be considered as the combination of all the forces.

The relation between a physical law and a force is not straightforward. Indeed, in simple cases, the magnitude of the force can simply be computed by filling in the parameters in the formula that represents the law. The gravitation on Earth can, for instance, be computed by filling in the mass m of the accelerated body in the formula $G = 9.8m$. The scope of this law is restricted, however. The law only applies near the surface of planet Earth. In a similar fashion, the Newtonian laws of gravitation only apply if the concerned velocities are not extremely high. In short, a law does not apply to a particular case if this case falls outside the scope of the law.[8]

[8] The scope of physical laws is discussed in [Toulmin, 1953, pp. 31 and 78].

The relation between a physical law and its instantiations, the forces that occur in a particular case, is similar to the relation between a rule[9] and the reasons that are based on it. If a rule applies to a particular case, the facts that make the rule applicable (that match with the rule conditions) become a reason for the rule conclusion.

A reason is comparable to a force; it pulls a rational audience towards acceptation of the conclusion for which it is a reason. A rule is comparable to a physical law. By itself a rule has no effects, but if it is applied to the facts of a particular case, it makes some of these facts into reasons. Similarly the law of gravitation by itself does not make things move, but applied to a concrete situation it makes one body attract another body.

Just as physical laws have a scope, rules have a scope too. Criminal laws do not apply to prescribed cases, some laws of contract do not apply in the case of force majeure, and inferior rules do not apply if they conflict with applicable superior rules.

The purpose of the comparison between a body that is subjected to forces, and an audience that is subjected to arguments, is to suggest that the conclusion from a set of premises is the outcome of the interactions of the reasons that can be based on these premises, and that the role of rules, which make facts into reasons, is comparable to that of forces, which make facts into causes.The comparison may be deceptive, because not all 'forces' that act on an audience represent reasons for accepting a particular conclusion. It is not my intention to suggest that a conclusion can validly be drawn from a set of premises, if the effect of adducing these premises is that the audience comes to accept the conclusion. That is why I wrote about the influences of reasons on a *rational* audience.

5.1 The forces metaphor and defeasibility of rule application

If reasoning is considered in the light of the forces metaphor, the defeasibility of rule application becomes a natural phenomenon. Let us define defeasibility in the case of rule application as follows:

The application of a rule is defeasible if it is possible that the conclusion of the rule is not drawn in a particular case, even though the conditions of the rule are satisfied in this case.

On the basis of the colliding forces metaphor, defeasibility can be explained in two ways. First it is possible that the case falls outside the scope of the rule. In that case the rule does not apply, even though its conditions are satisfied. If the rule does not apply, it does not generate a reason, and then the conclusion should normally not be drawn either. The second possibility is that the rule is applied and generates a reason for its conclusion, but that this reason is neutralized or outweighed by other reasons pleading in the opposite direction.

6 Forms of nonmonotonic reasoning

Let us say that two rules collide in a particular case if the instantiated conclusions of these rules are inconsistent. If rules are considered as statements that are true or false in possible worlds, theories that contain colliding rules that are both applicable are inconsistent. There are in principle two possibilities to withhold a theory with colliding

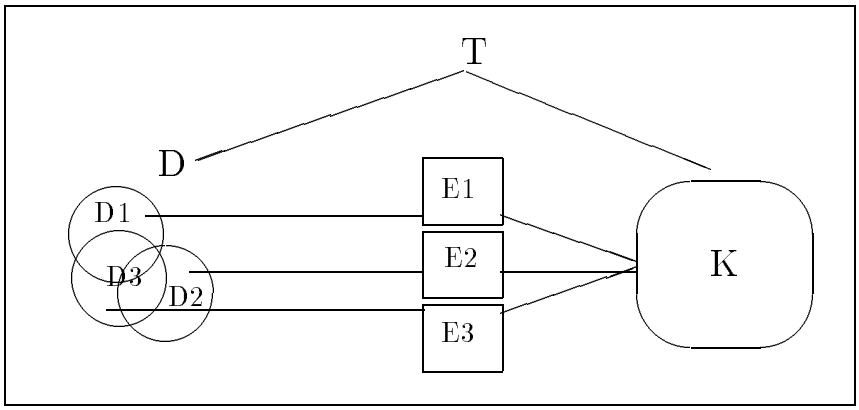
[9] For the present purposes I identify rules and principles. The reader may substitute the word 'principle' for 'rule' if she wishes to.

rules from being inconsistent.

One possibility is to leave a number of rules out of consideration, so that the result does not contain colliding rules anymore. This option is chosen by a number of nonmonotonic logics that deal with colliding rules by means of a kind of inconsistency handling. This approach is, in some form or another, taken in e.g. [Reiter, 1980; McCarthy,1980; Delgrande, 1988; Poole, 1988; Brewka, 1990; Geffner & Pearl, 1992]. I will globally characterize the way in which these logics deal with rules, necessarily skipping many details that make these logics different from each other.[10]

A theory T is considered as consisting of two parts (each of which can theoretically be empty), that is a consistent part of 'hard' knowledge and a part of defeasible knowledge (rules) in the form of so–called defaults. Let us call the hard knowledge K and the defeasible knowledge D. $T =< K, D >$.

If T as a whole is consistent, reasoning with T is dealt with by classical logic or in a similar way. If T is inconsistent, a new consistent theory T' is generated by adding a subset D' of D to K, so that $T' =< K, D' >$. This new consistent theory T' is then dealt with by classical logic or similarly.



Often it will be possible to combine many different subsets of D (say $D1..D3$) with K so that the result is consistent. For each of these subsets there is a corresponding set of sentences that can be derived from them. These different sets of derivable sentences are called the extensions of T. A sentence can be derived from a theory T if this sentence occurs in one or all of the extensions, or in one or all of the *preferred subsets* of the extensions of T.

The other option is to 'weaken' the conclusions that follows from rules, so that the results of rule application are not incompatible anymore. Such a weakening may consist of making the consequence of the application of a rule the existence of a *reason* for the rule conclusion. The reasons that result from the application of colliding rules still have to be 'weighed' to determine which conclusion follows. This option corresponds to the colliding forces metaphor or reasoning, and is adopted in Reason–Based Logic [Hage,

[10]The strive to capture a number of theories in one abstract description made it necessary to force these theories in the same terminological framework, which means that the terms employed here are not necessarily those that were used in the original descriptions.

1993; Hage & Verheij, 1994a; Hage & Verheij, 1994b].

An alternative form of weakening is to assign the conclusion of the rule a probability which is relativized to the rule conditions. If the probability of C, given A ($P(C|A)$) is 0.5, and the probability of C, given B ($P(C|B)$) is 0.01, this is not inconsistent, even if it is given that A and B. This form of weakening would, however, ask for a completely different interpretation of rules. Such an interpretation would certainly not do for most rules of law.

An advantage of the colliding forces metaphor is that it immediately suggests that the outcome of argumentation should be the result of weighing *sets of* reasons, rather than of individual arguments. Indeed, the outcome of colliding forces is also determined by the summation of all the present forces.

We find that the two metaphors for reasoning, and their corresponding notions of validity lead to different ways to consider nonmonotonic reasoning and consequently to different nonmonotonic logics to deal with rules.

Still another metaphor considers defeasible reasoning as a conflict between a number of competing arguments. The strongest argument wins, and its conclusion is also the conclusion from the set of all the premises from which the competing arguments were constructed [Pollock, 1992; Simari & Loui, 1993; Prakken, 1993; Vreeswijk, 1993]. This metaphor takes a place in between the forces metaphor and the container metaphor. With the forces metaphor, it shares its focus on argumentation. With the container metaphor it has in common that if an argument wins, its conclusion follows. This last idea is related to the idea that if a conclusion is entailed by a set of true premises, it must be true.

7 Conclusion and Reply to criticisms

The difference between rules and statements is reflected in two complementary metaphors for reasoning. Reasoning with statements that describe factual states of affairs is adequately dealt with by the container metaphor. The corresponding notion of validity is a semantic one: the conclusion of a valid argument should not contain information that is not available in the premises.

Reasoning with rules is covered by the metaphor of colliding forces. Rules generate reasons, which can conflict just like forces that pull a physical object in different directions. Only in the case of reasons it is not a physical object, but rather (the justified belief of the audience concerning) the correct conclusion that is being pulled. This metaphor is accompanied by the pragmatic notion of validity. An argument is valid if the type of reasons adduced in it is usually effective, if the reasons are recognized as such by the audience, and if the involved rules of inference stand up against criticisms.

The two metaphors lead to different approaches to reasoning with rules. In particular, the colliding forces metaphor does not consider reasoning with colliding rules as a kind of inconsistency handling. Inconsistency handling is not the right way to deal with colliding rules, because colliding rules are not inconsistent.

Some may object that my claim that the container metaphor cannot adequately deal with rule application has not been sufficiently substantiated yet. My main argument against logics based on this metaphor was that they cannot deal with defeasibility of rule application. However, there are logics, described in terms of possible worlds, that

can handle defeasibility. These logics work with weaker conditionals than the material implication, and do not validate Modus Ponens for these conditionals (e.g. [Morreau, 1994]).[11]Are not these logics counter examples to my claim?

The answer is negative, for two reasons. The first reason, which is the least important one, is that a weaker conditional may be suitable to express conditional statements, but not to express rules. The rule that thieves ought to be punished remains valid, even if we know that John, who is a thief, ought not to be punished because he is a minor. However, under these circumstances the conditional statement that if John is a thief, he ought to be punished, is false. Conditional statements are only true if there is no information to the effect that there are exceptional circumstances. Such information does not affect the validity of rules, which does not depend on circumstances that influence the applicability of a rule.

The second reason, which is more fundamental, is that conditional logics, such as the logic proposed by Morreau, have implicitly abandoned the container metaphor, although they still work with the notion of possible worlds. Their conditional sentences do not (only) give us information about the world in which they are true, but indicate the connection between facts of particular types, a connection which also depends on other facts that are not specified in the conditional itself.[12]In themselves, these conditional do not tell us anything about the facts that obtain in the world in which they are true, just as a single force does not tell us anything about the effects that are caused.

Because conditional logics are not based on the container metaphor, they are not a counter–example against my claim that logics based on the container metaphor are inadequate to deal with rule application. The same counts for logics, such as the logic of conditional entailment [Geffner & Pearl, 1992], which for the definition of their derivability relation make use of a preferred subset of the models of the premises. Actually these logics more or less do the same thing as conditional logics, which, in establishing the truth of the conditional, take only a preferred subset of the models of the antecedent of a conditional into account.

References

Brouwer, P.W. (1982). Over de toepassing van de tweewaardige logica in het rechtsdenken. Soeteman, A. en P.W. Brouwer eds. *Logica en recht, W.E.J. Tjeenk Willink, Zwolle.*

Delgrande, J. (1988). An approach to default reasoning based on a first-order conditional logic: revised report. *Artificial Intelligence 36, p; 63-90.*

Geffner, H. and J. Pearl (1992). Conditional entailment: bridging two approaches to default reasoning, *Artificial Intelligence 53, p.209-244.*

Hage, J.C. (1984). Enkele opmerkingen over geldigheid en gelding, *Algemeen Nederlands Tijdschrift voor Wijsbegeerte, deel 76, blz. 262-266.*

Hage, J.C. (1987). *Feiten en betekenis; een verhandeling over ontologie en praktische rede, PhD.-thesis, Leiden.*

Hage, J.C. (1993). Monological Reason Based Logic. *Proceedings of the fourth International Conference on Law and Artificial Intelligence, ACM-press, Amsterdam, p. 30-39.*

[11]Comparable logics are described in [Stalnaker, 1968; Lewis, 1973; Delgrande, 1988].

[12]A conditional statement implicitly tells us that in the world in which it is true, some rule is valid that connects its antecedent and its consequent, and that in this world there are no facts that prohibit the drawing of its conclusion. This is, however, not the meaning of the conditional.

Hage, J.C. R. Leenes, and A. Lodder (1994). Hard cases; a procedural approach. *Artificial Intelligence and Law, vol. 2, pp. 113-167.*

Hage, J.C. and H.B. Verheij (1994a). Reason-Based Logic: a logic for reasoning with rules and reasons. Paper accepted for *Law, Computers and Artificial Intelligence. Also to be obtained from the authors.*

Hage, J.C. and H.B. Verheij (1994b). Towards a logic for reasoning with norms. J. Breuker (ed.) *Proceedings of the ECAI-94 Workshop on Artificial Normative Reasoning, pp. 160-177.*

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programs. I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge, Cambridge University Press, pp. 91-196.*

Lakoff, G. and Johnson, M. (1980). *Metaphors we live by, The University of Chicago Press, Chicago, London.*

Lewis, D (1973). *Counterfactuals. Basil Blackwell, Oxford.*

McCarthy, J. (1980). Circumscription-a form of non-monotonic reasoning. *Artificial Intelligence 13, p. 89-116.*

Morreau, M. (1994). Prima facie and Seeming Duties, A.J.I. Jones and M.Sergot (eds.), *Deon '94, Proceedings of the second international workshop on deontic logic in computer science, pp. 221-251.*

Pollock, J.L. (1992). How to reason defeasibly, *Artificial Intelligence 57, pp. 1-42.*

Poole, D. (1988). A logical framework for default reasoning. *Artificial Intelligence 36, p. 27-47.*

Prakken, H. (1993). *Logical tools for modeling legal argument, PhD.-thesis Amsterdam.*

Quinton, A. (1974). *Utilitarian Ethics. The MacMillan Press, London.*

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence 13, p. 81-132.*

Simari, G.R. and R.P. Loui (1992). A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence 53, p.125-157.*

Soeteman, A. (1989). *Logic in Law. Remarks on Logic and Rationality in Normative Reasoning, Especially in Law. Kluwer Academic Publishers, Dordrecht.*

Stalnaker, R.C. (1968). A Theory of Conditionals, *Studies in Logical Theory, (N. Rescher ed.), Blackwell, Oxford.*

Toulmin, S.E. (1953). *The philosophy of science; an introduction, Hutchinson & Co., London.*

Toulmin, S.E. (1958). *The Uses of Argument, Cambridge University Press, London.*

Verheij, H.B. and J.C. Hage (1994). *Reasoning by analogy; a formal reconstruction, in this volume.*

Vreeswijk, G.A.W. (1993). *Studies in Defeasible Argumentation. PhD-thesis, Amsterdam.*

Acknowledgement

The research for this paper was partly financed by the Foundation for Knowledge Based Systems (SKBS), which seeks to improve the level of expertise in the Netherlands in the field of knowledge based systems, and to promote the transfer of knowledge in this field between universities and business companies. The author wishes to thank Bart Verheij, Henry Prakken and the referees of Jurix-94 for their comments on drafts of this article.