Rule Consistency

Jaap Hage

Department of Metajuridica, Maastricht University, The Netherlands
jaap.hage@metajur.unimaas.rl

Abstract

This paper develops the theory that a set of rules is consistent if it is not possible that: 1) the conditions of the rules in the set are all satisfied, 2) there is no exception to either one of the rules, and 3) the consequences of the rules are incompatible. To this purpose the notion of consistency is generalised to make it cover rules and is relativised to a background of constraints.

The theory is formalised by means of Rule Logic, a logic in which rules are treated as constraints on the possible worlds in which they exist. Rule Logic itself is introduced by giving a model-theory for it. It is characterised by means of constraints on worlds that are possible according to Rule Logic.

1 Introduction and overview

The knowledge base of a legal knowledge system often consists of a set of rules. The developer of such a knowledge base would like it to be consistent, in the sense that it does not lead to incompatible outcomes for a specific case. A legislator would, for similar reasons, like the legislation of a particular legal system to be consistent. The notion of consistency of rules is of practical interest, both for the development of legal knowledge systems and for the development of legal systems.

The purpose of this paper is to introduce and to develop a theory about the consistency of rules. There are at least three reasons why the consistency of rules differs from the consistency of descriptive sentences, all of which will be discussed in the following sections. First, many rules have a conditional structure, but their consistency cannot be treated as the consistency of conditional sentences. Second, the consistency of rules is relative to a set of constraints which determine which states of affairs can go together. Part of the complications in connection with rule consistency is that rules themselves can function as constraints relative to which consistency has to be judged. And finally, reasoning with rules is defeasible. There can be exceptions to rules that block the application of rules the conditions of which are satisfied. These exceptions can prevent threatening rule conflicts, thereby making seemingly inconsistent rules consistent. I will try to develop a theory of rule consistency that takes all these three aspects into account.

Jaap Hage, 'Rule Consistency', in: H.Jaap van den Herik et al. (eds), *Legal Knowledge Based Systems, JURIX 1999, The Twelfth Conference*, Nijmegen: GNI, 1999, p. 47-59.

2 Rules as conditionals

If the consistency of rules were the same as that of conditional sentences, the following two rules would be consistent:

Thieves are punishable.
Thieves are not punishable.

Their consistency would follow from that the following sentences are *not* inconsistent:

$\forall x(\text{Thief}(x) \rightarrow \text{Punishable}(x))$
$\forall x(\text{Thief}(x) \rightarrow \sim\text{Punishable}(x))$

Instead of being inconsistent, these sentences allow the derivation of

$\sim\exists x(\text{Thief}(x))$

Intuitively, however, a legislator should not be able to remove thieves from the world, merely by issuing both the rules that thieves are punishable and that they are not punishable.

The first conclusion to draw is, therefore, that a theory in which rules are treated like descriptive sentences, and that considers rule consistency as similar to the consistency of descriptive sentences, is unsatisfactory. There is reason to search for a notion of consistency that is especially suited to rules. A relevant intuition in this connection is that the consistency of rules should not depend on whether certain states of affairs obtain. We want, for instance, the rules that thieves are punishable and that they are not punishable to be inconsistent independent of whether there are thieves. If there are thieves, the two rules can, barring exceptions, be used to derive an inconsistency in the traditional sense, because then it can be derived that these thieves are both punishable and not. However, we want the inconsistency of the rules to be independent of whether there are facts that satisfy their conditions.

Yet, it is important for the consistency of rules whether the rules can be applied to the same case. For instance, the rules that thieves are punishable and that non-thieves are not punishable are intuitively consistent. The inconsistency of rules depends both on the incompatibility of the conclusions of the rules, and on the compatibility of the rule conditions.

It should, however, not be possible to make an inconsistent set of rules consistent by adding a rule with conditions that are incompatible with the conditions of the other rules in the set. For instance, the rules that thieves are punishable and that minors are not punishable should count as inconsistent, because there is a case (a minor thief) to which the rules attach incompatible consequences. This set of rules should not be made consistent by adding the rule that non-thieves are not punishable. To avoid this complication, the demand is made that a set of consistent rules does not contain an inconsistent subset. Or, in other words, a set of rules is inconsistent if it contains an inconsistent subset.

This leads me to the following provisional theory about rule consistency:

The rules in a set s are consistent if and only if it is not the case that there are a subset s' of s and a case[1] f such that
a. the facts in f satisfy the conditions of all the rules in s', and
b. to which the rules in s' attach incompatible consequences.

3 Consistency, compatibility and constraints

Descriptive sentences are called *consistent* if it is possible that they are all true. For instance, the sentences 'John is a thief' and 'John is a minor' are consistent, because it is possible that John is both a thief and a minor. In other words, because the *states of affairs* that John is a thief and that he is a minor are *compatible*, the *sentences* that express these states of affairs are *consistent*. The sentences 'John is a thief' and 'John is not a thief' are inconsistent, because it is not possible that John both is and is not a thief. It is the incompatibility of the states of affairs that John is a thief and that he is not a thief that makes the corresponding sentences inconsistent.

Compatibility of states of affairs is always relative to some background of *constraints*. The states of affairs that John is a thief and that he is not a thief are incompatible because of the constraint that a state of affairs cannot both obtain and not obtain. A similar constraint is that the compound state of affairs that John is both a thief *and* a minor can only obtain if *both* the states of affairs that John is a thief and that he is a minor obtain.

Such constraints are usually called *logical* constraints. Besides logical constraints, there are also other constraints. There are *physical* constraints that prevent somebody from being in two non-adjacent countries at the same time. *Conceptual* constraints make it impossible that something is both a square and a circle.

This is the occasion to introduce a terminological convention. The expressions 'compatible' and 'incompatible' will be used for *states of affairs* which can, respectively cannot go together relative to a set of constraints. The expressions 'consistent' and 'inconsistent' are used for both descriptive sentences and for rules, with different criteria for consistency. Sentences are consistent if and only if the states of affairs they express are compatible. The criteria for rule consistency are the subject of this paper.

What is possible depends on the constraints that are taken into account. I will develop this idea by means of the notion of a possible world. A state of affairs is possible (can obtain), if there is at least one possible world in which this state of affairs obtains. A world is possible relative to some set of constraints c, if the facts of that world satisfy the constraints in c.

The most fundamental constraints on a possible world are the logical constraints. These include that a state of affairs cannot both obtain and not obtain in a world. A world that is physically possible must satisfy the constraints of physics. These are the physical laws and everything that follows from them. For instance, in a physically possible world, it does not occur that there are two bodies with gravitational mass that do no exercise mutual gravitational forces. Neither does it occur in a physically possible world that something travels faster than the speed of light in vacuum.

1 A case in this connection is a set of logically compatible states of affairs which need not obtain actually.

In a legally possible world it does not occur that somebody is a thief and not punishable.[2] As this last example shows, the constraints on possible worlds can be the result of human culture. By adopting rules, humans can impose additional constraints on the world in which they live. In contrast to physical constraints, rule-based constraints are contingent in the sense that they are absent in a world in which these rules do not exist. But when they exist, they rule out certain combinations of states of affairs as impossible, and necessitate other states of affairs.
If the compatibility of states of affairs is relative to a set of constraints, this has implications for our provisional definition of rule consistency:

The rules in a set s are consistent relative to a set of constraints c if and only if it is not so that there are a subset s' of s and a case f such that
a. *the facts in f are compatible relative to c,*
b. *the facts in f satisfy the conditions of all the rules in s', and*
c. *the rules in s' attach consequences to f that are incompatible relative to c.*

I will treat logical compatibility of states of affairs as a special case, that is as compatibility relative to the empty set of constraints. In the second part of this paper, where the present theory is formalised, this idea will be made more precise. Let me illustrate the implications of the above theory of rule consistency by means of some examples.[3]

Example 1
rule_1 = *thief(*x*) ⇒ *punishable(*x*)
rule_2 = *minor(*x*) ⇒ *~punishable(*x*)

The rules 1 and 2 are logically inconsistent, because if John is both a thief and a minor, these facts satisfy the conditions of both rules, while the conclusions of the two are logically incompatible.[4] Notice that the inconsistency of the rules does not depend on the presence of the facts that John is a thief and a minor. These facts merely *illustrate* the inconsistency.

Example 2A
rule_1 = *thief(*x*) ⇒ *punishable(*x*)
rule_2 = *minor(*x*) ⇒ *protected(*x*)
rule_3 = *protected(*x*) ⇒ *~punishable(*x*)

The rules 1, 2 and 3 are logically inconsistent, because the rules 1 and 3 are logically inconsistent.

4 Rules as constraints

Rules function as constraints on the worlds, or on the normative systems, in which they exist. In the Netherlands the rule exists that thieves are

2 Exceptions to rules are disregarded at this stage of the presentation.
3 In these examples I use the formalism that will be explained in the second part of this paper. Because of its similarity to the language of predicate logic, I assume that this will cause no problems.
4 The constraints that are relevant for logical consistency are the constraints of some system of logic. Presently I assume the constraints of predicate logic.

punishable. As a consequence the states of affairs that somebody is a thief and that he is not punishable are, barring exceptions to the rule, incompatible according to Dutch law. In a legal system where this rule does not exist, these states of affairs might be compatible.

The phenomenon that rules can function as constraints on possible worlds has implications for the above theory of rule consistency. To illustrate this, I will adapt example 2A:

EXAMPLE 2B
rule_1 = *thief(*x*) ⇒ *punishable(*x*)
rule_2 = *minor(*x*) ⇒ *protected(*x*)
c = {rule_3: *protected(*x*) ⇒ *~punishable(*x*)}

The third rule of example 2A is removed from the set of rules that is evaluated as to its consistency, and added to the set c of constraints that govern the world in which the rules 1 and 2 are evaluated. The first thing to notice is that the remaining rules 1 and 2 are logically consistent. This is not surprising, because the inconsistency of the rules 1 to 3 in example 2A depended on the presence of rule 3. If this rule is removed from the set, the logical consistency is restored.

However, if the removed rule is added to the constraints relative to which consistency is evaluated, the rules 1 and 2 become inconsistent, since this constraint makes the states of affairs that somebody is punishable and that he is protected incompatible. So the rules 1 and 2 are logically consistent, but they are inconsistent relative to the set c of constraints which includes rule 3.

The rules 1 to 3 are logically consistent, but relative to rule 3 they are inconsistent. The following examples illustrate that it can even make a difference for the set of all rules, whether a rule is part of a set that is evaluated as to its logical consistency, or that this rule is taken as part of the constraints:

EXAMPLE 3A
rule_1 = *thief(*x*) ⇒ *punishable(*x*)
rule_2 = *minor(*x*) ⇒ *~punishable(*x*)
rule_3 = *minor(*x*) ⇒ *~thief(x)

EXAMPLE 3B
rule_1 = *thief(*x*) ⇒ *punishable(*x*)
rule_2 = *minor(*x*) ⇒ *~punishable(*x*)
c = {rule_3: *minor(*x*) ⇒ *~thief(x)}

We have seen in example 1 that the rules 1 and 2 by themselves are *logically* inconsistent. This logical inconsistency is maintained if rule 3 is added to the rules 1 and 2, because an inconsistent set of rules cannot be made consistent by adding to it.

However, the situation changes if rule 3 is taken *as the background* against which the consistency of the rules 1 and 2 is evaluated, as in example 3B. The conditions of the rules 1 and 2 are not compatible relative to a background consisting of rule 3. As a consequence the rules 1 and 2 are consistent *relative to rule 3*, even thought the rules 1 to 3 are logically inconsistent. Apparently it makes a difference whether a rule is considered as part of the set that is evaluated as to its logical consistency, or as part of the background for the consistency of the other rules.

In the examples 2A and 2B it did not matter for the consistency of the set whether rule 3 was part of the set, or part of the background, while in the examples 3A and 3B this was very relevant. This difference can be explained by pointing out that in the examples 3A and 3B, rule 3 made the *conditions* of the rules 1 and 2 incompatible, while in the examples 2A and 2B, rule 3 made the *conclusions* of the rules 1 and 2 incompatible.

5 Conditionless rules

Until now we have only considered conditional rules. There are, however, also rules without conditions, such as the rule that it is forbidden to steal. Such rules share some characteristics with rules that have conditions, in particular that they can have exceptions. For the evaluation of their consistency they seem a little different, however. The first part of the definition of rule consistency, that there is a set of compatible states of affairs that satisfies the conditions of all the rules, seems not to apply to conditionless rules.

This seeming complication is easily remedied, however, by treating conditionless rules as rules with conditions, where the conditions are always satisfied. If conditions which are always satisfied are denoted by the term *true, conditionless rules are represented as rules with *true as their condition part. For instance:

*true ⇒ *forbidden(stealing)

The rules that it is forbidden to steal and that it is permitted to steal are then easily shown to be inconsistent against the background of the constraint that an action cannot both be forbidden and permitted:[5]

rule_1	=	*true	⇒	*forbidden(stealing)
rule_2	=	*true	⇒	*permitted(stealing)
c = {*forbidden(*action*)			⇒	*~permitted(*action*)}

The inconsistency of the rules 1 and 2 against the background c is illustrated by any case, since any case satisfies the conditions of these two rules.

6 Exceptions to rules

It is not uncommon that two rules in a legal system attach incompatible consequences to a case. For instance, the rule that an owner is allowed to do anything he likes with his property collides with many rules that limit his property right. In such cases, the law contains a *prima facie* rule conflict. Many prima facie rule conflicts turn out not to be *actual* conflicts, because one of the prima facie conflicting rules is left out of application by making an exception to it. Two or more rules are in actual conflict when they *actually* apply to one and the same case, and attach incompatible consequences to this case.

I will use an example again to sharpen our intuitions concerning the effect of exceptions to the consistency of rules. Take the following three rules:

1: Thieves are punishable.
2: Minors are not punishable.
3: In case of minors there is an exception to the rule that thieves are punishable.

5 Some would want to include this constraint into the set of logical constraints. In general the example leaves a lot to be said concerning deontic logic. That is beyond the scope of this paper, however.

These rules interact in case of a minor who is a thief. If rule 3 is left out of consideration, the rules 1 and 2 are intuitively inconsistent, because they lead to incompatible results in case of a minor thief. Rule 3 prevents that rule 1 is applied, however, so that the prima facie rule conflict is not actualised.[6] Therefore the rules 1 and 2 are in my view consistent against the background of rule 3, although they are logically inconsistent.

The insight that rules can have exceptions which prevents them to come into an actual conflict leads to the following adapted version of the above theory of rule consistency:[7]

Let a minimal conflict set of rules *relative to a set of constraints c be a set m of rules such that:*

1. there is a case f such that

- *1..1. the facts in f are compatible relative to c,*
- *1..2. the facts in f satisfy the conditions of all the rules in m,*
- *1..3. the rules in m attach consequences to f that are incompatible relative to c.*

2. there is no set m' which is a proper subset of m, such that there is a case f' such that

- *2..1. the facts in f are compatible relative to c,*
- *2..2. the facts in f satisfy the conditions of all the rules in m',*
- *2..3. the rules in m' attach consequences to f that are incompatible relative to c.*

The rules in a set s are consistent relative to a set of constraints c if and only if for every minimal conflict set of rules relative to c s' which is a subset of s there is no case f such that

a. the facts in f are compatible relative to c,

b. the facts in f satisfy the conditions of all the rules in s',

c. there is no exception to either one of the rules in s'.[8]

The implications of the amendment to the theory of rule consistency which takes exceptions into account are illustrated by the following examples:

EXAMPLE 4
rule_1 = *thief(x) ⇒ *punishable(x)
rule_2 = *minor(x) ⇒ *~punishable(x)
c = {rule_3 = *minor(x) ⇒ *exception(*thief(x) ⇒ *punishable(x))}

Rule 3 holds that if somebody is a minor, there is an exception for him or her to the rule that thieves are punishable.

The rules 1 and 2 by themselves are logically inconsistent and a minimal conflict set relative to the empty set of constraints. Inclusion of rule 3 in the background makes that if the conditions of rule 2 are satisfied, there is an exception to rule 1, which takes the rule conflict

---

6 A logical account of the operation of exceptions to rules can be found in Hage 1997.

7 A more satisfactory theory would take into account that exceptions to rules should be minimised. In the unpublished longer version of this paper this is accomplished by a number of technicalities which go beyond the scope of this paper.

8 The author is grateful to Bob Brouwer for pointing out a difficulty with an earlier version of this definition.

away. As a consequence, the rules 1 and 2 are consistent relative to a background that consists of rule 3.

Exceptions can also make a consistent set of rules inconsistent:

Example 5
rule_1 = *thief(x) ⇒ *punishable(x)
rule_2 = *minor(x) ⇒ *~punishable(x)
c = {rule_3 = *minor(x) ⇒ *exception(thief(x) ⇒ *punishable(x)),
rule_4 = *second_offender(x) ⇒ *exception(rule_3)}

We have seen in example 4 that the rules 1 and 2 are consistent against the background consisting of rule 3. The addition of rule 4 to the background makes that there is no guarantee anymore that there is an exception to rule 1 in case of a minor. This is illustrated by the case that John is not only a thief and a minor, but also a second offender. In that case there is an exception to rule 3, and presumably no exception to rule 1. This illustrates that if rule 4 is added to the background, there are cases in which the conditions of both the rules 1 and 2 are satisfied, and in which these two rules are in actual conflict.

7 The language L_{RL}

The notion of rule consistency will be formalised by model-theoretic means, that is in terms of possible worlds. Intuitively, a set of rules is inconsistent if there is a possible world in which the conditions of all the rules are satisfied, and there is no exception to either one of the rules, while there is no possible world in which the conclusions of all the rules obtain. Variations on the notion of consistency are realised by different characterisations of possible worlds.

In order to give such a formalisation I need to be able to talk about states of affairs and rules in the logical language that I employ. For that purpose I will use a special logical language L_{RL}, the language of *Rule Logic*. L_{RL}, is essentially the language of first order predicate logic, augmented with some conventions.[9] I assume that all formulas of L_{RL} have an uppercase initial, while all terms (including function expressions) have a lowercase initial. Variables of L_{RL} are *italicised*. For the meta-language of L_{RL} I use schematic sentences and terms, unless the contrary is clear from the context.

I adopt the convention that all well-formed formulas of L_{RL} *express* states of affairs. For every state of affairs expressible in L_{RL} there is a term that *denotes* it. In the case of atomic formulas this term is constructed by replacing the uppercase initial of the formula expressing the state of affairs by the corresponding lowercase initial and prefixing the resulting string by an asterisk (*). E.g. if Thief(john) expresses the state of affairs that John is a thief, then *thief(john) denotes the same state of affairs. All terms that start with an asterisk are assumed to denote states of affairs. In this way the terms of L_{RL} are subdivided into terms that denote states of affairs and terms that denote other individuals.

Some kinds of states of affairs tend to go together, while other ones exclude each other. For instance, the states of affairs that x kisses y tends to

9 L_{RL} is a subset of the language of Reason-based Logic. This language is described and motivated in Hage 1997, ch. IV.

go together with the state of affairs that *x* touches *y*, and the state of affairs that *x* is a circle tends (very strongly) to exclude the state of affairs that *x* is a square. These relations between (usually generic) states of affairs are called *constraints* on states of affairs. Rules, including legal rules, can be seen as a special kind of constraints.
The symbol ⇒ is used to denote constraints in general and rules in particular. ⇒ is an infix functor with two parameters, both for (usually generic) states of affairs. The first parameter stands for the rule conditions, the second parameter stands for the rule conclusion. If a rule has no conditions, the condition part is taken by the term *true. Examples of well-formed rule expressions are:

*minor(*x*) ⇒ *~punishable(*x*), and
*true ⇒ *forbidden(stealing)

Rule formulations are according to these conventions not well-formed sentences of Rule Logic. From a logical point of view they are function expressions that denote individuals.

It is possible to make statements about rules, such as the statement that a rule exists. The predicate Exists serves to express that a rule exists. It is defined by the following sentence:

Exists(*rule*) ≡ def. ∃*x*(*x* = *rule*)

Finally, the language L_{RL} has a one-place predicate Exception that ranges over rules and expresses that there is an exception to the rule in question for the case to which the rule is instantiated. For instance:

Exception(*thief(john) ⇒ *punishable(john))

8 Model theory for Rule Logic

Central in the model theory for rules is the notion of a constraint. Logical constraints hold in general for all *logically possible worlds*. These are, in the present context, the constraints of predicate logic, augmented with one additional constraint that characterises the logic of rules. Together, these constraints on all logically possible worlds are called the constraints of *Rule Logic*. I will treat rules as local constraints. Rules define what is possible and impossible *in a concrete world*, that is, in the world in which these rules exist.

The model-theoretic characterisation of worlds that are possible according to Rule Logic runs as follows:

CONSTRAINTS ON WORLDS THAT ARE LOGICALLY POSSIBLE ACCORDING TO L_{RL}
Let L_{RL} be the language of Rule Logic. $L_{RL} = \{S_1, S_2, ..., S_n\}$, where S_1 ... S_n are all the well-formed sentences of L_{RL}.
Let S_i be a sentence in L_{RL}, and let $*sa_i$ denote the state of affairs that is expressed by S_i. $*sa_i$ is then a state of affairs that is *possible relative to* L_{RL}.[10]

10 L_{RL} may be thought of as the conceptual scheme by means of which worlds are 'captured'.

Let the set SA be the set of all states of affairs that are possible relative to L_{RL}, and let W be the power set of SA. Intuitively, W stands for the set of all worlds, the content of which is expressible in L_{RL}. Every w ∈ W is a subset of SA.

There are no other constraints on the states of affairs that are elements of the worlds in W. There are, for instance, worlds in W in which the states of affairs *p and *~p both obtain. Such worlds are possible relative to L_{RL}, but they are not logically possible according to Rule Logic. Worlds that are logically possible are subject to a number of additional constraints. The set of these logically possible worlds is denoted by W_{RL}.

CONSTRAINTS ON WORLDS THAT ARE LOGICALLY POSSIBLE ACCORDING TO RULE LOGIC

1. if *p ∈ w then *~p ∉ w, and if *~p ∈ w, then *p ∉ w.
2. *p & q ∈ w if and only if both *p ∈ w and *q ∈ w.
3. *p ∨ q ∈ w if and only if either *p ∈ w, or *q ∈ w, or both.
4. *p → q ∈ w if and only if either *p ∉ w, or *q ∈ w, or both.
5. *p ≡ q ∈ w if and only if either both *p ∈ w and *q ∈ w,
   or both *p ∉ w and *q ∉ w.

These constraints correspond to the traditional constraints of propositional logic stated in terms of relations between states of affairs.

6. *∃*x*(r(*x*)) ∈ w if and only if there is an individual a, such that *r(a) ∈ w.
7. *∀x(r(*x*)) ∈ w if and only if there is no individual a, such that *r(a) ∉ w.

These constraints give the traditional meaning of the quantifiers, again stated in terms of states of affairs.[11]

A constraint that is characteristic for Rule Logic is the constraint that if the conditions of an existing rule are satisfied and there is no exception to this rule, the consequences of this rule obtain. Let *conditions/σ and *conclusion/σ denote the states of affairs expressed by respectively the conditions and the conclusion of a rule with their variables instantiated according to substitution σ. Then the above mentioned constraint becomes:

8. If *exists(*conditions ⇒ *conclusion) ∈ w, and
   *conditions/σ ∈ w, and
   *exception(*conditions/σ ⇒ *conclusion/σ) ∉ w, then
   *conclusion/σ ∈ w.

Finally there is a constraint to guarantee that terms that denote states of affairs expressed by logically equivalent sentences are co-referential:

9. If and only if for all worlds w ∈ W_{RL} it holds that *p ≡ q ∈ w, then *p = *q.

11 To gain simplicity at the cost of precision, the formulations of the constraints 6 and 7 do not deal with compound formulas, or the use of quantifiers or function expressions within the scope of the quantifiers.

9 Compatibility of states of affairs

Given the model theory for Rule Logic, it is possible to give a formal characterisation of rule consistency. The starting point is the characterisation of compatible states of affairs:

RELATIVE COMPATIBILITY OF STATES OF AFFAIRS
Let c be a set of constraints, and let W_c be the set of worlds $w \in W_{RL}$, such that for every constraint $c_i \in c$, it holds that $*exists(c_i) \in w$. The states of affairs in a set s are then said to be compatible relative to the set of constraints c if and only if there is some set of states of affairs $s' \in W_c$ such that $s \subseteq s'$.

LOGICAL COMPATIBILITY OF STATES OF AFFAIRS
The states of affairs in a set s are logically compatible if and only if they are compatible relative to the empty set of constraints: $s \in W_{\emptyset}$.

Let me illustrate this by means of the following example:

Example 6
s = { *thief(john), *~punishable(john),
*exception(thief(john) ⇒ punishable(john))}
c = { *thief(*x*) ⇒ *punishable(*x*)}

The first two states of affairs in s are logically compatible. They are, however, incompatible relative to c, because the conditions of the rule that thieves are punishable which is an element of c, are satisfied in all worlds in which the states of affairs *thief(john) obtains. This rule leads to the conclusion that John is punishable. However, the addition of the exception to this rule reinstates the compatibility of the states of affairs in s, because this exception blocks application of the rule in c.

10 The consistency of rules

By means of the notions of logical compatibility of states of affairs and compatibility of states of affairs relative to a set of constraints, it is possible to give a formal characterisation of rule consistency.

Let $r = \{r_1 \ldots r_n\}$ be a finite set of n rules, where $r_i = *conditions_i \Rightarrow *conclusion_i$, for i = 1 to n.

Let $s = \{\sigma_1 \ldots \sigma_n\}$ be a set of n instantiations for the variables that occur in r, where σ_i is applied to the variables in r_i. For instance, let r_3 be *thief(*x*) ⇒ *punishable(*x*), and let σ_3 be {*x* -> john}.

Then the instantiation of r3 by means of σ3, $inst(r_3, \sigma_3)$, is

*thief(john) ⇒ *punishable(john).

Let $I_{conditions}(r, \sigma)$ be the set of the instantiations by means of σ of the conditions of all rules in r. That is:

$I_{conditions}(r, \sigma) = \{inst(conditions_1, \sigma_1), \ldots inst(conditions_n, \sigma_n)\}$.
Let $I_{conclusion}(r, \sigma)$ be $\{inst(conclusion_1, \sigma_1), \ldots inst(conclusion_n, \sigma_n)\}$.

Let $I_{\sim exception}(r, \sigma)$ be
{ *~exception(inst(conditions$_1$, σ_1) ⇒ inst(conclusion$_1$, σ_1)) ...
*~exception(inst(conditions$_n$, σ_n) ⇒ inst(conclusion$_n$, σ_n))}.

Then the following definition captures the notion of rule consistency relative to a set of constraints:

RELATIVISED RULE CONSISTENCY:
The rules in the set s are consistent relative to a set of constraints c, if and only if it is not so that there is a set s' ⊆ s and a set of instantiations σ, such that
a. the set $I_{conditions}(s', \sigma) \cup I_{exception}(r, \sigma)$ is compatible relative to c ∪ r,
b. the set $I_{conclusion}(r, \sigma)$ is incompatible relative to c ∪ r.

The compatibility of the joint rule conditions and conclusions, and the absence of exceptions to the rules is judged against the background of both the set of constraints and the rules themselves, because the rules that are evaluated as to their consistency also impose constraints on the world in which they exist.

Logical consistency of rules
The rules in the set s are logically consistent, if and only if it is not so that there is a set s' ⊆ s and a set of instantiations σ, such that
a. the set $I_{conditions}(s', \sigma) \cup I_{exception}(r, \sigma)$ is compatible relative to r,
b. the set $I_{conclusion}(r, \sigma)$ is incompatible relative to r.

Let me re-use some examples of the sections 3 and 4 to illustrate these definitions:

Example 1
rule_1 = *thief(*x*) ⇒ *punishable(*x*)
rule_2 = *minor(*x*) ⇒ *~punishable(*x*)

The rules 1 and 2 are logically inconsistent, as is illustrated by the set of instantiations {σ1, σ2}, where σ1 = σ2 = {*x* → john}, with c = ∅.

Example 2B
rule_1 = *thief(*x*) ⇒ *punishable(*x*)
rule_2 = *minor(*x*) ⇒ *protected(*x*)
c = {*protected(*x*) ⇒ *~punishable(*x*)}

The rules 1 and 2 are also inconsistent relative to c because there can be no instantiation of *x* that makes the instantiations of the states of affairs *punishable(*x*) and *protected(*x*) co-obtain in a world in which the constraint in c exist.

Example 3B
rule_1 = *thief(*x*) ⇒ *punishable(*x*)
rule_2 = *minor(*x*) ⇒ *~punishable(*x*)
c = {*minor(*x*) ⇒ *~thief(x)}

The three rules in this example taken together are *logically* inconsistent. However, moving the third rule to the background changes the situation. The conditions of the rules 1 and 2 are not compatible relative to c, because there can be no instantiation of *x* that makes the states of affairs *thief(*x*) and *minor(*x*) co-obtain in a world in which the constraints in c exist. Therefore the rules 1 and 2 are consistent against the background of c.

Example 4
rule_1 = *thief(x) $\Rightarrow$ *punishable(x)
rule_2 = *minor(x) $\Rightarrow$ *~punishable(x)
c = {*minor(x) $\Rightarrow$ *exception(thief(x) $\Rightarrow$ punishable(x))}

The rules 1 and 2 are consistent relative to c, because there can be no instantiation of *x* that makes the state of affairs *minor(*x*) obtain in a world in which the constraint in c hold, and in which the state of affairs *exception(thief(*x*) ⇒ punishable(*x*)) does not obtain. As a consequence, there can be an instantiation of x such that *thief(*x*) and *minor(*x*) co-obtain in a world in which the rules 1 and 2 both exist in combination with the constraint in c.

11 Related and future research

To my knowledge, little work has been done on the characterisation of the consistency of rules. Most often, a seemingly related, but nevertheless altogether different question is discussed, that is the consistency of norms, where norms are conceived as deontic sentences. I have tried to use non-deontic examples in this paper, to illustrate that the issues at stake have nothing to do with deontics. Den Haan (1996) contains a discussion of 'normative conflict', which deals with deontic norms, but which nevertheless takes an approach related to the one of this paper. Her work is directed toward practical application in computer science, and is theoretically less developed. Moreover, it does not take exceptions to rules into account, at least not exceptions in the sense of this paper.

An obvious extension of this work is to include principles in the characterisation of consistency. This leads to a major complication, because colliding principles are not necessarily inconsistent if there is no exception to either one of them. Instead they might lead to reasons pointing in different directions (see Hage 1997). Another extension would be to take into account that the exceptions to rules should be minimised. This extension is included in the original version of this paper.

References

Den Haan 1966
N. den Haan, *Automated Legal Reasoning*, Doctoral Thesis, Amsterdam, 1996.
Hage 1997
J.C. Hage, *Reasoning with rules*, Dordrecht: Kluwer Academic Publishers, 1997.