

File ID 136082
Filename Chapter 6: The normative order

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type Dissertation
Title Legal theory, sources of law and the semantic web
Author A.W.F. Boer
Faculty Faculty of Law
Year 2009
Pages IV, 321

FULL BIBLIOGRAPHIC DETAILS:

<http://dare.uva.nl/record/305854>

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use.

Chapter 6

The Normative Order

6.1. Introduction

The function of the institution, to create and formalize normative order, is different from its structures and rules of change. The function of artifacts is generally more important than how they work: there are for instance considerably more people who know what a VCR is good for and how it is operated, than people who know how it works internally.

The legislator believes that by shaping the institution in certain ways certain beneficial effects are brought about in brute reality. The rules of the institution directly or indirectly reflect preferences of the legislator with respect to the behaviour of other agents and assumptions of the legislator about the preferences and abilities of other agents; The relevant mechanisms were described in sections 4.5.2, 4.6, and 4.7.

When we interpret the source of law in terms of agent roles and scripts, of obligation, or of subjunctive betterness we are in effect taking into account the intentions and expectations of the legislator.

Legal knowledge representation does not merely reflect the institutional character of the rules, but often also tries to capture the intended normative order: creating it is the function of the institution. The most obvious way in which the legislator creates normative order is through normative rules: in legal knowledge engineering these are often represented in deontic logics.

Section 6.2 presents the representation of normative rules. The normative rule *qua* normative rule presents us with little difficulty, but also fails to address planning in compliance with the rules.

While chapter 4 only discussed the effect of rules on behaviour in isolation, and chapter 5 discussed logical form and applicability, this chapter also introduces and discusses several phenomena – normative conflict (see section 6.3 and 6.3.1) and contrary to duty norms (see section 6.4) – that frequently occur in sets of norms *when interpreted as a coherent normative order*.

When normative rule n is considered in isolation a classification into cases in $|\exists \text{allowedBy}.\{n\}|$ and cases in $|\exists \text{disallowedBy}.\{n\}|$ suffices to guide behaviour; Because it is relative to the individual rule n only no such thing as normative conflict arises. When normative rules are however considered as a coherent set used to order alternatives in a concrete decision problem, normative rules can expose a betterness relation that allows for gradations of goodness, as the example of the contrary-to-duty norm shows. This betterness relation moreover need not be unambiguous even in legislation from a single legislator, as the existence of normative conflict shows.

A coherent explanation of these phenomena has to appeal to either the account in terms of normative positions, or in terms of a betterness relation, and one has to take the position that certain sets of normative rules are formalizations of coherent ideas of an intended normative order. As pointed out in chapter 4, these accounts are abstract theories *fitted on* normative rules.

This betterness relation cannot be a relation between OWL DL classes, however, as suggested by the informal explanation in section 4.6, but must be a straightforward OWL property to be named \preceq . Section 6.5 discusses this OWL property as if it were a new modal logic-based logic of betterness, in keeping with the interpretation of OWL

DL as a labeled multimodal logic in section 3.4. Because the account would only apply to specific sets of normative rules, the \preceq property is subscripted \preceq_n to make this distinction. The account in terms of betterness *at least* applies to each individual normative rule, but has little added value if only applied to individual rules.

Note that the same OWL translation issue would apply for a classical deontic account in terms of O and P modal operators, which would normally be interpreted as $\forall.P$ and $\exists.P$ for some relevant property P . Because OWL DL lacks reflexivity, required for a classical deontic logic, this would be a dead end.

Section 6.6 discusses the more general problem of having to choose between legal rules; choice rules allow or disallow certain choices. Since the choice between legal rules is often motivated by the ascribed intended normative order, i.e. the normative conflict, this chapter is the right place to introduce the choice rule (see section 6.6.3), which is a special kind of applicability rule. This chapter does not however try to give an account of when and why choice rules are introduced in cases of normative conflict: the applicability of a choice rule does *not* depend on a recognition of normative conflict. Section 6.6.2 points out an example where such choice arises without any form of conflict.

Section 6.7 discusses the intended normative order from a number of perspectives: the legislator's, the court's, and those of addressees of the source of law and those involved in transaction with them, who base preferences and expectations on the prevailing normative order as understood by them.

The rules of the institution have an instrumental function for those that engage in social interactions governed by those norms that goes beyond their interpretation as constraints on behaviour, as explained in section 4.5.2 and section 4.8. In section 6.7.1 the interpretation of rules as found in the source of law as ingredients for action scripts is discussed.

Since the legislator is neither omnipotent nor omniscient in his abilities to bring about normative order in brute reality, his attempts to do so will inevitably fail occasionally.

Intentions have a habit of coming to the foreground when attempts and failures matter: sometimes even courts appeal to the intentions of the legislator to explain how legislation should be applied (cf. for instance [19]). Perceived failures of the legislator – commonly in the form of perceived normative conflict – creates the space for courts to get involved in creating normative order themselves, as section 6.7.2 explains.

The courts are the legitimate authors of choice rules.

Section 6.7.3 considers the values the legislator tries to promote through the intended normative order: these unsurprisingly also take the form of propositions of subjunctive betterness, the *norms of analysis*. Section 6.7.3 points out a weakness in the prevailing view of comparing legislation based on ontology integration; Comparative law in actuality compares intended and actual normative orders relative to selected social scripts (as noted in section 4.5.1), instead of comparing institutional ontologies. If knowledge engineers limit themselves to the institutional interpretation of legal rules, no meaningful comparison can be made.

6.1.1. Contrast and Subjunctive Betterness

The reading of obligation $O(\alpha \mid \beta)$ given in section 4.6 is *if β , it would be better if α* , as opposed to $\neg\alpha$.

If we take the account of obligation to represent either a single normative rule or a selected set of them we have to make a refinement: $O_i(\alpha \mid \beta)$ means *if β , it would be better if α* , as opposed to $\neg\alpha$, with reference to axis i , which is the standard used for comparison.

In section 6.2 we will distinguish the readings *if $\beta \sqcap \neg\alpha$, it would have been better if $\beta \sqcap \alpha$* , and *if $\beta \sqcap \alpha$, no $\beta \sqcap \neg\alpha$ would have been better*. The same reading will be applied to preferences and values in section 6.7.

The normative rule involves four ingredients: the case to which it is applicable, the case which is disallowed, the case which is allowed, and the axis of reference being the identifier of the norm itself. Similar structures are found in *comparison theory*, *psychological geometry*, and *measurement theory*.

Levinson's comparison theory (cf. [188]) deals with *rhetorical comparison* or *simile*. Generally, a rhetorical comparison compares a thing A (the *primum comparandum*) to a thing B (the *secundum comparatum*) on the basis of a quality C (the *tertium comparationis*): A is like B with respect to C , or A is as C as B , versus A is different from B , A is less C than B , etc. Just like in deontic reasoning, the axis of reference or tertium comparationis (in deontic reasoning the norm) is more often than not missing in similes.

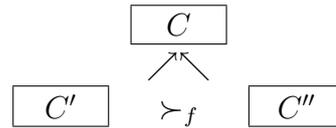
Kelly's *psychological geometry* (cf. [171]) is based on the supposition that we explain the world with theories built up from *personal constructs* that help us to find relevant contrasts (e.g. hostile vs. friendly, movable vs. immovable, etc.) relative to an *axis of reference*. According to Kelly "A person's [*mental*] processes slip into the grooves which are cut out by the mechanisms he adopts for realizing his objectives" (cf. [171], p. 49): to explain the situation is to make distinctions using these personal constructs. Kelly explicitly contrasts the better thing against its worse alternative(s). Measurement theory is similarly based on contrasts to define scales of measurement (cf. [242] for an overview and relevant references).

Repertory grids and associated methods (cf. [113]) for *knowledge acquisition* are based on psychological geometry: one for instance takes a topic concept, imagines three salient instances of it, and then tries to think of a distinguishing criterion present in two instances, but clearly absent in the third one, and then classifies these instances into two contrasting subconcepts along the exposed axis. To the subcategories the same procedure can be applied to produce a kind of taxonomy.

The criteria people come up with usually have an evaluative flavour: they distinguish a category that is "good" and a category that is "bad" for some purpose. The goodness of the category is obviously not a property of the category, but a judgment relative to an intended purpose of applying the construct to something. Constructs represent assessment criteria, and the taxonomies produced in this way reflect the way in which these criteria are prioritized: they are in effect handcrafted decision trees – perhaps even a kind of *scale* to assess things against – rather than true taxonomies, combining subsumption and betterness as organizing principles for knowledge

representation.

The result of such knowledge acquisition exercises are triangles between a general concept C and two subtypes C', C'' of it that 1) are disjoint ($C' \equiv \neg C''$), 2) completely partition the subsuming concept ($C \equiv C' \sqcup C''$), and 3) generally indicate that instances of C' are better than instances of C'' ($C' \succ C''$) for some purpose, or for the realization of some function f , graphically represented as:



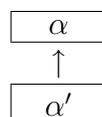
Interestingly, one of the mistakes commonly made by students who learn to build ontologies in a description logic like OWL DL is that they assume that subclasses are disjoint and fully partition the superclass by default, and are surprised that the reasoner fails to come to the expected conclusions (cf. [233]; also known from personal experience and that of colleagues).

Psychological geometry should certainly not be advanced as a method for designing ontologies: the reader will recognize that psychological geometry maps out the epistemological purposes of concepts and properties rather than a conceptualization of a domain detached from its immediate context of use. Psychological geometry is for instance a good method for building simple decision trees by hand, and structured comparison of sets of three instances is a good conversational tool for knowledge acquisition. It is also a suitable basis for a representation of subjunctive betterness.

In LKBS relevant contrasts also play a central role. Particularly interesting is for instance the notion of case factors favouring defendant or plaintiff in case law (cf. [18, 19] for a clear formulation, or section 6.7.2). A factor is not a plain fact: the facts of the case present in the description given in a verdict are present exactly *because* they favour one of the sides, and they are therefore factors. The justification of the case decision obviously reveals selective attention to those facts deemed to be (potentially) relevant for the decision, and they are relevant for the decision because they are relevant for the normative order the legal system intends.

Since repertory grids and decision trees are more or less recognizable conventions, and moreover seem to have a certain intuitive validity, this chapter, in particular section 6.2, will use Kelly's triangles liberally as a conceptualization of betterness. Note however that Kelly's triangles suggest a preference relation on propositions, and is therefore an epistemic account that requires reification of concepts. In OWL DL this reification is not allowed.

This chapter uses some notational conventions in addition to those introduced in chapter 3 for the use of Kelly's personal construct triangles. Because subsumption relations play a key role in this discussion, the following graphical convention for displaying a subsumption relation $\alpha' \sqsubseteq \alpha$ is occasionally used:



The set $|\phi|$ is the set of OWL individuals i such that $M, i \models \phi$. Another notational convention is used to make clear that a statement $|\beta \sqcap \alpha| \succ |\beta \sqcap \neg\alpha|$ is an ordering on α and $\neg\alpha$ within the context of β :

$$\beta : \boxed{\alpha} \succ \boxed{\neg\alpha}$$

An alternative notation is $|\beta \sqcap \alpha| \succ |\beta \sqcap \neg\alpha|$ or the following:

$$\boxed{\beta \sqcap \alpha} \succ \boxed{\beta \sqcap \neg\alpha}$$

Personal construct triangles are however only used for explanatory purposes.

6.2. Representation of Normative Rules

The normative rule is a legal rule that constrains assertions of the form `disallows(n, a)` and `allows(n, i)`. The properties `allows` (inverse `allowedBy`) and `disallows` (inverse `disallowedBy`) represent a relation between legal rules and actions. Because of these domain restrictions, and the obvious disjointness of legal rule and action, the properties are irreflexive and asymmetric.

The two properties are disjoint, and subproperties of `appliesTo` (inverse `applicable`). If a rule allows something, it does not disallow it. If a rule allows or disallows something, it is also applicable to that thing.

The ontology fragments introduced here are largely part of the *LKIF* ontology (cf. [35]), although this chapter includes some new notions.

Normative rules are modeled by *constitutive* rules that make the normative rules *apply* to actions, and qualify them as allowed or disallowed. These rules are distinguished from other constitutive rules by their function: they separate the good from the bad, and the worse from the better, from the perspective of the legislator. Their function is to influence choice: the addressee of the rule is to choose allowed alternatives over disallowed ones. Normative rules are intended to have a *deontic choice* effect.

Constitutive rules normally constrain assertions of the form `constitutes(b, i)`, where b is a brute thing and i an institutional one. One could think of the forms `disallows(n, a)` and `allows(n, i)` as alternative renderings of `constitutes(i, v_n)`, `constitutes(i, u_n)`, where v_n is a violation of n and u_n a use of n . This representation would however be needlessly verbose and complicated.

The representation of normative rules proposed here is based on a slightly different mapping from deontic operators to the ordered set $\{allowed, silent, disallowed\}$ than the one in [265], but is otherwise inspired by it. Just like in the treatment of applicability given in section 5.2.2 I am however interested in representing explicitly *which* legal rule was violated.

The purpose of the representation is create the following mapping:

$$v(c) = \begin{cases} \textit{disallowed} & : O(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \neg\alpha \mid \\ \textit{disallowed} & : F(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \alpha \mid \\ \textit{allowed} & : O(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \alpha \mid \\ \textit{allowed} & : F(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \neg\alpha \mid \\ \textit{allowed} & : P(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \alpha \mid \\ \textit{silent} & : P(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \neg\alpha \mid \\ \textit{silent} & : \textit{otherwise} \end{cases}$$

The difference is mainly relevant in relation to so-called *compliance conflicts* (cf. section 6.3). The view that prohibitions and obligations give rise to a *strong permission*, contrary to the views of Kelsen in [172], Lindahl in [189], and Valente in [266], who believe that the obligation gives rise to a *weak permission*, will be defended in section 6.3.1.

The representation of normative rules models each legal rule by multiple constitutive ones. All rules state something *about* n instead of purporting to be *the* representation of n , taking into account the criticism expressed in section 2.2.2, and easily accommodate alternative representations of the things allowed or disallowed. This distinction between the legal rule and the logical propositions describing it was not explicitly made in [265].

Proposition 26. For each legal rule n interpreted as an obligation $O_n(\alpha \mid \beta)$ or prohibition $F_n(\neg\alpha \mid \beta)$, create three necessary conditions and three indicators (r_1, r_2, r_3):

$$\begin{aligned} & \{n\} \sqsubseteq \forall \textit{appliesTo}.\beta \\ & \{n\} \sqsubseteq \forall \textit{disallows}.\beta \sqcap \neg\alpha \\ & \{n\} \sqsubseteq \forall \textit{allows}.\beta \sqcap \alpha \\ (\text{DEFAULT } r_1 & (\textit{known}.\beta)(\textit{free}.\exists \textit{applicable}.\{n\}) \\ & (\textit{assume}.\exists \textit{applicable}.\{n\}) \\ (\text{DEFAULT } r_2 & (\textit{known}.\beta \sqcap \neg\alpha)(\textit{free}.\exists \textit{disallowedBy}.\{n\}) \\ & (\textit{assume}.\exists \textit{disallowedBy}.\{n\}) \\ (\text{DEFAULT } r_3 & (\textit{known}.\beta \sqcap \alpha)(\textit{free}.\exists \textit{allowedBy}.\{n\}) \\ & (\textit{assume}.\exists \textit{allowedBy}.\{n\}) \end{aligned}$$

Proposition 27. For each legal rule n interpreted as an obligation $P_n(\alpha \mid \beta)$, create two necessary conditions and two indicators (r_1, r_2):

$$\begin{aligned} & \{n\} \sqsubseteq \forall \textit{appliesTo}.\beta \\ & \{n\} \sqsubseteq \forall \textit{allows}.\beta \sqcap \alpha \\ (\text{DEFAULT } r_1 & (\textit{known}.\beta)(\textit{free}.\exists \textit{applicable}.\{n\}) \\ & (\textit{assume}.\exists \textit{applicable}.\{n\}) \\ (\text{DEFAULT } r_2 & (\textit{known}.\beta \sqcap \alpha)(\textit{free}.\exists \textit{allowedBy}.\{n\}) \\ & (\textit{assume}.\exists \textit{allowedBy}.\{n\}) \end{aligned}$$

The syntactical operations translate normative rules represented by the source of law, interpreted as obligatory and permissive rules, respectively. Just like the formalism in [265] these should not be interpreted as more general accounts of obligation and permission. All rules depend on only three ingredients: n, β, α . It is trivial to generate them syntactically from a user interface implementing just one representation that involves all three ingredients, for instance the classical $O_n(\alpha \mid \beta)$.

Recall from section 4.7 that normative statements in principle always address actions, although the intended category of actions may only be described in terms of a relevant result and not the way in which it was brought about. The formulation n : *in situation β you ought to do α* for instance becomes:

$$\begin{aligned}
& \{n\} \sqsubseteq \forall \text{appliesTo} . \exists \text{situation} . \beta \\
& \{n\} \sqsubseteq \forall \text{disallows} . (\exists \text{situation} . \beta) \sqcap \neg \alpha \\
& \{n\} \sqsubseteq \forall \text{allows} . (\exists \text{situation} . \beta) \sqcap \alpha \\
& (\text{DEFAULT } r_1 \text{ (known} . \exists \text{situation} . \beta) (free} . \exists \text{applicable} . \{n\}) \\
& \quad \text{(assume} . \exists \text{applicable} . \{n\}) \\
& (\text{DEFAULT } r_2 \text{ (known} . (\exists \text{situation} . \beta) \sqcap \neg \alpha) (free} . \exists \text{disallowedBy} . \{n\}) \\
& \quad \text{(assume} . \exists \text{disallowedBy} . \{n\}) \\
& (\text{DEFAULT } r_3 \text{ (known} . (\exists \text{situation} . \beta) \sqcap \alpha) (free} . \exists \text{allowedBy} . \{n\}) \\
& \quad \text{(assume} . \exists \text{allowedBy} . \{n\})
\end{aligned}$$

The proposition $\exists \text{situation} . \beta$ should be interpreted as *doing something in situation β* .

Instead one may find other types of constraints on action, for instance a modification of the action, n : *when doing A you ought to do it in a A' manner*, or indication of a change, for instance n : *in situation S , S' should be initiated* or n : *situation S should be terminated*, that are all self-evident.

This representation is very similar in form to Valente's in [265], although [265] expressed the opinion that norms apply to situations.

Straightforward case assessment can be performed based on these constitutive rules alone. We can determine which known rules allow, and which known rules disallow, something s through a query $M_{\mathcal{DL}} \models \text{allows}(n, s)$, $M_{\mathcal{DL}} \models \text{disallows}(n, s)$, for all n in the abox. In [265] a criterium for recognizing conflict (informally discussed in section 6.3.1) between normative rules is supplied and a simple conflict resolution method is given. This method could in principle be applied here, which we will not do.

The approach presented here has the advantage over [265] that it is completely represented in the description logic: relevant subsumption relations between components of normative rules are automatically computed.

This would however require us to make a distinction not made so far if the conflict resolution method in [265] would be used: the norms subsume each other only with respect to their propositional content.

Subsumption between normative rules would be a bigger problem. Let n and m be two different norms. It is not sensible to think that $\{m\} \sqsubseteq \{n\}$: this would mean

that $m = n$. So there must be a generalized norm M and N , such that $\{n\} \sqsubseteq N$ and $\{m\} \sqsubseteq M$, and $M \sqsubseteq N$, and to these all these axioms about content apply, while specifics of $\{m\}$ and $\{n\}$ do not. M and N are the *norms qua qualification*. These must strictly speaking be different rules: surely N is an abstraction of $\{n\}$.

It is important to realize (as pointed out in section 3.4) that in OWL logical constants – in reality uniform resource identifiers – like n are not bijective. Given two norms n and m , it is permitted to state that $\{m\} \equiv \{n\}$ which will only succeed if the combination of both norms is a consistent norm. This line of exploration will however not be followed any further in this book.

The use of constitutive rules instead of institutional ones for normative rules is in first instance a principled one. Whether some state of affairs is allowed or disallowed can hardly be considered to be defining of the structures a legal institution can take on. The institution can certainly take on forms disallowed by some normative rule. It is not ontology, but an opinion on the relative desirability of the things qualified.

There are however certain logical relations between allowed and disallowed, *within the same axis of reference*, that are ontological in character: the same thing is not allowed and disallowed, and things allowed are better than things disallowed, relative to the same axis of reference.

This is a reason to place the allowed vs. disallowed distinction in its own institutional reality, just like Searle does by distinguishing between the institution of chess and the institution of competitive game playing to which winning and losing belongs (cf. [248]). One also cannot win and lose *the same game*.

There is also a practical precautionary reason to reject the interpretation of normative qualification rules as (potentially) institutional rules. Normative rules are clearly subject to applicability conditions and an arbitrary requirement that can function as exclusion grounds: inference to violation of a normative rule is therefore at least potentially defeasible without the existence of normative conflict.

The normative rules are constitutive, but the notion of *normative conflict* is *not* addressed as a standard default reasoning feature as is often the case in nonmonotonic deontic logics. The rules only state that some things are (dis)allowed by n ; No claim is made about a normative order as a whole (in keeping with Makinson's argument quoted in section 2.2.2), and normative rule n presumably never conflicts internally. Normative conflicts only give rise to defeasibility subject to certain assumptions.

The only type of logical inconsistencies to be expected without making assumptions about normative order involve contradictions between for instance a normative rule n : *blue cars are not allowed in this parking area* and an applicability rule *rule $\{n\}$ only applies to red cars* or some other requirement. The normative rule is subject to exceptions, but these are different in character than normative conflicts.

The method given in [265] did not address applicability conditions and exceptions at all. Instead these had to be expressed as more specific normative rules, forcing a normative conflict.

When we use a set of constitutive rules as proposed above, we do not account in any form for an interpretation in terms of betterness. The qualifications allowed and disallowed however have a logic of their own, because they are evaluative concepts.

As stated above, but also in section 6.1.1 about contrasts, we expect the set of

things that are allowed $|\exists\text{allowedBy}\{n}\ |$ and the set of things that are disallowed $|\exists\text{disallowedBy}\{n}\ |$ by a single normative rule to be disjoint. This property is inherent in the syntactical translation proposed.

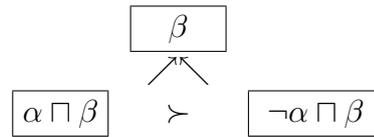
We moreover expect that things allowed are better than things disallowed, relative to the same axis of reference. This expectation, important for complying with the normative rules in planning applications is a lot harder to enforce in OWL DL. For now we write down this expectation as follows, where \succ_n is a relation *better relative to axis of reference n*:

$$|\exists\text{allowedBy}\{n}\ | \succ_n |\exists\text{disallowedBy}\{n}\ |$$

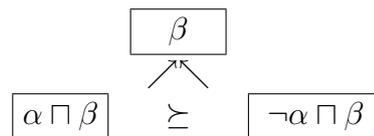
In planning we generally have to comply with multiple relevant rules. If we interpret these rules as *a set* of ordering constraints on alternative plans – in lieu of actions to which the rules apply – we have to deal with two important complications: firstly that the betterness relation is not antitransitive, which would be convenient, and – more importantly – that it is also not actually asymmetric.

Section 6.3 informally discusses normative conflict, a special property of some occurrences of symmetry in the betterness relation, and section 6.4 discusses contrary-to-duty norms, which are cases of apparent transitivity of the betterness relation.

In the following sections I will interpret a normative rule $O_n(\alpha | \beta)$ or $F(\neg\alpha | \beta)$ as:



The permissive normative rule cancels an obligation or prohibition, hence $P_n(\alpha | \beta)$ is interpreted as:



Note that it is the practical predicament of planning, and expectations about the rationality of the legislator, that suggests that certain *sets* of normative rules *should* represent a coherent betterness relation. The \prec and \preceq relation must in reality be a relation between individuals, and be restricted to for each individual normative rule n to domain $|\exists\text{disallowedBy}\{n}\ |$ and range $|\exists\text{allowedBy}\{n}\ |$.

Section 6.5 proposes a betterness relation that *can* be used in OWL DL on sets of norms. It does not allow for normative conflict, which is in my view not commensurable with accounts of constraints in planning in terms of obligation. The existence of normative conflict rather proves that the account in terms of obligation does not fit on a certain set of normative rules, in the sense discussed in section 2.3.1. The contrary-to-duty obligation on the other hand is a normal thing that should be accounted for. In section 6.5.1 normative conflict, and the circumstances in which it can be said to arise, is discussed in relation to the formalism proposed in section 6.5. The solution

for dealing with normative conflict will only be introduced in section 6.6.3, and this solution does bring back defeasibility through the back door through the choice rule.

Large parts of the following three sections were published in [44, 43, 35], although this chapter also contradicts some of the conclusions of these publications.

6.3. Normative Conflict

An instrumental concept in explaining conflicts between norms is that of *realizability*; A norm is *realized* if the state of affairs it allows or disallows is the case. There is a conflict between a pair of norms if they are not *jointly realizable*.

A clear logical formulation of normative conflict is not easy to give: very different conceptions of what constitutes a normative conflict are given in the relevant literature. This section gives an informal exposition of types known in literature.

A distinction is usually made between so-called conflicts of disaffirmation and compliance conflicts. Lindahl (cf. [189]) defines disaffirmation as follows: “a relation between two norms of different deontic mode, one being permissive and the other mandatory”.

A disaffirmation conflict is in our context a circularity between a permission and either an obligation or prohibition. This occurs if a state of affairs is simultaneously allowed and disallowed. The first such situation – a conflict of disaffirmation between $O(\alpha|\beta)$ and $P(\neg\alpha'|\beta')$ where $\alpha' \sqsubseteq \alpha$ and $\beta' \sqsubseteq \beta$ – can be represented as follows:

$$\begin{array}{ccc} \beta : & \boxed{\alpha} & \succ & \boxed{\neg\alpha} \\ \uparrow & \uparrow & & \downarrow \\ \beta' : & \boxed{\alpha'} & \prec & \boxed{\neg\alpha'} \end{array}$$

This is called a *disaffirmation of an imperative* (cf. [30]). Intuitively it is meant as an exception and takes precedence to the primary obligation, but this is not necessarily the case as we will see in section 6.6. Some simple examples:

(1) Parking is prohibited in this neighbourhood. (2) License holders are allowed to park in the designated bays.

(1) Walking on the lawn is prohibited. (2) Park maintenance department employees are permitted to mow the lawn.

The second such situation – a conflict of disaffirmation between $P(\alpha|\beta)$ and $O(\neg\alpha'|\beta')$ where $\alpha' \sqsubseteq \alpha$ and $\beta' \sqsubseteq \beta$ – can be represented as follows:

$$\begin{array}{ccc} \beta : & \boxed{\alpha} & \succ & \boxed{\neg\alpha} \\ \uparrow & \uparrow & & \downarrow \\ \beta' : & \boxed{\alpha'} & \prec & \boxed{\neg\alpha'} \end{array}$$

This is called a *disaffirmation of a permission*. Intuitively this case should be handled in the same way as the previous one, with the disaffirming norm taking precedence. A simple example:

1. Parking is allowed in the parking bays.

2. Parking in the parking bays is not allowed for cars wider than 1,85m.¹

Disaffirmation of permissions and imperatives is a very useful tool for legislators. It allows the legislator to create exceptions to rules in sources of law, without amending sentences with yet another sentence fragment. Since sentences in legislation are already too long as it is, according to many, this is certainly a useful thing. Disadvantage of doing this is that it makes it harder to apply rules, because the task of checking whether there are any such exceptions anywhere in the legal corpus is left as an exercise to the reader.

Another purpose of the disaffirmation is to influence the legislation of another legislator without actually touching its legislation. If legislator A is superior to legislator B, a disaffirmation in the legislation of A will effectively amend the legislation of B. This function was explained in general terms in relation to *delegation*. Because the amendment now ends up in a completely different part of the relevant legal corpus, this type of exception will be hard to find.

This does not mean that disaffirmation conflicts are always merely a design feature of legislation, or that the disaffirmation conflict is the same thing as an exception. There are several other types of disaffirmation conflicts that have no intuitive solution, or clear purpose. The first such situation – a conflict of disaffirmation between $O(\neg\alpha|\beta')$ and $P(\alpha'|\beta)$ where $\alpha' \sqsubseteq \alpha$ and $\beta' \sqsubseteq \beta$ – can be represented as follows:

$$\begin{array}{ccc} \beta : & \boxed{\alpha'} & \succ \boxed{\neg\alpha'} \\ \uparrow & \downarrow & \uparrow \\ \beta' : & \boxed{\alpha} & \prec \boxed{\neg\alpha} \end{array}$$

The following two simple rules exemplify this case:

1. Using network facilities in the classrooms is prohibited.
2. Using WiFi in the university building is permitted.

It is not clear which of the two rules is an exception to the other one, granted that we share the belief that using WiFi is subsumed by using network facilities, and being in the classrooms is subsumed by being in the university building. The second such situation is created by reversing the deontic modalities in the example:

1. Using network facilities in the classrooms is permitted.
2. Using WiFi in the university building is prohibited.

This is a conflict of disaffirmation between $O(\neg\alpha'|\beta)$ and $P(\alpha|\beta')$ – where $\alpha' \sqsubseteq \alpha$ and $\beta' \sqsubseteq \beta$ – and it can be represented as follows:

$$\begin{array}{ccc} \beta : & \boxed{\neg\alpha'} & \succ \boxed{\alpha'} \\ \uparrow & \uparrow & \downarrow \\ \beta' : & \boxed{\neg\alpha} & \preceq \boxed{\alpha} \end{array}$$

These situations are certainly conflicts from a *joint realization* point of view, and would certainly have been recognized as such by Lindahl (viz [189]) or Hill (viz. [147]). Arguably they belong to the Hill's *intersection conflicts* (cf. [105, 147]).

¹Inspired by the SUV amendment of 20 October 2004 of the town council of Nijmegen.

These cases defy Valente's formalization of normative conflicts (cf. [266]), which would not be able to determine which one is an exception to the other one. Any solution would rely on the distinction between settled context and that which is obliged that is typical for an ought-to-do perspective on norms (cf. section 4.7.1).

A special case is the *explicit disaffirmation*: the permission and the imperative disaffirm each other directly. This obviously happens if and only if $\alpha' \equiv \alpha$ and $\beta' \equiv \beta$. All four of the previous characterizations are applicable to this pair of norms. Intuitively this does not make sense at all. It is however in principle valid, provided that the two norms have been enacted by different legislators that are apparently disputing jurisdiction.

The compliance conflict is logically less transparent. The compliance conflict is defined by Lindahl as a relation between two mandatory norms, both of which are individually realizable, but not jointly. Simply put, the compliance conflict gives you a choice where none of the options is allowed, or all of them are obliged.

This situation – a *compliance conflict* between $O(\alpha|\top)$ and $O(\neg\alpha'|\top)$ where $\alpha' \sqsubseteq \alpha$ – can be represented as follows:

$$\begin{array}{ccc} \boxed{\alpha} & \succ & \boxed{\neg\alpha} \\ \uparrow & & \downarrow \\ \boxed{\alpha'} & \prec & \boxed{\neg\alpha'} \end{array}$$

This essentially makes it a special case of disaffirmation. True compliance conflicts appear to be rare. Consider the following interesting real-world example of something which appears to be a compliance conflict: The Amsterdam police at some point ordered nightclub owners to keep emergency exits locked to keep drugs out, while the fire department ordered the same nightclub owners to keep them unlocked to allow for escape in case of disaster. Newspapers considered this a compliance conflict, and this one does indeed invite a logical representation that would make it one.

Whether this is a compliance conflict in the sense of “all alternatives are disallowed” however depends on how we conceptualize the alternatives of the nightclub owners. Strictly speaking, the nightclub owners did have the possibility of admitting much less customers, so that no emergency exits are required, or even to stop running a nightclub. Surely the correct reading of the obligation is that one should not initiate the situation of hosting a large number of customers and having one's emergency exits locked, or unlocked respectively. It only appears to be a case of “all options are disallowed” if one considers certain propositions to be part of a settled context, i.e. one does not have to consider alternatives that involve not admitting the same number of customers in one's nightclub as one did before. But where does this allowance for not considering that alternative come from? This kind of argument would not fly in many other contexts.

The treatment of compliance conflicts in automated reasoning systems as logical inconsistency is not really adequate, because the judgment call will depend on the inevitability, legality, and desirability of the settled context, and whether the agent appears to have intentionally or recklessly created a situation of perceived compliance conflict. Many perceived compliance conflicts are not compliance conflicts in a logical

sense, but only become one in a specific decision making setting in a settled context. This point of view will be developed further in section 6.3.1.

The conflicts or circularities between norms discussed in this section so far do not depend on the particulars of the case at hand in this formulation. The subsumption relationship between two norms does not come into existence in the context of a specific case. The simple fact that the same behaviour is permitted by one norm and prohibited by another, does not in any way imply a conflict. Translating norms to subjunctive betterness, and therefore conflicts to circularities, makes this clear, but the idea that a conflict arises between two norms with respect to a specific case is widespread.

In [266], Valente proposed a computation of the *prime implicant* (cf. the definition of prime implicant in section 3.3 or [215]) of the conflict. While Valente was right to observe that not any pair of norms of different deontic mode applying to the same situation point to conflict, it is unnecessary to involve an actual situation to be assessed in the computation, as observed in [286]. There must be in fact a straightforward subsumption relation between (relevant components of) the rules.

This means that to check for conflicts in a normative system, one does not need a batch of cases describing situations. Conflicts can be found – and usually resolved – using the norms alone. This cuts down dramatically on the amount of computation required for the solution in [266], because subsumption can be calculated offline and the results of it stored in a database for use by LKBS. Also, this point of view, which requires some more explanation, will be developed further in section 6.3.1.

As I have argued in [30], it is also extremely useful to recognize *affirmation of an imperative* and *affirmation of a permission* as a tool for structuring legislation. These closely related concepts are also based on subsumption, but do not involve a circularity or conflict. An affirmation of an imperative assumes two norms of the same mandatory deontic mode in a subsumptive relationship, and affirmation of a permission involves two norms of permissive mode. Affirmation as a logical instrument plays a role in the so-called *ne bis in idem* principle explained in section 6.6.

Example 12. An imperative for designing ships for instance states that equipment of any kind on ships should be designed “in such a way as to minimize potential damage of mishandling”. This norm is also *affirmed* in many places by other, more detailed, imperatives relating to specific bad user interfaces of specific types of equipment. It is for instance not allowed to use a single piping system with faucets to transport both water for the fire pumps and fuel oil for the propulsion system. Mistakes with faucets are easily made in a panic. It is likely the more general norm has been added because new ways of mishandling equipment are invented regularly as technology progresses, and designers are admonished to think of these scenarios before the ship is built. (example from [30]).

The previous characterizations of conflict are adequate when we consider only pairs of normative statements from law, but there are arguably other situations which are also conceptualized as a conflict that are not covered by this characterization. In some cases a pair of normative statements is realizable, but only in a way that is not satisfactory with respect to the intentions of the legislator. Perhaps we can tentatively

refer to this as the *conflict with legislator's intentions* scenario. One such scenario, constructed by Elhag et al. (viz. [105]), involves two permissions:

There seem to be other types of conflict as that between the permission for A to live in a certain house and a permission for B to destroy that same house. These conflicts need our attention and have to be embodied in a theory on normative conflicts.

It is trivial to think of similar situations involving two different agents and two norms. This situation is actually jointly realizable (to the detriment of A), and neither of the agents is confronted by a circularity in making his choice. In addition, both agents are free to act or not to act on the permission. As such, it is not really a conflict between the two involved norms, but everyone would agree that the resulting situation is awkward if both agents act on the permissions.

Another situation of perceived conflict arises if the legislator issues a norm that is *impossible to realize* or its realization does not depend on any agent choice. The legislator cannot order that dice should always land on six eyes, it cannot prohibit volcanoes from exploding, it cannot amend the law of gravity, it cannot prohibit fatal accidents, etc. If you ought to do something, you must have the ability to do that thing.

A last oddity is the lack of ability to apply a normative rule, of recognizing that it is violated or complied with. In the Netherlands there is a crime in the books, grievous blasphemy², that passed into desuetude when a court asked to apply it argued, with reference to the constitutional freedom of religion, that it found it impossible to decide what attributes are to be considered grievous when applied to a deity if the defendant denies having had the intention to grieve a deity. This is the more general situation of an indicative condition that never arises in brute reality, since blasphemers generally speaking do not have the intention to grieve the deity itself if the intention was to grieve in the first place. The rule was however never removed from the corpus since it is deemed to still have a function as moral guidance.

6.3.1. Compliance and Brute Reality

The traditional deontic axiom that states that everything that is obliged is allowed has always been a point of contention in legal theory and legal knowledge engineering. Kelsen in [172], Lindahl in [189], and Valente in [266] have stressed that obligations and prohibitions, besides classifying certain behaviours as disallowed, only render a *weak permission* to do the opposite of what is disallowed. In this view only the explicit *strong permission* classifies behaviours as allowed. Hence the following value function v mapping from descriptions to the ordered set $\{allowed, silent, disallowed\}$ (compare to the first table in section 6.2):

²Wetboek van Strafrecht, art. 147

$$v(c) = \begin{cases} \textit{disallowed} & : O(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \neg\alpha \mid \\ \textit{disallowed} & : F(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \alpha \mid \\ \textit{silent} & : O(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \alpha \mid \\ \textit{silent} & : F(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \neg\alpha \mid \\ \textit{allowed} & : P(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \alpha \mid \\ \textit{silent} & : P(\alpha \mid \beta) \wedge c \in \mid \beta \sqcap \neg\alpha \mid \\ \textit{silent} & : \textit{otherwise} \end{cases}$$

What this comes down to is that if you approach a T-junction and you see a sign saying that turning left is disallowed, then you have only a weak permission to turn around or turn right. Kelsen, Lindahl, and Valente do however recognize the existence of *compliance conflicts* (cf. section 6.3). If turning right or around is also prohibited, there is a conflict. These are however not considered conflicts between the implicit classification of the opposite of going left as allowed, and the classification of going right or turning around as disallowed, but as cases of joint unrealizability: two opposite behaviours have the same classification of disallowed. Valente's theory uses this notion of opposite to define the compliance conflict. There is no disagreement that one obligation may be used as a valid argument to violate another obligation.

The problem here is first and foremost a difference in conceptualization of conflicts, and secondly of what is the *opposite* of a behaviour.

If normative statements are not conditionalized, i.e. a statement $F\alpha$ is allowed to occur, the notion of a strong permission from a prohibition for instance is bizarre: the permission of the opposite $P\neg\alpha$ would cover almost everything. A prohibition stating that one is not allowed to walk on the grass surely doesn't allow walking anywhere else? The rejection of the implied permission because of the open-ended nature of the logical complement of the behaviour that is disallowed is however purely an esthetic one: the notion of "opposite behaviours" with the same classification creates exactly the same logical problem, but less obviously. This is however a red herring.

The fundamental difference between the accounts given in section 6.3 and [266] is the operationalization of conflict. The operationalization of section 6.3 is based on a logical conflict between two norms, applied to the same behaviour, regardless of the case at hand. The compliance conflict and the conflict of disaffirmation are really two faces of the same thing.

Valente's operationalization is far more general, and would indeed give bizarre results if obligations and prohibitions yielded strong permissions. In Valente's operationalization, the two involved norms do not have to stand in a subsumption relation to each other. In Valente's system, norms are applied not to a behaviour but to the *description of the case*: the thing classified by the norm is the *prime implicant* (cf. section 3.3) of the *generic case* description in the norm. This case description may describe (the results of) two or more different behaviours. Moreover the notion of prime implicant would only apply to the *description* of a behaviour, while the norms apply to the behaviour, at least in the view developed here.

What is required is that there is such a prime implicant case c which is classified as disallowed by some norm, while its negation is also a prime implicant case of another norm that classifies it as disallowed. Alternatively, for conflicts of disaffirmation,

required is that there is a prime implicant case c of two norms, one yielding disallowed and the other allowed. These two norms do not however have to be logically related.

The restriction to prime implicants obviously is obviously necessary to weed out false positives like “today I went to vote (which is strongly permitted) and then I shot everyone at the voting station (which is prohibited)”. But there is also a more profound truth found here. The question is whether the cases that would be covered by the prime implicant criterion, but do not involve a subsumption relation between normative statements, are true conflicts. This book takes the point of view that this is *not* the case, but there is a big problem with the restriction of irreconcilable *conflict* to subsumption only. The prime implicant criterion does pick up a category of cases that will sometimes be considered conflicts, but it also picks up false positives.

Clearly, the criterion “more specific” is wider than logical subsumption. See for instance also [168] for the same observation relating to prioritized defaults and the specificity criterion: intuitively one may want to consider a preference for *a film directed by Almodavar* as more specific than a dislike of *a Spanish film*, even though 1) they clearly aren’t in a subsumption relation to each other and 2) the whole notion of *a Spanish film* becomes underspecified when we consider Spanish directors, producers, cast, script, locale, etc. More specific in this sense is perhaps better understood as a refined level of granularity (as intended in section 2.3.1): there is a level of granularity at which the two norms would appear to be in conflict. This expanded notion of specificity is not reconcilable with the interpretation of obligations and prohibitions as overly general strong permissions, when considered in combination with conflict resolution strategies like *lex superior* and *lex posterior*³.

It is essential to note that the fact that two legal rules are in “irreconcilable conflict” cannot possibly be inferred from the rules themselves. Assuming that institutional rules and requirements cannot participate in a conflict, the things that any pair of indicative rules apply to only stand in a subsumption relation to each other relative to a specific *theory of brute reality*. This theory of brute reality may be an *ontology* of brute reality, in which case there is no problem with a translation of specificity to logical subsumption, but the involved situation recognition will in many cases also involve ampliative reasoning, abstraction, etc. The framework sketched in chapter 3 does not account for subsumption between autoepistemic rules, and it is not clear what such a subsumption theory would look like when applied to the problem of detecting irreconcilable conflicts between legal rules. An approach to default subsumption has been sketched by [274], but adopting this very broad approach would seem to generate false positives.

The legislator does not supply a theory of brute reality with his legal rules, and therefore cannot possibly foresee what combinations of normative rules will at some point in time be construed as being in “irreconcilable conflict”. The nightclub example in section 6.3 shows the problem. There is no logical justification for considering this a compliance conflict according to [172, 189, 266] or this book, unless one models the involved situation so that they meet the logical formulation of the conflict, which just hides the problem.

³Confer sections 5.1 and 6.6. Conflicts solved by *lex specialis* will generally not lead to problems since the overly general permission will be defeated.

The question is why one would take the trouble to reformulate the rules so that they appear to be in conflict. Is the compliance conflict just one logical principle, akin to the conflict of disaffirmation, or is it something else? The compliance conflict in a classical logical sense appears to be a very rare event, or even a non-event. But there is another soft principle at work which can also be described under the heading of compliance conflict:

Proposition 28 (Compliance Dilemma). In any decision problem, there must be a *known* option or set of options that is the allowed choice in that decision problem.

The selection of the appropriate menu of alternatives is of great importance. Given a set of specific options, the rules may impose a number of ordering constraints on them. Circularities may arise, but do not in themselves constitute a conflict. A problem arises if all conceivable alternatives are disallowed, or strictly worse than inconceivable, fictional alternatives.

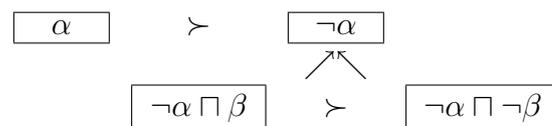
It is not good enough to claim – as a matter of ontology – that “there exists a better alternative that is allowed”. This better alternative must also be imaginable; There is a *burden of proof* on the prosecutor to supply such an option. If it isn’t, courts may latch on to a conspicuous circularity and use the principles that they normally use to resolve normative conflicts to construct a theory that allows them to conclude that no violation has taken place. This theory construction in the simple case involves *strengthening the antecedent* (cf. section 6.5.1) of the more generally applicable rule involved in the circularity.

It however makes a difference which decision problem we are considering. If proposition α is true in all imaginable alternatives, α is part of the *settled context* in that decision problem. To make a case that α is a violation one has to reconceptualize the decision problem, for instance by placing it back in time, when α was still preventable. Obligations that only apply in situations where the settled context already includes a violation are called *contrary-to-duty obligations*.

6.4. Contrary to Duty Obligations

The *contrary-to-duty* (*CTD*) obligation arises in a sub-ideal situation, brought about by the violation of a primary obligation. This is not allowed in standard deontic logic because it is based on a modal distinction between the actual world and the ideal one, and treats normative sentences as constraints on the ideal world. Treatment of obligations as a proposed betterness ordering of types of situations orders these worlds.

The relation between the primary obligation and *CTD-obligation* can be represented according to the following schema:



The CTD-obligation in the bottom row only distinguishes cases in which $\neg\alpha$ is already the case, in both allowed and disallowed situations.

It is very important to realize that the relation between the primary obligation and the CTD-obligation is not one of compliance or disaffirmation conflict. As long as you don't get into the sub-ideal situation, the CTD-obligation is silent – neither complied with nor violated. The CTD-obligation also does not imply a permission to enter into the sub-ideal situation.

The Chisholm paradox (viz. [77]) is an instructive example of how to use subjunctive betterness statements to order actions and to analyze complex sets of logical relations between primary obligations and CTD-obligations. The Chisholm set consists of the following norms:

1. $O(\alpha|\top)$
2. $O(\beta|\alpha)$
3. $O(\neg\beta|\neg\alpha)$

The logical paradox arises in standard deontic logic when α is the case. α reads “a man goes to the assistance of his neighbours”, and β reads “the man tells his neighbours that he will come”. Opinions are divided (viz. [265]) over whether this paradox already exists in natural language, or only in the logics that give rise to it. If one reads norms as statements of subjunctive betterness, one takes the position that the paradox merely disqualifies certain logical representations of the Chisholm situation.

The Chisholm set translates to the following set of preferences:

1. $|\alpha| \succ |\neg\alpha|$
2. $|\beta \sqcap \alpha| \succ |\neg\beta \sqcap \alpha|$
3. $|\neg\beta \sqcap \neg\alpha| \succ |\beta \sqcap \neg\alpha|$

The first preference expresses a preference for both $|\beta \sqcap \alpha|$ and $|\neg\beta \sqcap \alpha|$ over $|\beta \sqcap \neg\alpha|$ and $|\neg\beta \sqcap \neg\alpha|$. As for instance observed in (cf. [269]) a good formalization of ordering alternatives enforces the following ordering constraints:

$$\boxed{\alpha \sqcap \beta} \succ \boxed{\alpha \sqcap \neg\beta} \succ \boxed{\neg\alpha \sqcap \neg\beta} \succ \boxed{\neg\alpha \sqcap \beta}$$

An obligation is not cancelled by adding a second obligation telling us what to do if the first one is violated. If you tell your neighbours that you will come to their assistance, and don't do it, then you have violated two obligations. As a general rule, CTD-obligations do not require a choice between the involved norms. Straightforward deontic logics based on an accessibility relation interpreted as *ideality* often have a problem with this construct: such logics do not order the alternatives, but pick the ideal or, in some logics, minimally subideal one.

Only because the Chisholm situation involves a third norm ordering the two situations not ordered by the first two, we are able to construct a complete ordering. This is not generally the case with CTD-obligations.

Time adds another interesting twist to the Chisholm situation. Note that the transitions that can be effected by an agent's actions are only those from $|\neg\alpha \sqcap \neg\beta|$ to either $|\alpha \sqcap \neg\beta|$ or $|\beta \sqcap \neg\alpha|$, and from $|\beta \sqcap \neg\alpha|$ to $|\alpha \sqcap \beta|$. The choice for β can only be realized *before* the choice for α .

This makes a temporal reading, where norms only become visible at certain time points and norm 2 only becomes *visible* because norm 1 was violated, impossible (cf. [269]). This shows that the Chisholm example, although a bit contrived, has been chosen very well. A more straightforward and intuitive temporal example, by Prakken and Sergot in [232], admonishes one to keep one's promises, and to apologize if one doesn't.

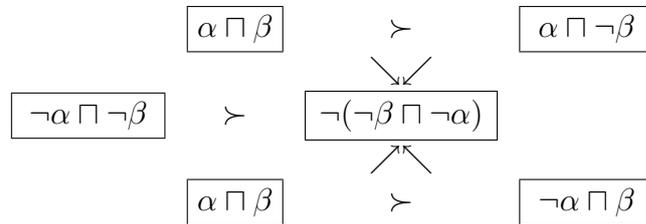
The Reykjavik set (cf. [78]) requires a partial ordering of the involved types of action. The Reykjavik set consists of the following norms:

1. $O(\neg\alpha \sqcap \neg\beta | \top)$
2. $O(\beta | \alpha)$
3. $O(\alpha | \beta)$

α reads "tell Reagan the secret", and β reads "tell Gorbachov the secret". This one is slightly more complicated to translate to preferences. The Reykjavik set translates to the following set of preferences:

1. $|\neg\alpha \sqcap \neg\beta| \succ |\neg(\neg\beta \sqcap \neg\alpha)|$
2. $|\beta \sqcap \alpha| \succ |\neg\beta \sqcap \alpha|$
3. $|\beta \sqcap \alpha| \succ |\beta \sqcap \neg\alpha|$

This shows us it is possible that there are more than one CTD-obligations pertaining to one primary obligation. The relation between this primary obligation and the two CTD-obligations can be represented in the following schema:



If CTD-sets get even more complex, it is obviously no longer possible to represent them graphically in this way.

The concept $|\beta \sqcap \alpha|$ is less cumbersome than $|\neg(\neg\beta \sqcap \neg\alpha)|$, but for the sake of making clear where it comes from the statement is not rewritten. The first preference clearly expresses a preference for $|\neg\beta \sqcap \neg\alpha|$ over $|\beta \sqcap \alpha|$, $|\neg\beta \sqcap \alpha|$, and $|\beta \sqcap \neg\alpha|$. The imposed ordering is a partial one:

$$\boxed{\neg\alpha \sqcap \neg\beta} \succ \boxed{\alpha \sqcap \beta} \succ \boxed{\neg\alpha \sqcap \beta} \quad \boxed{\alpha \sqcap \neg\beta}$$

The last two situations are left to personal preference for either Reagan or Gorbachov. Note that the transitions that can be effected by an agent are those from $|\neg\alpha \sqcap \neg\beta|$ to either $|\alpha \sqcap \neg\beta|$ or $|\beta \sqcap \neg\alpha|$, and from those to $|\alpha \sqcap \beta|$. There is no ordering constraint on choices between α and β .

Although temporal readings of CTD structures are ruled out, one might still speculate that the primary obligation and the contrary-to-duty obligation involve a substantially different decision point. In all examples given the decision points can in

principle be separated. The decision to tell or apologize is substantially different from the decision to come or break a promise. This intuition may be dispelled by Forrester's situation (viz. [111]), in which one is admonished to kill gently, if one kills. Killing gently is directly subsumed by killing, so it is not possible to kill gently while not killing. α and β are not at all independently realizable in this case.

The CTD situation should not be considered merely theoretical. Real legislation does actually create many situations reminiscent of the Chisholm and Forrester situation. Contrary-to-duty imperatives are quite common, although they are usually far more complex to analyze in legislation. All norms regulating punishment by the legal system, and contractual remedial and reparational obligations and liabilities arising from contract violations follow this general pattern.

The CTD situation is merely a recognition in legal theory of the fact that betterness relations are transitive. Since classical deontic logics are based on a distinction between what is the case and what would be *ideally* the case, the CTD situation forms a problem. Many modern deontic logics have addressed this issue; Some do so by explicitly modeling a preference relation.

6.4.1. Duty Conditional on Violation

There is a very different way of handling (some kinds of) CTD-obligations that should be mentioned. Instead of viewing a CTD-obligation as an obligation arising from the sub-ideal situation in which one norm has already been violated, one could see the CTD-obligation arising from the violation itself. Consider the following possible formulations of norms about apologizing when breaking promises in natural language:

1. One ought not to break a promise.
2. One may break a promise if one apologizes.
3. One ought to break promises and apologize.
4. If one breaks a promise, then one ought to apologize.
5. If item 1 is violated, then one ought to apologize.

Item 2 is clearly a permission that is an disaffirmation of item 1. Item 3 creates a compliance conflict with item 1. Item 4 is a CTD-obligation and at first glance expresses the same thing as item 5, but still works somewhat differently in a legal system.

Item 4 is only active as a CTD-obligation if item 1 exists. If the text of item 1 changes, then the preference expressed in item 4 changes as well. Governatori et al. (cf. [123]) solve this by the introduction of an operator $O(\alpha) \otimes O(\beta)$, which is read as " $O(\beta)$ is the reparation of the violation of $O(\alpha)$ ". This makes the obligation $O(\beta)$ inapplicable to any situations in which the opposite of α is the case, but $O(\alpha)$ is nevertheless not violated. The operator is therefore not suitable for replacement of a more general way of dealing with CTD-obligations. Governatori et al. may however have a valid argument that some CTD-obligations do actually require the violation. Arguably this is the case in the following set:

1. One ought not to break a promise.
2. One may break a promise if one pays a \$100 fee to the government.

3. If item 1 is violated, then one ought to apologize.

The question is whether item 1 is still violated if one pays the fee. Does one still need to apologize? We have a problem if that is not the case: we need to distinguish between CTD-obligations that are only applicable if the primary obligation is, and CTD-obligations that are applicable regardless of the primary obligation. This is a case of confluence of a CTD-relation on the logical level, and an applicability condition, as described in section 5.2.3.

Apparently the primary obligation does not need to be applicable if item 3 would have been formulated as “If the conditions of item 1 are met, then one ought to apologize.”, but there is no guarantee that it will be read this way.

This type of *duty conditional on violation* is for instance quite frequently encountered in contracts.

That the CTD-like duty conditional on violation n_3 only applies to violation of some other rule n_1 can be represented as:

$$\{n_3\} \sqsubseteq \forall \text{appliesTo}. (\exists \text{disallowedBy}. \{n_1\})$$

This necessary condition on n_3 simply that n_3 only applies to violations of n_1 . Contrary to the previous phenomena this pattern does not appeal to a perceived deontic choice effect.

6.5. Betterness in OWL DL

The contrary-to-duty situation is clearly a normal feature of legislation, and more importantly also a normal feature of decision problems, which indicates that \prec and \preceq are transitive. One would expect this of such relations. It presents a problem for simple modal deontic logics, which equate violation with inconsistency, and cannot deal with situations involving violations. Inability to deal with contrary-to-duty situations simply disqualifies certain logics.

The normative conflict is a different issue. It is a fundamental problem with the applicability of the obligation terminology on decision problems. In the presence of normative conflict, one can no longer be sure that one can strengthen the antecedent β of weaken the consequent α of an obligation $O(\alpha \mid \beta)$ (cf. for instance [268] for a discussion). There are of course many different ways to complicate deontic logic to accommodate normative conflict, but it is simply the concept of obligation, as understood in deontic logic, which no longer fits well on the decision problem. Interestingly, since permission is *intended* to cancel obligations, which it doesn't do in standard deontic logic, we can also conclude that the way standard deontic logic models permission makes little sense, as has been observed by many before.

Trying to model a simple deontic logic in OWL DL is therefore not very helpful. Even for a simple deontic logic OWL DL would moreover create problems, since OWL DL lacks the ability to enforce reflexivity on relations.

There are generally two routes one can take to address these issues:

1. Betterness (or preference) instead of ideality is a better fit for contrary-to-duty situations, and moreover connects better to decision problems like planning;

2. Defeasibility is an obvious explanation of the relations between obligation and permission, and of compliance conflict.

In this book I recommend both, but in different contexts. From an institutional perspective, normative rules are clearly defeasible rules as already indicated. Normative conflict is one of the reasons why normative rules are considered defeasible, dependent on an equally defeasible assumption about the pair of normative rules. Section 6.6.3 will discuss applicability rules that make the applicability of one normative rule dependent on the applicability of another one.

Obligation, just like preference, in first instance relates to decision problems. It is an account of choice between alternative plans. It is for this purpose that I introduce a property \succeq to model a relation between plans or actions. The notion of an *actual situation*, as distinguished with ideal ones, will however play no role. *LKIF* (see appendix to [35]) includes the same property.

Each normative rule posits its own unique betterness relationship, which can thought of as a norm in the proper contexts: a property \succeq_n is inferred from $\{n\}$. There is an expectation that these relationships can be combined into a single one for sets of rules in some way, resulting in an integrated theory of what the legislator wants us to do. The expectation in a standard deontic logic is very strong: a single \succeq can model each set of norms, and is besides that antitransitive. I do not attempt in this section to give a systematic account of these expectations.

Proposition 29. For each legal rule n interpreted as an obligation $O_n(\alpha \mid \beta)$ or prohibition $F_n(\neg\alpha \mid \beta)$, create the following terminological axioms:

$$\begin{aligned} \exists \text{allowedBy}.\{n\} &\sqsubseteq \neg \exists \preceq_n . \exists \text{disallowedBy}.\{n\} \\ \exists \text{disallowedBy}.\{n\} &\sqsubseteq \exists \preceq_n . \exists \text{allowedBy}.\{n\} \end{aligned}$$

Proposition 30. For each legal rule n interpreted as an permission $P_n(\alpha \mid \beta)$, create the following terminological axiom:

$$\exists \text{applicable}.\{n\} \sqcap \neg \exists \text{allowedBy}.\{n\} \sqsubseteq \exists \preceq_n . \exists \text{allowedBy}.\{n\}$$

Because we represent each normative rule by both an indicator and a requirement, $\exists \text{allowedBy}.\{n\} \equiv \alpha \sqcap \beta$ and $\exists \text{disallowedBy}.\{n\} \equiv \neg\alpha \sqcap \beta$ are “almost” true, barring applicability conditions that defeat the indicative rules. For discussion purposes we will assume that applicability of n already settled. Note however that continuous application of indicative rules will be necessary.

$O_n(\alpha \mid \beta)$ – barring applicability restrictions on n – therefore means:

$$\begin{aligned} \alpha \sqcap \beta &\sqsubseteq \forall \preceq_n \alpha \sqcup \neg\beta \\ \neg\alpha \sqcap \beta &\sqsubseteq \exists \preceq_n \alpha \sqcap \beta \end{aligned}$$

If the obligation is for instance *when driving a vehicle one should keep to the right of the road*, the interpretation in terms of subjunctive betterness is:

When driving a vehicle and keeping to the right of the road, all equal or better alternatives would have involved not driving a vehicle or keeping to the right of the road. When driving a vehicle and not keeping to the right of the road, an equal or better alternative would be driving a vehicle and keeping to the right of the road.

The combination of both statements is only satisfied by considering alternatives involving not keeping to the right while driving strictly worse than those that do involve keeping to the right, or by not considering alternatives which involve driving a vehicle at all.

$P_n(\alpha \mid \beta)$ means – barring applicability restrictions on n :

$$\neg\alpha \sqcap \beta \sqsubseteq \exists \preceq_n \alpha \sqcap \beta$$

If the permission is for instance *In an emergency situation, drivers of an emergency vehicle are allowed to race a red light if circumstances permit*, the interpretation in terms of subjunctive betterness, involving some interpretation to separate the condition from the thing permitted, is:

When an emergency vehicle stops for a red light while there is an emergency situation and circumstances permit racing the red light, an equal or better alternative is racing the red light.

As pointed out in section 4.6, there is a difference between permissions that offer a free choice between two alternatives, and those that remain silent on the other option. If the phrasing of the text does not suggest such a freedom of choice, one should go for this interpretation. Indifference of the legislator simply involves a permission in both directions, that is adding:

When there is an emergency situation and circumstances permit racing the red light and an emergency vehicle races the red light, an equal or better alternative is that the emergency vehicle stops for the red light.

The legislator may not intend this.

In order to make a number of formal observations about the family of \succeq properties I present it here as if it were a self-contained labeled modal logic. In actuality we are however dealing with simple transitive OWL DL properties. The property \prec and $=$ are pseudo-properties that do not play an actual role in OWL DL.

As pointed out in section 3.4, OWL DL can be understood in terms of modal frames. Let M be a modal frame of the form $M = \langle W, N, \{\preceq_n\}_{n \in N}, V \rangle$ where W is a set of worlds, N a set of preference theories, the \preceq_n are transitive relations, and V is a propositional valuation. Each $n \in N$ is a (consistent) subtheory in the larger preference theory, and can be intuitively understood as a kind of (institutional) agent that imposes its own preferences on worlds $w \in W$. In a description logic context the worlds should be understood as the individuals in a model (cf. the examples in section 3.4.2), and specifically as actions or the alternative plans in a menu.

The preference theory represents the *norm* implied by one normative rule, or if preferred by a set of normative rules provided they are consistent with each other. A set of preference theories describes a normative order. The notion of *preference* and *preference theory* as used in this section is strictly used for compatibility with relevant preference theory literature (generally [138, 241, 267, 100, 102, 101]): what we are

dealing with are subjunctive propositions about *betterness*, made by a legislator, that are intended to play a role as motivating preferences in decisions.

As for instance Sagoff explains (cf. [241]) preferences are generally attributed to *agent* roles (see i.a. section 4.5.2); Human beings cannot be attributed a single comprehensive agent role, and the set of preferences ascribed to them in order to explain their behaviour need not be consistent. The concept of attributing mental attitudes to normative systems as if they were agents has been used before in computer science & Law (cf. for instance [28]). The observation Sagoff makes applies here: certain sets of normative rules may describe a coherent normative order as a preference theory, but the totality of rules of an institution does not have to, and often doesn't.

Hansson in [138] gives a state-of-the-art account of obligation based on preference theories. This is not such an account; The normative rule only imposes constraints on plans when alternatives are considered that are allowed or disallowed by the normative rule. The proposal given here does not allow for evaluation of normative positions in the way that a deontic logic does, but it may be combined with such an account.

Read $x \preceq_n y$ as “preference theory n considers y as least as good as x ”. Read $x \prec_n y$ as “preference theory n considers y better than x ”. This is shorthand for $(x \preceq_n y) \sqcap \neg(y \preceq_n x)$. Read $x =_i y$ as “preference theory n is (explicitly) indifferent between y and x ”. This is shorthand for $(x \preceq_n y) \sqcap (y \preceq_n x)$. In this book the label n is omitted in some cases.

The language L is a set of terminological axioms, implicitly joined by a \sqcap , in a description logic *TBox* (terminological box) of the form $\phi_1 \sqsubseteq \phi_2$ where ϕ_1 and ϕ_2 are concepts conforming to the syntax in 3.4. As usual $\forall \preceq$ is defined in terms of $\exists \preceq$ as follows: $\forall \preceq \phi \equiv \neg \exists \preceq \neg \phi$. The truth definition for this language, and semantic notions like frame, satisfiability, and validity, are standard (see [24]):

$$M, w \models \exists \preceq_n \phi \text{ iff } \exists w' : w \preceq_n w' \text{ and } M, w' \models \phi$$

This says that ϕ is true in at least one alternative of w which n considers at least as good as w . The preference modality $\exists \preceq_n$ constrains a preference order at the level of worlds, for a preference theory n .

Note that the frame axiom for reflexivity, which would allow me to turn \preceq into a reflexive relation, is absent: OWL DL does not allow for reflexive frames. Addition of this axiom would permit use of the *S4* system used by i.a. Van Benthem et al. as a preference logic (cf. [267]), and as a building block for more complex deontic logics like for instance Boutilier's *CT4O* (cf. [56]) or Van der Torre's *2DL* (cf. [268]).

Reflexivity is a common property of mereological, topological, spatial, and temporal relations. Some workarounds have been proposed in i.a. [125] (submissions 26 and 33), and the OWL specifications propose some alternative methods for approximating reflexivity in the definition of parts and wholes. The OWL 2 proposal (cf. section 3.4) does add reflexivity, but is not the standard yet. Acceptation of OWL2 would be reason to reconsider this formalization.

One of the arguments made for the original omission of reflexive frames, besides the apparent computational attractiveness of doing so, is that its primary function is as an abbreviation anyway: instead of stating that for some w there is some w' such that wRw' that has some property we are looking for, we have to state that w has the

property we are looking for or there is some w' such that wRw' that has the property we are looking for. In [234] Rescher made the point that the property of reflexivity is in most use cases actually ontologically suspect: nothing is a “part of itself” and well-formed wholes should consist of at least two parts. In a mereological, topological, spatial, or temporal context this observation is not of great importance. If identity is not regarded as a limit case of parthood this does not appear to invalidate important results.

Rescher’s point clearly also holds for the \preceq relation: it is odd to say that something is as good as itself, or that we are indifferent between something and itself. Since we are mostly concerned with the irreflexive counterpart of \preceq the limitations of OWL DL can be accepted.

It is unfortunate that this limitation of OWL DL necessitates re-evaluation of existing formal results before they can be incorporated into the Semantic Web.

As stated, the obligation $O_n(\alpha \mid \beta)$, or a prohibition $F_n(\neg\alpha \mid \beta)$ translates to:

$$\begin{aligned}\beta \sqcap \alpha &\sqsubseteq \neg\exists \preceq_n \beta \sqcap \neg\alpha \\ \beta \sqcap \neg\alpha &\sqsubseteq \exists \preceq_n \beta \sqcap \alpha\end{aligned}$$

This says that in a world where $(\beta \sqcap \alpha)$ is true, there is no world equal or better accessible where $(\beta \sqcap \neg\alpha)$ is true, and in a world where $(\beta \sqcap \neg\alpha)$ is true, there is some world equal or better accessible where $(\beta \sqcap \alpha)$ is true, according to i . It is possible to include other worlds where $(\beta \sqcap \neg\alpha)$ is true in the ordering as long as one does not claim they are better than a world in which $(\beta \sqcap \alpha)$ is true, and obviously a world where $\neg\beta$ is true can be placed anywhere in the ordering.

Because the peculiar frame-based syntax of OWL will produce a triad C, C', C'' in a taxonomy, as discussed in section 6.1.1, there will often be good reason to simply state the subsumption triples $C' \sqsubseteq C, C'' \sqsubseteq C$, and a disjointness triple $\neg C' \equiv C''$, in addition to the $C' \sqsubseteq \neg\exists \preceq_n C''$, $C'' \sqsubseteq \exists \preceq_n C'$. In other words, α as differentiating characteristic, between C' and C'' can be left implicit.

The permission $P_n(\alpha \mid \beta)$ translates to:

$$\beta \sqcap \neg\alpha \sqsubseteq \exists \preceq_n \beta \sqcap \alpha$$

This says that in a world where $(\beta \sqcap \neg\alpha)$ is true, there is some world equal or better accessible where $(\beta \sqcap \alpha)$ is true. Intuitively this may suggest that it is advisable to choose $(\beta \sqcap \alpha)$ over $(\beta \sqcap \neg\alpha)$, but that is not intended. It is obviously consistent to add an edge from a world in which $(\beta \sqcap \alpha)$ is true to a world where $(\beta \sqcap \neg\alpha)$ is true, establishing that both alternatives are of equal value as proposed in section 4.6.

Note that the characterization of permission is just to alert us to any *conflicts of disaffirmation* between obligations and permissions. Trying to strengthen the definition of permission within the expressive limits of OWL DL (for instance $(\beta \sqcap \neg\alpha) \sqsubseteq (\exists \preceq_n (\beta \sqcap \alpha) \sqcap \forall \preceq (\alpha \sqcup \neg\beta))$) is not going to solve the problem that we are dealing with an underspecified frame that includes unintended orderings.

The relation \preceq is not complete. Some worlds are genuinely incomparable, and they must be if \preceq is to be a reasonable interpretation of permission. This means that the \preceq

relation does not meet the property of *trichotomy*, requiring that it should be possible to choose between any two alternatives, as is often required of preference relations. Hansson defends the general invalidity of the trichotomy property in [138], but this argument is hardly universally accepted. The relation covered here therefore does not match with the mainstream conceptions of *revealed* preferences as explained in section 2.4.1. As pointed out, limitations of OWL offered a practical reason to reject trichotomy.

Proposition 31 (No trichotomy). Some worlds are incomparable.

Proof. Assume the obligations $O_n(\alpha \mid \top) \sqcap O_n(\beta \mid \top)$. $\mid \phi \mid$ is the set of worlds w such that $M, w \models \phi$. Assume a $x \in \mid \neg\alpha \sqcap \beta \mid$ and a $y \in \mid \alpha \sqcap \neg\beta \mid$. Observe that both $x \preceq_n y$ and $y \preceq_n x$ are inconsistent. \square

Hansson tried to explain deontic logic in terms of *ceteris paribus preferences* [102, 100, 101], i.e. *all else being equal* preferences: these also do not mesh well with revealed preference. Subjunctive betterness as encountered in formal constitutive rules, which are obviously formulated prior to the phenomena they describe as pointed out in section 4.1, has little to do with revealed preferences. The notion that preferences can be revealed is based on the assumption that choices between alternatives are observable, which is highly suspect, as pointed out in section 2.4.1.

To be viable as a deontic knowledge representation system, the system has to meet some requirements. Minimally, the system should capture two properties: that what is obliged should be permitted and the impossible should not be obligatory.

The intuition for the first property is that you cannot be obliged to do something without at the same time being permitted to do that something. The property is usually expressed by way of the following axiom: $O\phi \rightarrow P\phi$. This characterization is not defined in terms of the frame axiom $\forall \preceq \phi \rightarrow \exists \preceq \phi$, but it does enforce it:

Proposition 32. What is obligatory is permitted. The axiom $O_n(\alpha \mid \beta) \rightarrow P_n(\alpha \mid \beta)$ is true.

Proof. Trivial from $((\beta \sqcap \alpha) \sqsubseteq \neg\exists \preceq_n (\beta \sqcap \neg\alpha)) \sqcap ((\beta \sqcap \neg\alpha) \sqsubseteq \exists \preceq_n (\beta \sqcap \alpha))$ and $(\beta \sqcap \neg\alpha) \sqsubseteq \exists \preceq_n (\beta \sqcap \alpha)$. \square

Note that it is perfectly possible to weaken the obligation so that it does not meet this property (to appease the critics mentioned in 6.3.1). The obligation translates to two statements in this system, connected here by a \sqcap . Omitting the preference proposition $((\beta \sqcap \neg\alpha) \sqsubseteq \exists \preceq_n (\beta \sqcap \alpha))$ will do the trick. This is however not a feature I would want to include.

A nice feature of this system is that it does not only make the impossible obligation $O_n(\neg\alpha \mid \alpha)$ inconsistent, but will also do something intuitively meaningful with the unintuitive and meaningless obligation $O_n(\alpha \mid \alpha)$ mentioned by [195].

Proposition 33. The impossible is not obligatory: $\neg O_n(\neg\alpha \mid \alpha)$ is true.

Proof. Observe that \perp follows from $((\neg\alpha \sqcap \alpha) \sqsubseteq \neg\exists \preceq_n (\neg\alpha \sqcap \neg\alpha)) \sqcap ((\neg\alpha \sqcap \neg\alpha) \sqsubseteq \exists \preceq_n (\neg\alpha \sqcap \alpha))$. \square

Proposition 34. The meaningless obligation $\neg O_n(\alpha \mid \alpha)$ gives rise to an incoherent concept $(\alpha \sqcap \neg\alpha)$ which evaluates to the empty set represented by concept OWL:NOTHING. Concepts such as this will generally be marked as ill-formed by an OWL aware editor like Protégé.

The propositions so far do not require a deeper understanding of OWL DL subsumption relations between propositions. Contrary-to-duty situations and normative conflict are also classical tests of the viability of a knowledge representation system for normative rules. In their most straightforward formulation these tests are also passed. In section 6.5.1 we will however see that when we take OWL DL subsumption into account this formalization shows peculiar behaviour, when compared to deontic logics.

The contrary-to-duty situation discussed in section 6.4 present no problem, as is shown with the Chisholm paradox.

Proposition 35 (Chisholm). The sentences $O_n(\alpha \mid \top)$, $O_n(\beta \mid \alpha)$, $O_n(\neg\beta \mid \neg\alpha)$ are only satisfied by the ordering identified by [268].

Proof. If $\mid \phi \mid$ is the set of worlds w such that $M, w \models \phi$, then the desirable ordering is represented as $\mid \neg\alpha \sqcap \beta \prec_n \neg\alpha \sqcap \neg\beta \prec_n \alpha \sqcap \neg\beta \prec_n \alpha \sqcap \beta \mid$. Verify for each set $\mid \phi \mid$ that:

1. that it cannot be empty if some other set $\mid \phi' \mid$ in the ordering is nonempty and $\mid \phi' \mid \prec_n \mid \phi \mid$;
2. that $x \preceq_n y$ is not possible for any $x \in \mid \phi \mid$ and $y \in \mid \phi' \mid$ and $\mid \phi' \mid \prec_n \mid \phi \mid$.

□

The essence of the Chisholm paradox is of course not that to comply with the obligations you have to choose a world $w \in \mid \alpha \sqcap \beta \mid$. The knowledge representation system should be able to establish that if only a settled context $\mid \neg\alpha \mid$ is given, i.e. all considered alternative worlds $w \in \mid \neg\alpha \mid$, then the optimal choice is a $w \in \mid \neg\alpha \sqcap \neg\beta \mid$.

As mentioned before, the conflicts of disaffirmation and compliance described in detail in section 6.3 do raise a contradiction, as they arguably should. As argued before it is important to be able to calculate all prima facie obligations, even if we apply a method to resolve them automatically. The axiom that captures this notion for conflicts of compliance is generally called the *no-dilemma assumption*.

Proposition 36 (No dilemma). There are no conflicting obligations. The obligations $O_n(\alpha \mid \beta)$ and $O_n(\neg\alpha \mid \beta)$ are inconsistent: $\neg(O_n(\alpha \mid \beta) \sqcap O_n(\neg\alpha \mid \beta))$ is true.

Proof. If $M, w \models \beta \sqcap \alpha$ then $M, w \models \exists \preceq_n (\beta \sqcap \neg\alpha) \sqcap \neg \exists \preceq_n (\beta \sqcap \neg\alpha)$ if both statements are true. □

The conflict between a permission and an obligation follows a very similar pattern. As stated earlier the obligation consists of two separate statements. One captures the permission for a choice that it entails as $(x \preceq_n y)$, and the other, negative, statement can be thought of as blocking a permission for a competing choice $(\neg(y \preceq_n x))$.

Proposition 37 (Disaffirmation). The obligation $O_n(\alpha \mid \beta)$ and permission $P_n(\neg\alpha \mid \beta)$ are inconsistent: $\neg(O_n(\alpha \mid \beta) \sqcap P_n(\neg\alpha \mid \beta))$ is true.

Proof. If $M, w \models \beta \sqcap \alpha$ then $M, w \models \exists \preceq_n (\beta \sqcap \neg\alpha) \sqcap \neg \exists \preceq_n (\beta \sqcap \neg\alpha)$ if both statements are true. \square

How to deal with normative conflicts that *do* occur is the subject of the following subsection. For now I will end this presentation with some remarks.

The modal notion of bisimulation makes some sense in this interpretation and can be used to establish that two preference structures are the same from a modal standpoint. This notion is central to the semantics of comparison of preference structures but we will not develop this line of possible research here.

It may be useful in comparing preference structures to include ‘silent’ preferences in analogy to silent transitions in weak bisimulation of state transition systems. As will be pointed out in section 6.7.3, comparison as intended in comparative law involves three systems: the systems M_1 and M_2 to be compared, and the norm(s) of analysis used to decide whether they are sufficiently ‘the same’. It is possible to check whether M_1 is a weak bisimulation of M_2 , but also which one is the “best” simulation of a third set of norms M_3 .

Decidability in LKBS use contexts – if we leave the application of defeasible rules out of consideration – is taken for granted, as it is represented in OWL DL (cf. [158]). OWL DL has been used in practice on very large knowledge bases (viz. [193]), although these exhibit other structural properties than legal knowledge bases built with this method. It is known that certain mereological and temporal constructs do not scale very well, and that is also the case for the betterness relation central to the conceptualization of norms proposed here: it is very taxing on OWL reasoner implementations. The question of decidability of this representation *including* the defeasible rules is ill-defined, since the logic of applying these was intentionally left underspecified.

6.5.1. Betterness and Normative Conflict

In legal theory there is a lively interest in the question what pairs of normative rules – assuming we would a priori expect them to form a coherent normative order – can be in normative conflict. Conflict plays a role on two different levels: there is the true normative conflict on logical grounds, and there are mundane cases where a pair of normative rules presents us with a planning dilemma. Planning dilemmas are not normative conflicts.

Section 6.3 addressed the notion of normative conflict as a relation between two normative rules that are *necessarily* in normative conflict on terminological grounds. In addition, legal practice recognizes conflicts that only occur in certain categories of cases, as section 6.3.1 observed. This latter category can, as a consequence of the observed ontological commitments for the representation of normative rules, not be considered candidate reusable knowledge about sources of law, since the construction of the conflict involves ampliative reasoning of an unspecified nature.

In principle the existence of normative conflict must be affirmed by another source of law to be of direct relevance here. This would be generally case law.

The betterness interpretation of normative rules does provide some useful tools for the detection and resolution of conflicts, which will be discussed in this section.

The first such tool is the epistemic interpretation of what violation is, as an alternative to the straightforward derivation of the institutional qualification.

According to the betterness reading of normative rules, if you choose an alternative even though another alternative was available that is strictly better in n , then you have violated n . This can be interpreted in two ways, depending on whether we use a straightforward model theoretic interpretation of OWL DL or the abox of the OWL DL reasoner as the epistemic model of a theory (as explained in section 3.4.2):

Terminological: the formalization of the normative rule n entails – as a matter of ontology – that there exists a better or equal alternative, and asserting that the alternative a under consideration is (also) better or equal to that alternative (i.e. that they are of equal value) would give rise to inconsistency; or

Epistemic: the abox names some other individual alternative a' , and it is not possible to consistently assert $(a' \preceq_n a)$ (or $\preceq_n(a', a)$ in abox notation).

The terminological option has no added value over the institutional interpretation of normative rules. It is in fact only more complicated than merely establishing the proposition $\exists \text{disallowedBy}.\{n\}$.

The epistemic alternative is important for a variety of reasons. When planning, the comparison between alternatives is generally made between those explicitly considered. The best plan, among those conceivable, is one of which it is possible to assert that it is better or equal than all alternative plans. In addition, in planning we typically deal with settled contexts.

The restriction to a *settled context* relates to the observation in section 6.3.1: we only have to order the alternatives that are known to be feasible. In practical planning there are two reasons to consider certain propositions part of the settled context:

- they are already true in our interpretation of the situation and cannot – to one's knowledge – be undone anyway, or they are inevitably going to become true according to our interpretation of the situation; or
- one already settled on violation of the law, but plans to do so without any other violations.

The violation without a settled context can be understood as a violation in settled context \top .

Proposition 38 (Violation). $u \preceq w$ is a violation of preference theory n in settled context α if and only if $w \prec_n u$ and $M, w \models \alpha$ and $M, u \models \alpha$.

The object of using the preference theory for planning is to order alternatives in such a way that the best alternative remaining does not violate rules, or no additional ones beyond the ones already violated in the settled context.

An important feature of this formalization is that it only takes into account *orderings of alternatives* that are against the rules. It is silent on decision problems where

all available alternatives are against the rules, since these do not involve preferring an alternative that is allowed over one that is disallowed. This distinction however only becomes relevant for planning if an epistemic interpretation is chosen.

In some cases a preference theory based on normative rules may present the planning agent with a planning dilemma.

Proposition 39 (Planning Dilemma). Both $u \preceq w$ and $w \preceq u$ are violations of the relevant preference theory in the chosen settled context.

The planning dilemma is *not* a normative conflict, or significant evidence of it. The planning agent has to choose between violating one of two normative rules. A normative conflict is only deemed to exist if one of the normative rules is not applicable due to the fact that the other also applies.

There is in principle no logical problem with substantially the same behaviour being allowed by some normative rule n_1 and disallowed by n_2 . This inconsistency could only arise if one makes the following assumption about a *set* of normative rules N even though N contains conflicting rules such as the pair $\{n_1, n_2\}$: if behaviour b is disallowed according to some $n \in N$ then behaviour b is also disallowed by N , and if behaviour b is allowed according to some $n \in N$ then behaviour b is also allowed by N .

Some *collections* of normative rules may be expected to result in a coherent account of normative order in terms of a preference theory partially ordering alternatives. If this account is inconsistent, it exposes normative conflicts. If the explanation of a normative order as a single set of equally important propositions fails for the analyzed collection, then an alternative conceptualization of the collection must be found. The construction of alternative conceptualizations of coherent normative order is however beyond the scope of this section: it is addressed in the next one. What is of interest here is how one knows that there is a normative conflict in a set of rules.

The strategy to be used is composing a single compound normative rule out of the set of rules, and investigating the resulting preference theory.

Proposition 40 (Compound Normative Rule). Let n be any normative rule, and n_c be a compound normative rule. For any $n \in n_c$, if behaviour b is disallowed by n then behaviour b is also disallowed by n_c , and if behaviour b is allowed by n then behaviour b is also allowed by n_c . The following must be an axiom of any coherent compound normative rule n_c : $\{n_c\} \equiv \forall \text{allows}.\forall \text{disallowedBy}.\neg\{n_c\}$.

The membership criterium for the set of rules one is interested in is of course application-specific. Let us for instance assume that the normative rules represented by work w , which is a source of law, is an attempt to formalize a coherent normative order. Construct a compound normative rule n_w that combines the normative rules in w :

$$\exists \text{disallowedBy}.\exists \text{representedBy}.\exists \text{metalex:embodies}.\{w\} \sqsubseteq \exists \text{disallowedBy}.\{n_w\} \quad (6.1)$$

$$\exists \text{allowedBy}.\exists \text{representedBy}.\exists \text{metalex:embodies}.\{w\} \sqsubseteq \exists \text{allowedBy}.\{n_w\} \quad (6.2)$$

If one would want to devise a strategy to construct compound normative rules automatically, one would use indicative rules with a ($\text{free}.\exists\text{disallowedBy}.\{n\}$) condition.

If making these assertions, in addition to the coherency axiom, does not introduce inconsistency, then there is no normative conflict within the set.

Proposition 41 (Terminological normative conflict). If a compound normative rule contains terminological normative conflict, it will not be coherent.

Finding *pairs* of rules that pose normative conflict is a diagnostic problem, that can be solved with a systematic approach to composing hypothetical compound normative rules in an OWL DL description classifier. This is done offline, i.e. not in the production LKBS, since the results of this computation are a static knowledge source to the LKBS. The result of such a diagnosis can be stored in the form of choice rules, discussed in section 6.6.3. Choice rules constructed using such a method do not however reflect a source of law, and are of no direct interest to this book.

The terminological normative conflict criterium as considered here does not however cover all types of normative conflict discussed in section 6.3.

Logical inconsistencies are obviously always an artifact of the chosen formalization. Certain classes of deontic logics can be for instance be classified by whether they have axioms for *strengthening the antecedent* ($O(\alpha_1 \mid \beta_1)$ to for instance $O(\alpha_1 \mid \beta_1 \sqcap \beta_2)$) and *weakening the consequent* ($O(\alpha_1 \mid \beta_1)$ to for instance $O(\alpha_1 \sqcup \alpha_2 \mid \beta_1)$) (cf. for instance [268] for an overview of deontic logics categorized by these features). Depending on what strategy one chooses, one is more or less likely to find logical contradictions between normative propositions.

These features always attract criticism. Certainly in law, where the presence of normative conflict is a matter to be explained, the absence of normative conflict is often considered a feature of legal deontic logics. The normative conflict however only arises from the deontic interpretation of normative rules; Normative conflict can in that sense be considered an essential feature of deontic logic proper. Because the formalization given here is only a basic *ordering* one, which treats the antecedent of the obligation very differently from deontic logics, a transparent comparison with deontic logics that can be classified as having the features of *strengthening the antecedent* or *weakening the consequent* is not trivial. The behaviour of the antecedent is best illustrated by an extended example.

Example 13. Consider a simplistic example: Assume that $\text{MOWLAWN} \sqsubseteq \text{ONLAWN}$ (hence $\neg\text{ONLAWN} \sqsubseteq \neg\text{MOWLAWN}$) and $\text{TUESDAY} \sqsubseteq \text{WORKDAY}$ are true.

Consider the following disaffirmation conflict: It is disallowed to walk on the lawn ($F_n(\text{ONLAWN})$ or $O_n(\neg\text{ONLAWN})$), but it is allowed to mow it ($P_m(\text{MOWLAWN})$). Alternatively consider the compliance conflict version: It is disallowed to walk on the lawn ($F_n(\text{ONLAWN})$ or $O_n(\neg\text{ONLAWN})$), but it is *obliged* to mow it ($O_m(\text{MOWLAWN})$).

Both pairs of two normative rules become logically inconsistent in the chosen representation if we would consider \preceq_m and \preceq_n as being part of a more general betterness relation \preceq_{nc} :

$$\neg\text{ONLAWN} \sqsubseteq \neg\exists \preceq_n \text{ONLAWN} \quad (6.3)$$

$$\text{ONLAWN} \sqsubseteq \exists \preceq_n \neg\text{ONLAWN} \quad (6.4)$$

$$\text{MOWLAWN} \sqsubseteq \neg\exists \preceq_m \neg\text{MOWLAWN} \quad (6.5)$$

$$\neg\text{MOWLAWN} \sqsubseteq \exists \preceq_m \text{MOWLAWN} \quad (6.6)$$

Axiom pairs 6.3 and 6.6, *and* 6.4 and 6.5 are inconsistent with a theory of brute reality that recognizes that $\text{MOWLAWN} \sqsubseteq \text{ONLAWN}$ and $\neg\text{ONLAWN} \sqsubseteq \neg\text{MOWLAWN}$. Subsumption between consequents appears to do the expected job here, for both cases.

The addition of an antecedent (for instance $O_n(\neg\text{ONLAWN} \mid \text{WORKDAY})$) makes no difference as long as both normative rules have the *exact same* antecedent.

Pairs of the type $O_n(\neg\text{ONLAWN} \mid \text{WORKDAY})$ and $P_m(\text{ONLAWN} \mid \text{TUESDAY})$, where the antecedents subsume each other, but the consequent is exactly the same also make no difference. Subsumption between antecedents also does its job, and as expected results in a reversal

Pairs however cannot vary in both the antecedent *and* consequent. Consider the axioms from the example above, but with the condition WORKDAY or, more specific, TUESDAY added, resulting in $O_n(\neg\text{ONLAWN} \mid \text{WORKDAY})$ and $O_m(\text{MOWLAWN} \mid \text{TUESDAY})$:

$$\text{WORKDAY} \sqcap \neg\text{ONLAWN} \sqsubseteq \neg\exists \preceq_n \text{WORKDAY} \sqcap \text{ONLAWN} \quad (6.7)$$

$$\text{WORKDAY} \sqcap \text{ONLAWN} \sqsubseteq \exists \preceq_n \text{WORKDAY} \sqcap \neg\text{ONLAWN} \quad (6.8)$$

$$\text{TUESDAY} \sqcap \text{MOWLAWN} \sqsubseteq \neg\exists \preceq_m \text{TUESDAY} \sqcap \neg\text{MOWLAWN} \quad (6.9)$$

$$\text{TUESDAY} \sqcap \neg\text{MOWLAWN} \sqsubseteq \exists \preceq_m \text{TUESDAY} \sqcap \text{MOWLAWN} \quad (6.10)$$

Obviously this pair of normative rules will present us with dilemmas with respect to a certain class of behaviours in which TUESDAY is the case, but the two norms are not *necessarily* in conflict: there is no *logical* inconsistency in the chosen formalization. Neither does the reversal of WORKDAY and TUESDAY result in inconsistency.

Proposition 42. A terminological normative conflict between a pair of normative rules only exists if either the antecedents or the consequents of the deontic propositions that were translated to normative rules are equivalent.

The conclusion must be that the given formalization finds only a subset of the types of normative conflict discussed in section 6.3. It differs substantially from the informal presentation based on sets representing the extension of concepts used in that section, resulting in an important deficiency. It however does so on the safe side: the normative conflicts found are definitely to be considered normative conflicts.

The most obvious way to address this deficiency is to expand the set of normative rules in the compound normative rule, by inventing new ones. Let a pair of normative rules n, m of interest for instance express $O_n(\neg\text{ONLAWN} \mid \text{WORKDAY})$ and $O_m(\text{MOWLAWN} \mid \text{TUESDAY})$; Given $\text{TUESDAY} \sqsubseteq \text{WORKDAY}$ it makes sense to conjecture that a rule $O_n(\neg\text{ONLAWN} \mid \text{TUESDAY})$ exists, based on n . Among the

many arbitrary – and defeasible – conjectures one could make of this kind, this one is interesting because of the existence of *m*.

Given this deficiency – which does not affect practical planning and dilemma in planning – there is still a case to be made for adding a deontic logic on top of this proposed interpretation for the purpose of discovering more normative conflicts.

In summary, this section constructs two general characterizations of the problem of conflict or dilemma. A planning dilemma exists if two alternative actions are ordered in opposite directions by two normative rules. If only the alternative actions in the abox are considered, one obtains an epistemic interpretation of violation and dilemma. There is also a more specific criterium for true normative conflict: two normative rules are intended to be part of one single coherent normative order, but cannot be conceived of as a single compound normative rule. The normative conflict necessarily exists between both normative rules. Somewhere in between these two criteria is a class of normative conflicts that are only normative conflicts for a specific category of cases, and, as far as we should be concerned, and *only when a source of law states it is*.

Sources of law express resolutions of normative conflicts in the form of choice rules. These are introduced in the next section.

6.6. Choice between Legal Rules

We all know that beliefs can sometimes be wrong, so intelligent beings need to be able to revise beliefs when they acquire new information that contradicts their old beliefs. Reasoning systems modeling this phenomenon are called *belief revision* systems.

One common way of determining which beliefs should be surrendered is to use a so-called *epistemic entrenchment* ordering (viz. [208]) discussed in section 2.4. This ordering expresses the idea that some of our beliefs are more fundamental than others; It is a “betterness” ordering on beliefs.

This preference ordering is distinct from the betterness ordering that describes a normative order. The preference ordering on beliefs guides us in surrendering beliefs and adopting new beliefs as our understanding of a situation improves. The preference ordering alluded to by the normative order orders situations and actions and guides us in determining whether an action is desirable, and how a situation is to be avoided or reached.

Legal reasoning, like most domains of reasoning, involves both types of ranking the better over the worse. Since both types of reasoning are often successfully assumed to be independent, i.e. no deontic preference ordering ever follows from an epistemic preference ordering and vice versa, we often limit our attention to each type of reasoning individually (as pointed out in section 2.4.1).

The mapping between brute and institutional realities occasionally turns out to be less than perfect. Our beliefs about brute reality are generally speaking epistemically deeper entrenched than our beliefs about institutional reality: we generally assume, in the context of LKBS, that institutional facts only exist as long as it is consistent to believe they exist. In the case of law this means the following: If application of the

whole set of rules leads to absurd results, we must obviously choose not to apply one of the rules.

6.6.1. Lex Posterior, Specialis, and Superior

Makinson and Gardenförs postulated in [117] that there is a tight connection between belief revision and nonmonotonic logics. Belief revision leads to *temporal nonmonotonicity*, i.e. the set of beliefs does not grow monotonically with time. Default reasoning leads to *logical nonmonotonicity*, i.e. the set of consequences does not grow monotonically with the set of premises.

In law we find a parallel process: we assume that over time legal institutions refine the rules as they get better at capturing the intended normative order. In legal theory we find the principles of *lex posterior derogat legi priori* (*lex posterior*) and *lex specialis derogat legi generali* (*lex specialis*) (cf. generally [265, 189]).

The *lex posterior* principle entails that, in case of irreconcilable conflict, the later legal rule will take precedence over the earlier legal rule, and the *lex specialis* principle entails that, in case of irreconcilable conflict, the more specific legal rule will take precedence over the more general legal rule. There is also the third principle of *lex superior*, or *lex superior derogat legi inferiori*, which states that the higher legal rule will take precedence over the lower legal rule (cf. generally [265]).

The *lex specialis* and *lex posterior* principles *describe* certain phenomena of reasoning in general. They do not sanction the preference of the newer over the older, or the specific over the general, but merely observe that it usually *is* so.

The *lex specialis* and *lex posterior* principles do not have to be explained in many words to laymen; They will be naturally applied even by children in contexts outside law. The *lex specialis* principle, as a kind of logical nonmonotonicity, is based on the principle of parsimony in communication (cf. section 2.5). We expect the reader or listener to infer – to some degree – that our more specific statements are *exceptions* to the more general ones if they appear to be in conflict. The *lex posterior* principle, as a kind of temporal nonmonotonicity, is based on the assumption of *improvement*: if later and earlier statements from the same source appear to be in contradiction, the reader or listener will generally speaking assume that the later statement reflects a better understanding of the issue by the author than the earlier statement, and that the author of the statement is aware of the fact that he is revising an earlier statement. We take for instance for granted that the latest publication discussing a theory reflects the best understanding of that theory.

As I explained in [48], the situation is more complicated in law when reasoning over cases that happened in the past. This situation is discussed in the context of version management in section 5.3. In normal circumstances we would never consider applying the scientific theories of the 10th century to the reconstruction of an event in the 10th century, but in law we are often asked to apply the rules of the past to a case of the past.

Logical and temporal nonmonotonicity have arguably become explicit in law, as the legal principles of *lex specialis* and *lex posterior*, because law can manipulate these principles by for instance instructing the reader to not apply them. It is exactly because

legal reasoning occasionally deviates from the intuitions embodied by *lex specialis* and *lex posterior* that in the context of law people will start doubting their common sense judgment.

Conceptually, application of the principles is easy to understand. Any provision obviously has a history – when did it enter into the norm system and when where changes last made to the text – and it has propositional content. A *lex specialis* ordering between two provisions can be discovered by comparing the provisions content-wise. A *lex posterior* ordering between two provisions can be discovered by comparing the history of the provisions.

The *lex superior* principle is also ‘intuitive’ enough, but it is a design principle for complex, layered legal systems. An act of parliament for instance typically takes priority over a royal decree, which takes priority over a municipal decree, etc. This hierarchy is not something which is discovered by application of the principle to a pair of provisions. It has been purposefully designed into the legal system by a legislator who foresaw the possibility of having to choose between the rules. Provisions overruling the normal activity of *lex specialis* and *lex posterior* usually define a preference ordering on sources of law that can be used for choosing between norms. The *lex superior* ordering is a consequence of the provisions that manipulate the application of *lex specialis* and *lex posterior*.

It is only natural to wonder whether legislators cannot invent additional principles beyond the three mentioned. These would however all be classifiable as instances of *lex superior* choice rules. An odd example (*viz.* [260]) is for instance to leave the choice which rule to apply to the defendant. The implementation of the *lex superior* principle as a rule would in this case be dependent on the decision making setting, but still be classifiable as an instance of *lex superior*.

The logic of the legal system dictates that *lex superior* should take precedence over *lex specialis* and *lex posterior*. The relative priority of the *lex posterior* and *lex specialis* principles among themselves is however not necessarily settled. The reason for this is that they are implicitly assumed to reinforce each other. Assuming that the legislator refines his expressed preferences over time, is aware of his own acts in the past, and intends it’s new provisions to be compatible with the existing corpus, it is only reasonable to expect that new preference statements refine the existing system. The *lex posterior* ordering and the *lex specialis* ordering are in other words expected to point the same way in most cases, while the *lex superior* ordering is expected to point in the opposite way.

In the case where an older rule is more specific than a newer rule, the older rule seems to amend the newer one – which is absurd – and there are good arguments for precedence of both principles: the legislator could have repealed the older rule but didn’t, and the older rule is not likely to have been intended as an amendment to the first one. To the author the first argument is more convincing, and therefore *lex posterior* should defer to *lex specialis*⁴. The law however claims priority over what we consider common sense: the status of this type of conflict has to be verified for each case where it is encountered.

⁴But [266] for instance takes the opposite point of view.

Absurd is the case of a symmetric disaffirmation conflict where even the *lex posterior* principle doesn't offer a solution because the involved provisions were adopted at the same time. This case cannot be solved by interpreting the intent of the legislator, since a rational legislator cannot possibly have intended to contradict himself. In the Netherlands this case would for instance be handled by a detailed provision⁵ that proposes that the order in which the monarch signs⁶ and the sequential order of publication is decisive.

Any criterion will do: For example, one U.S. court held that when two amendments adopted at the same time are irreconcilable, the one receiving the greater number of votes in the Senate takes priority over the other (viz. [260]). The absurdity of this situation becomes clear if one realizes that at least some senators voted to support two amendments with the exact opposite propositional content. This exotic principle has not been accepted in jurisprudence anywhere (including the US).

Typical for law is however that there is assumed to be some rule that decides which rule to choose in the case of conflict between a pair of rules. There is no freedom of choice in the matter. Once such a principle has been decided on for a pair of rules, it is subject to *stare decisis* and forms a precedent, an *epistemic obligation* – just like any other rule – to decide in conformance with the new rule.

6.6.2. Ne bis in idem

We do not however only have to choose between rules if they are in conflict. There is another legal principle that can also be conceived of as having to choose between rules that does not involve a conflict.

In this case we have to choose between the application of two rules even though no inconsistency, nor a circularity between alternatives, has arisen.

Affirmation of obligations (cf. 6.3) does not generally require a choice between the involved norms, but there are real world situations where this choice is demanded by law. In criminal law it is a generally accepted rule that you cannot be prosecuted twice for substantially the same facts. Since there may be several crime descriptions involving for instance unlawful killing, coercion, or defamation, one will occasionally be required to choose between norms with an obligatory character in this area of law.

Ne bis in idem means that no legal action can be instituted twice for the same cause of action. Closely related is *double jeopardy* in common law jurisdictions, which is limited to criminal law. The reason why we have to choose between rules in this case is that the institutional facts created by these two rules cannot be jointly applied to the same thing because doing so would be unjust. This rule is however often limited to criminal law: it is perfectly valid to prosecute for reckless driving while giving an administrative fine for speeding for substantially the same behaviour, provided that speeding is part of another domain of law exempt from the *ne bis in idem* rule vis-à-vis criminal law. The *ne bis in idem* principle only functions in areas of law where it is

⁵ *Aanwijzingen voor de Regelgeving* 1990, art. 173a.

⁶ Contrary to what one might think this order is no longer unique: certain decisions with the status of royal decree, naturalization decrees, are signed once per batch. Apparently no problems are expected here.

specifically declared as a positive principle: it is for instance found in the European Human Rights Convention.

The *ne bis in idem* principle usually applies to an affirmation relationship between two normative rules. For instance insult is an abusive attack on one's honour or dignity, and defamation is an insult involving a false accusation or malicious misrepresentation. In Dutch criminal law one can be prosecuted for either one or the other, but not both. The Netherlands is one of the few jurisdictions that (still) consider these behaviours crimes in addition to considering them torts; A criminal prosecution does not however prevent a (civil) tort action.

Nearly universal is the distinction between murder and manslaughter, or even more grades of increasing seriousness of responsibility for someone's death. There appears to be a kind of informal choice rule that says that one usually prosecutes for the offense with the greater punishment, and this rule often happens to coincide with *lex specialis*: the crime with the more specific description is the one punished more. This suggests an implicit order $\text{DEFAMATION} \succ \text{INSULT}$, $\text{MURDER} \succ \text{MANSLAUGHTER}$.

This is however a prosecution policy which arises naturally from the personal preferences of public prosecutors and victims and not a general principle. If someone is for instance the victim of a defamation, but establishing that this is the case would bring additional harm⁷, the victim will generally prefer a charge with insult instead of a defamation and has the freedom to choose in his own interest.

In this case the operative principle is that the decision maker generally has a freedom of choice between the rules, unless the law states otherwise. It therefore also means that a constitutive *legal act* is performed. If the freedom of choice is restricted by a specific rule, one is (epistemically) obliged to decide in conformance with the rules.

6.6.3. Representation of Choice Rules

The issue of choice between legal rules only comes up for indicative rules. Indicative default rules are applied only, producing a result, if it is consistent to do so, and in some cases only if one intends to bring about that result. Terminological and requirement rules necessarily apply.

If application of an indicative rule makes the knowledge base inconsistent, we retract a some default consequent of an indicative rule and decide that the indicative rule that sanctions it not applicable. Normally speaking, an important epistemic constraint regulates this revision activity in law: the revision must be consistent with choice rules.

Choice rules impose a kind of partial order over defaults by making the applicability of legal rules depend on the application of other legal rules. Choice rules are applicability rules, and just like all other legal rules discussed in this book, they are based in a source of law.

⁷Take for instance the scenario of a married man who visits prostitutes, and is publicly accused of visiting underage prostitutes. Regardless of who has the burden of proof, establishing the truth of the matter will bring additional harm.

In the previous sections a number of different perspectives on choice between rules, and a number of different *reasons* to do so were discussed. There is difference between choice between two indicative rules:

1. because of an inconsistency in institutional reality if both are applied to the same fact;
2. because of incoherence in the intended normative order ascribed to the legislator as evidenced by normative conflict or the dilemma, if both indicative rules are normative; and simply
3. because they cannot be applied to the same fact according to the rules or a decision making policy, even though doing so would lead to no inconsistency or incoherent normative order.

That application of both rules would lead to inconsistency is not of direct relevance to the choice: if the choice rule says that two rules cannot be both applicable, then they necessarily are not. No inconsistency or conflict is required as a condition.

The application of normative rules was implemented in section 6.2 by the standard default rules that are also used for other indicative rules. A pair of rules that is in normative conflict will however in the normal interpretation of normative rules as constitutive rules that derive the institutional fact *allowed* or *disallowed* with the rule itself as axis of reference never be inconsistent with each other: there is in principle no logical problem with substantially the same behaviour being allowed by r_1 and disallowed by r_2 . This inconsistency would only arise if one makes an assumption about a set of rules R even though R contains conflicting rules: if behaviour b is disallowed according to $r \in R$ then behaviour b is also disallowed by R . Set R can be conceived of as a compound normative rule. The underlying assumption, disproven by the verifiable existence of normative conflict, is that R , conceived of as a simple set of rules, formalizes a coherent normative order. If this assumption does not hold, one can retreat to the position that set R , which cannot be represented by a compound normative rule, *and* a set of choice rules formalizes a coherent normative order.

As pointed out in section 3.3, prioritized default reasoning with indicative rules and OWL DL can be conceived of as the use of an *incision function* for belief base revision. The purpose of the incision function is to specify an epistemic entrenchment theory on beliefs, to decide which belief should be retracted in case of inconsistency. Choice rules do not define such an incision function, but constrain it by imposing a noncircular partial order $\{N, \prec\}$ on rules. How deeply an institutional belief is entrenched is determined by the rule that generated it. Even a comprehensive catalogue of such ordering constraints would still leave freedom in the construction of coherent models of institutional reality for the same case that can i.a. be used for tactical purposes in argumentation. The partial order does not replace a sound incision function.

The notion of a noncircular partial order on legal rules representing their priority, and propositions that describe this order is common in computer science & law. See for instance [133, 179, 266, 286, 230] for examples of use of this explanatory device. There are however considerable differences in approach.

[266, 286] for instance only apply it to normative rules, and develop, for normative rules of limited expressiveness only, a method to compute the application of *lex*

specialis to normative conflict (which is criticized in section 6.3.1). The method of [286] subscribes to the notion that a pair of rules must be necessarily in conflict: the partial order on rules can therefore be completely compiled out in advance using a terminological concept of conflict, and the principles of *lex specialis*, *lex posterior*, and *lex superior*. In [266] the computation of normative conflict depends on the case (cf. section 6.3.1), and conflicts are resolved as they are found. A property that both approaches share is that a determined order of priority between two rules is *never* conditional on anything.

The accounts found in [133, 179, 230] potentially apply to any pair of legal rules, and they moreover allow for the possibility that priority between two rules depends on the setting in which the priority relationship is determined because a choice has to be made. This allows for rules that determine the priority between two legal rules for a specific category of cases, which was suggested as a conceptualization of *intersection conflict*.

Neither of these approaches makes a conceptual distinction between “normal” de-feasibility of legal rules and the resolution of perceived normative conflict, and they moreover all depend on the notion that choice only follows from conflict or inconsistency.

Hage and Verheij (i.a. [133]) also take a strong position with respect to the status of (court) decisions in general, including the decision to give one rule priority to the other. Such decisions in their opinion cannot have the status of a legal rule. Principles such as *lex specialis* can be said to be part of positive law, but the application of the principle in a decision is not. Here we do not adopt this point of view, which is rather philosophical: for LKBS development a court decision that explicitly orders a pair of rules will always be considered safer ground than the opinion that *lex specialis* applies to them and imposes a certain ordering. This issue is addressed in more detail in section 6.7.2.

If one takes a cautious approach to representation of sources of law one does not try to precompile a partial order based on theoretical arguments on how such priorities arise, but simply represents the ordering as it is made explicit over time by actual decisions and case law, at least by the ones which of one believes that they have a *stare decisis* effect. This is not automatically the case when freedom of choice is involved, like in the *ne bis in idem* case, or when a decision maker chooses between rules to resolve a compliance dilemma (*intersection conflict*). For a reusable representation of sources of law there is of course a good argument to be made for this. Being overly cautious however leads to ignoring the underlying normative order, and makes legal institutions fundamentally incomparable on the functional level, as explained in section 6.7.

The representation solution chosen here is in the tradition of Kowalski in [179], who came to a similar solution for applicability propositions as this book, and treat choice rules as applicability rules.

The incompatibility between pairs of legal rules as applied to the same behaviour can be enforced by applicability rules. If we for instance say that rule r_1 is defeated by r_2 when applied to the same fact because r_2 is more specific, this fact can be represented as *rule r_1 is only applicable if r_2 is not applicable, according to r_s* :

$$\{r_1\} \sqsubseteq \forall \text{appliesTo} . (\forall \neg \text{applicable} . \{r_2\} \sqcap \exists \text{applicable} . \{r_s\})$$

Rule r_s is not the general principle that justifies the inference, but the specific institutional fact – ideally backed up by a source of law that represents it – that r_1 is only applicable if r_2 is not applicable.

The attractiveness of considering the choice principles as generators of applicability rules is that a generous amount of expressiveness is available. It is for instance easy to state that rule r_3 is defeated by r_4 when applied to civil servants because r_4 is more specific when applied to civil servants:

$$\{r_3\} \sqsubseteq \forall \text{appliesTo} . (\neg \exists \text{actor} . \text{CIVILSERVANT} \sqcap (\forall \neg \text{applicable} . \{r_4\} \sqcap \exists \text{applicable} . \{r_{cs}\}))$$

This immediately gives us a solution for cases of *intersection conflict* in which we believe that a specific decision to choose between the rules *will* have effect on future cases, but only in relation to a specific category of cases, as considered in section 6.3.1. Compare for instance dynamic priority in [230]); Dynamic priority cannot be solved by a straightforward partial ordering $\{N, <\}$ on legal rules only.

Constraints of these type enforce the incompatibility of pairs of rules, but they do not enforce an order for an *incision function* (cf. section 3.3). They perform the ontological function of constraining which models are models of institutional reality. Since application of an indicative rule is an ampliative operation it is consistent to say that r_1 is applicable as long as no one has applied r_2 . These rules indicate that two rules cannot be both applied to the same thing, because of general defeasibility, normative conflict, or otherwise.

What we want to encode are statements of the form *applying r_2 would be better than applying r_1* , or *applying r_2 when the case is about a civil servant would be better than applying r_1* . So far we have however always translated what could be considered an *epistemic obligation* – an instruction on “how to think” – to straightforward axioms, constraints, or default rules. Only in the case of obligations relating to behaviour we consider it necessary to make a distinction between compliance and violation.

If we would want to represent this meta-level domain of applying rules, and consider application of an indicative rule an action, we can simply represent them in the same way as we would represent other normative rules, in a new meta-level domain that consists of a case, facts of the case, rule applications, in a specific order, and resulting facts. No LKBS would however ever take this approach, and so far we have settled on an approach which takes `applicable(CONSTITUTIVETHING,RULE)` as the representative primitive and not the reified `APPLICATION` in relation to a `RULE` and `CONSTITUTIVEFACT`.

In the *LKIF* ontology (cf. [35]) I have provided the ingredients for this approach. Assuming that we commit to doxastic-style beliefs, we can consider alternative beliefs one could hold with respect to the same thing. This is basically the same approach as the ordering of behavioural alternatives. If a proposition **proposes** the existence of something, then the `LEGALRULEAPPLICATIONPROPOSITION` proposes something of the form `∃applicable.LEGALRULE`. The proposition that proposes this is a different

object from the one about which this belief is held. The choice rule then becomes a standard normative rule applicable to $\exists\text{proposes}.\exists\text{applicable}.\{r_1, r_2\}$ ⁸:

$$\begin{aligned} \exists\text{proposes}.\exists\text{applicable}.\{r_1\} &\equiv \neg\exists\text{proposes}.\exists\text{applicable}.\{r_2\} \\ \exists\text{proposes}.\exists\text{applicable}.\{r_1\} &\sqsubseteq \exists \preceq_{r_s} .\exists\text{proposes}.\exists\text{applicable}.\{r_2\} \\ \exists\text{proposes}.\exists\text{applicable}.\{r_2\} &\sqsubseteq \forall \preceq_{r_s} .\neg\exists\text{proposes}.\exists\text{applicable}.\{r_1\} \end{aligned}$$

Any observation made about normative rules also applies here, including the problem of normative conflict. Note that the CIVILSERVANT could also be added effortlessly, as well as conditions on the believer in or the medium of the proposition.

The presence of \preceq between two beliefs is sufficient to make clear that they are alternative explanations about the same underlying thing.

Of course, if one does not take a cautious approach to representation of sources of law one does want to try to precompile a partial order based on theoretical arguments on how such priorities arise, as – amongst others some of whom were mentioned earlier – I have done in [286]. For one thing one obviously wants to express known lex superior orderings that apply to whole classes of rules R_1 and R_2 . Because we do not know that any pair of these cannot be applied together – it is even extremely unlikely – we cannot apply them as explained above. We *can* however ensure that no ordering is made that goes against lex superior, and – as explained earlier – lex superior must necessarily take priority: the lex superior ordering is always *allowed*:

$$\exists\text{proposes}.\exists\text{applicable}.R_1 \sqsubseteq \exists \preceq_{r_s} .\exists\text{proposes}.\exists\text{applicable}.R_2$$

Since these lex superior propositions on whole classes of legal rules do explicitly occur in sources of law, the above could very well be a real legal rule.

The same is not true for lex specialis and lex posterior. These are, as stated before, descriptions of the strategies used in defeasible reasoning in general.

Lex specialis causes specific problems: it is not always clear which one of two rules is the more specific one, certainly if the two rules do not logically contradict each other. Specificity suggests some relation with logical subsumption, but two things that have a subsumption relation cannot be in conflict.

Although attempts to formalize the meaning of lex specialis, such as in [286, 265], are certainly interesting from a legal theoretic point of view, automated application of the lex specialis principle has only little relevance to the representation of sources of law. If lex specialis is only the recognition in the legal field of a pre-existing set of generic purpose epistemic habits that the readers of texts bring with them, as argued before, then the underlying mechanism is a more general one involved in text understanding, i.e. decoding information contained in the text.

This mechanism appeals to common sense knowledge – including knowledge with an ampliative function – that the readers bring to the text, and the conclusion of the

⁸That we have asserted $r_1 \neq r_2$ becomes essential here.

application of the principle is relativized to a context if the argumentation involves ampliative inference.

For a representation of a source of law that is used in LKBS in a variety of settings, only the representation of the conclusion of the application of the principle is relevant, and then only if it represents the policy of the LKBS user. In section 6.7.2 a case will be made for the thesis that setting such policies is a major quasi-legislative function of the courts.

6.7. Elaborating on the Intended Normative Order

As pointed out in section 4.7 the objective of this book is to represent legal rules while making minimal epistemic commitments to the way of using these legal rules. At the same time we have to account for two primary uses of legal rules, being planning and plan recognition as pointed out in section 4.1.

Planning involves generating and comparing alternative plans, the execution of which involves behaviour, while situation and action recognition involve generating and comparing consistent alternative explanations of observed behaviour. Both are subject to legal rules: normative rules directly address planning, while all legal rules directly address explaining behaviour in legal terms. Sometimes they have the role of constraints, ruling out certain alternative plans or explanations, sometimes they create viable alternatives, or at least make alternatives that would otherwise be extremely unlikely or undesirable more salient because the behaviour of other agents has become more predictable. Generally speaking, the objective of planning has little to do with the legal rules: law plays only a minor role in people's lives.

Sometimes a rule can be applied if one intends to bring about a certain legal result by proposing that one is executing a plan to that effect (one undertakes a *constitutive act* as explained in section 4.2; One tries to construct a plan that meets $\exists \text{executes}.\exists \text{causes}.\exists \text{initiates}.\text{INTENDEDLEGALPOSITION}$). This effect, intended by the legislator, remains unexplained if we do not consider the larger normative order intended by the legislator.

Chapter 5 proposes a representation of all legal rules, but only the function of normative rules has been elaborated on. By qualifying certain propositions as allowed or disallowed, the legislator attempts to make behaviours occur more, respectively less, often than they did before the normative rule was created. The legislator articulates a preference, in the form of a rule, with the intent that it will motivate the actions of its addressees. The legislator in the Netherlands for instance has publicly announced that he prefers us to drive at the right hand side of the road: underlying this rule are presumably preferences for 1) drivers driving on the same side to prevent accidents and generally increase the capacity of the roads, as well as the speed at which drivers can safely drive it, 2) drivers not having to come to a consensus whether we should keep to the left or right, and 3) not having to switch road side when you drive into a neighbouring country.

Normative rules have an important function for bystanders, regardless of whether the allowed behaviours are also preferred by the bystander: because the legislator is generally successful in making disallowed behaviours occur less often, the bystander

expects that the allowed alternative is much more likely than the disallowed alternative, therefore reasons that it should therefore be (epistemically) preferred as a prediction, and decides based on that assumption. This applies as much to the driver who expects the other driver coming from the left to yield as it applies to the driver who has to yield and who expects that failure to do so may firstly result in an accident, and secondly in additional negative consequences for him made salient only because of the existence of rules that facilitate organizing punishment of his violation.

Non-normative rules are only relevant to situation recognition, with or without a following planning stage that uses the normative rules. Strictly speaking they only mean something in relation to their capacity to trigger normative rules.

The notion that the normative rules are supposed to form a coherent normative order involves an abstraction (to a form of [183]’s up-down metaphor): the \prec_n relation is a relation between domain $|\exists \text{disallowedBy}\{n\}|$ and range $|\exists \text{allowedBy}\{n\}|$. Taking the normative order intended by the normative rules, including the associated choice rules, at face value, decision problems involving the normative order can be classified into the following four categories⁹:

Assessment: Does an action a violate normative order \mathcal{N} ?

Envisioning: What categories of cases does normative order \mathcal{N} distinguish, and how does it order them?

Ordering the alternatives: Given a settled context description \mathcal{C} , and a menu \mathcal{A} , and an $a_1 \in \mathcal{A}$, $a_2 \in \mathcal{A}$, order a_1 and a_2 in accordance with normative order \mathcal{N} .

Choosing the best alternative: Given a context description \mathcal{C} , and a menu \mathcal{A} , choose all $a_1 \in \mathcal{A}$ for which there exists no $a_2 \in \mathcal{A}$ such that $(a_1 \prec a_2) \in \mathcal{N}$.

The two differentiating features are whether one is using the normative system with or without a settled context, the situation features present in each alternative, and whether one is ordering or minimizing alternatives.

Assessment can be taken to be a variety of choosing the best alternative (an action violates the normative order if it is *not* in the set of best alternatives). A major difference however exists between the concept *disallowed because the normative order implies that there must be a strictly better alternative* and *disallowed because you chose an alternative that is strictly worse than another alternative you (must have) considered*. This is the essence of the *compliance dilemma* as presented in section 6.3.1. Whether to use the first or second conceptualization depends to a large extent on the knowledge representation language one uses: the first criterion is a classical one, while the second one is epistemic, and undoubtedly the better account of how norms are actually used.

Classification systems in LKBS will usually use the classical conceptualization (cf. generally section 4.6) – some behaviours are disallowed and the alternatives are irrelevant – with the risk of creating compliance dilemmas. For planning one would generally want to consider the second approach, which allows one to take into account compliance dilemmas and resolve them, but of course with the risk that alternatives are overlooked that others, that ascribe an intention to your behaviour, did not overlook.

⁹loosely based on [268]

On the logical level we also have to distinguish ordering either alternatives or sets of alternatives. In a normative theory that does not contain conflicts, this distinction is of less importance than it would be in a true logic of preference, because we are sure – given the definitions of deontic operators in terms of preference – that the preferences are all properly partitioned (if $P \succ Q$ then $|P| \cap |Q| = \emptyset$). Because we are assured of this, as long as we are able to determine in which order to consider application of conflicting normative statements these do not pose a complex problem, but one which will not be addressed in detail in this book.

Exposing the normative order intended by the legislator is however not merely a matter of envisioning and ordering the categories of cases that the legislator's rules distinguish, as sections 4.5 and 4.2 pointed out: normative order always exists, regardless of whether it is formalized, and the legislator often just tinkers with the details of an existing social order in society. Sometimes the actions of the legislator only confirm what is already *normal*, by formalizing it, or perform minor repairs in existing voluntary normative order, and sometimes the legislator only tries to create the right conditions for acting in one's own interest without obstructions or harming others (e.g. facilitating a free market). Institutional reality and the intended normative order are different things.

The legislator acts on an existing normative order, or more accurately on his preferred explanation of the existing normative order, and performs legislative actions with the intent to change it. The same is true on the level of the institution, as pointed out in the previous chapter: the sources of law are not a blueprint of the institution, but a log book of relevant legislative changes to it. The legislator relies on three important mechanisms to effect changes in the normative order: public announcement of the rules, the unique ability to effectively organize punishment, which underlies the normative rule, and the widely held preference for predictable social interactions among the addressees of his legislative acts (the scripts of section 4.5).

The legislator's acts obviously sometimes *fail* to achieve the intended effects on society. Besides that, the legislator is usually not a single natural person: it is highly unlikely that the actual people involved (for instance a parliament) share a coherent theory of an ideal normative order. The legislator only has intentions in the same sense that a company has a strategy: we are only likely to take the notion seriously if written evidence of the strategy is published in the name of the entity.

Institutional reality works separately from normative order, but some notions do appeal to normative order, or the intentions of the legislator: rights, compliance dilemmas, conflicts, unrealizability of norms, etc. An intended normative order is ascribed to the legislator, because we need to identify when the legislator fails to create the intended normative order. Section 6.7.2 will show that the courts go beyond the legal rules in the interest of creating a coherent normative order.

A subjectivist would argue that the way we motivate our actions is beyond rational criticism. This especially applies to the motives of the legislator, who is not bound to any requirements of rationality and not a rational agent in the normal sense. Suber (cf. [260]) expresses this view when he observes that the legislator can in principle abrogate any common sense knowledge or widely accepted logical inference and replace it with legal fiction. The legislator is obviously not omnipotent or omniscient, but correction

of the legislator should be achieved through the democratic process and not the legal system.

Action can however still be rightly criticized as irrational. The first kind of mistake is when action is based on false beliefs. If one accepts that the action a I intend to do in actual situation s and the action I would intend to do in the same hypothetical situation s are the same, which is widely accepted, then I should be willing to change my behaviour if false beliefs I hold are corrected by someone else. If I learn that I am actually in situation s' and I already believed that in a hypothetical situation s' I would intend to do a' instead of a , then I will intend to do a' instead of a . Actions are after all understood to be precisely those events that are causally explained by the intentions of the actor (confer section 4.2).

The second kind of mistake depends on the possibility of a difference between my revealed preferences and my beliefs about the preferences that motivate my behaviour. Section 2.4.1 explained the silliness of equating revealed preferences and motivating preferences in the form of a joke: revealed preferences are ascribed, while motivating preferences are policies or *rules* one adopts. If there are such rules, then it is possible to correct them by reflecting on the effects of one's actions.

As noted in section 6.5), a preference theory represents the *norm* implied by one normative rule, or if preferred a set of normative rules provided they are consistent with each other. A set of preference theories, and a preference theory about choosing between them, describes a normative order. Normative rules are interpreted as subjunctive propositions about *betterness*, made by a legislator, that are intended to play a role as motivating preferences in decisions.

To explain the difference between these two views on preferences, we have to delve a bit deeper into the notion of choice. An *induced decision rule* $\mathcal{C}(\mathcal{A})$ for a menu \mathcal{A} selects an alternative from the menu. Decisions can be conceived of as applying an induced decision rule. A rule $\mathcal{C}_\succ(\mathcal{A})$ can be directly induced from a preference relation \succ if it is completely preorder (complete and transitive) and asymmetric:

Definition 1. (Induced Choice Rule) Given a relation \succ , the Induced Choice Rule \mathcal{C}_\succ is defined by: $\mathcal{C}_\succ(\mathcal{A}) = \{x \in \mathcal{A} : x \succ y \text{ for all } y \in \mathcal{A} \text{ and } y \neq x\}$

That a preference relation is transitive and asymmetric is a reasonable assumption to make of a rational agent. There are however two different ways of looking at the relation between preference and choice. One theory, let's call it the *revealed preference* theory, postulates that preferences are revealed by choice between alternatives. That is:

Revealed preference : $a_1 \succ a_2$ if the agent chooses alternative a_1 over alternative a_2 .

That choice always entails a conscious comparison of *all* alternatives and therefore a preference is not entirely obvious; people enthusiastically choose between equally valuable products on a daily basis, and for instance show no signs of distress when confronted with the choice between a well known and an unknown alternative in the supermarket (cf. [138] for a discussion). Revealed preference theory solves this problem by postulating that if a comparison would result in preferring a_1 over a_2 then I already

prefer a_1 over a_2 but didn't *know* it yet. Our preferences can be discovered through comparisons. A corollary of this thesis is that preference orderings over alternatives are always complete.

The other theory, let's call it the *motivating preference* theory, postulates that we can be truly indifferent to alternatives while still making choices. That is:

Motivating preference : the agent chooses alternative a_1 over alternative a_2 if $a_1 \succ a_2$.

An incomplete preference ordering cannot be used directly to induce a decision rule. Instead we can only say that a certain decision rule is consistent with a *preference theory* (cf. section 6.5), or is a *bisimulation* of it. It is for instance possible to induce a choice rule based on the assumption that any choice between alternatives to which the original choice rule is indifferent will do.

It would seem odd to distinguish *motivating indifference* – revealed indifference certainly doesn't exist – but such a notion is useful in law. It is obviously meaningful if one adopts motivating preference *and chooses* between incompatible incomplete decision rules like one does with (legal) choice rules because one is not able to formulate a single coherent one. Permissions play the role of a motivating indifference, but *only* in the sense of restricting how we can extend the normative order by formulating additional normative rules. The addressee is only bound by \succ , not by \succeq . As long as a legislator values freedom of choice, the legal system will always formulate incomplete choice rules.

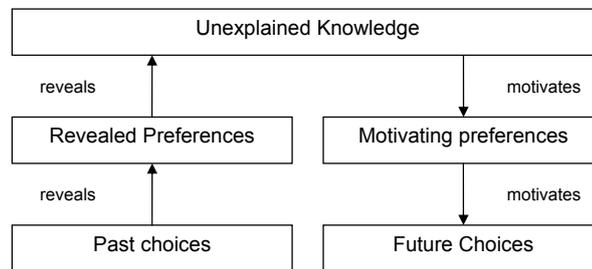


Figure 6.1. The problem with the idea that the preferences are revealed by choices, and the choices motivated by preferences is that these two categories of preferences cannot refer to the same thing.

Revealed and motivating preferences are clearly different (see figure 6.1): we can in fact speak of a legislative monitoring cycle: to legislate involves reflecting on choices one saw people making in the past, formulating hypotheses on why these choices fail to bring about the intended normative order, and formalizing new rules to guide choices in the future.

One can of course criticize the coherence of the normative order as created by the legislator, or point out that this and that is against some important value. The idea that one can simplify the system by isolating core values from it is however simplistic. Attention for undesirable outcomes in specific cases tends to increase the complexity of the law.

To understand why, we need to appeal to the property of preference independence (cf. for instance [240]). Hansson ([138]) distinguishes between preferences among incompatible alternatives, which he refers to as *exclusionary preferences*, and preferences among compatible alternatives, which he refers to as *combinative preferences*. When I choose a red car over a non-red car, I am exhibiting an exclusionary preference, because the alternatives are incompatible. My preference for having a cheap car rather than a red car, by contrast, is combinative, because the alternatives are compatible. The norms in legislation are all examples of a particular type of statements of exclusionary preference: $|\alpha| < |\neg\alpha|$ and $|\alpha| \cap |\neg\alpha| = \emptyset$. If these are preferentially independent, they are easily composed into a coherent system. If not, we will be faced with the situation of having to choose between them on occasion.

But in law the property of *preference independence* clearly often does not hold¹⁰. Given propositions P and Q , it is not generally the case that Q is preferentially independent from P . What is understood by preference independence is easiest explained by an example: Let P mean *running over a pedestrian in your car* and Q mean *calling 911 and telling them you ran over a pedestrian*: observe that it is entirely possible and reasonable that P is disallowed, and Q is allowed if P and disallowed if $\neg P$. This is essentially what happens in the contrary-to-duty situations discussed in section 6.4.

This strictly speaking rules out the use of many valuation tools used in decision theory for choice between alternatives like additive value functions, multiplicative utility functions, and other economic concepts based on the assumption of preference independence (cf. [169, 240]). Decision making subject to legal constraints therefore does not lend itself very well for approaches based on the *expected utility hypothesis*. The use of decision support methods like Multi-attribute Utility Theory (MAUT; cf. [104]) is ruled out by this criterion.

This lack of independence between evaluation dimensions makes it hard to conceive of underlying values in law that are not just defeasible-motivating-preferences-in-a-context, just like the original legal norms found in normative rules. Specific norms may sometimes affirm more general ones – the notion of *affirmation of an imperative* (cf. 6.3) – but legal norms, regardless of whether they are dressed up as fundamental principles or values, are always defeasible if we interpret them as a coherent normative order. The intentions of the legislator cannot be reduced to a number of independent core values that happen to be expressed in context-specific settings in the form of a specific norm.

Still, this does not discourage anyone from ascribing underlying preferences (not just intentions) to the legislator based on his observed actions, to analyze the normative-order-creating actions of the legislator using another, smaller, set of norms – norms of analysis – or for courts to attribute mistakes to the legislator and to fill in the gaps left in the implied normative order over time.

As alluded to in section 4.5.2, preferences are generally attributed to *agent* roles in a script; Human beings cannot be attributed a single comprehensive agent role, and the total set of preferences ascribed to them in order to explain their behaviour need not be consistent. For normative rules the same is true: normative rules apply to an

¹⁰I am not claiming that it does hold in some other domain, just that in law it *clearly* doesn't.

agent role.

6.7.1. Preferences and Scripts

When an agent makes a decision what to do, he has to deal with uncertainty. He has to weigh the utility of the foreseen possible outcomes of the alternative actions he can perform with their likelihood. This is the standard economic formulation of a decision problem, as solved by the expected utility hypothesis (see section 2.4.1).

The most important wild card an agent has to deal with in social interaction is the behaviour of other agents. The outcome of his actions are usually most sensitive to the decisions of the other involved agents: the preferences they have determine the epistemic preferences he should have towards recognizing their intentions. Even in a deterministic planning environment like chess this is shown to be very hard, while chess has a relatively low degree of complexity compared to potential social interaction between agents.

In real world planning a relaxed version of the problem is solved: because other agents available as resources in the planning problem are clearly not obviously in direct competition with the problem solver, the planner is limited to executing a number of social *scripts* that usually work: the alternative outcomes of an action are reduced to the one that normally occurs. This is considerably closer to what humans do, as discussed in the reflection on normal behaviour in section 4.5.

In the Hohfeldian scenario of section 4.8.2¹¹ this means that two expected scenarios are open for *a* who offers something for sale to *b*:

$$\begin{array}{c} \text{OFFERFORSALE}_a / \neg \text{OFFERACCEPTED}_b \\ \text{OFFERFORSALE}_a / \text{OFFERACCEPTED}_b / \text{SUPPLY SOLD ITEM}_a / \text{PAY}_b \end{array}$$

There is a norm n_1 which tells us:

$$\begin{array}{c} | \text{OFFERFORSALE}_a / \text{OFFERACCEPTED}_b / \text{SUPPLY SOLD ITEM}_a | \succ_{n_1} \\ | \text{OFFERFORSALE}_a / \text{OFFERACCEPTED}_b / \neg \text{SUPPLY SOLD ITEM}_a | \end{array}$$

There is a norm n_2 which tells us:

$$\begin{array}{c} | \text{OFFERFORSALE}_a / \text{OFFERACCEPTED}_b / \text{SUPPLY SOLD ITEM}_a / \text{PAY}_b | \succ_{n_2} \\ | \text{OFFERFORSALE}_a / \text{OFFERACCEPTED}_b / \text{SUPPLY SOLD ITEM}_a / \neg \text{PAY}_b | \end{array}$$

Agent *a* prefers compliance with n_1 and expects and prefers compliance of *b* with n_2 . Agent *a* moreover prefers the first scenario over the second, while agent *b*'s choice is unpredictable.

¹¹ Assuming a particular order here, and simplifying the representation. Since these are institutional actions, a real world action of paying may for instance constitute accepting the offer. In addition the *deadline* problem of section 4.7.2 is left unaddressed here.

In the case of normative rules this observation for instance leads to the prediction that people generally comply with the norms, but this doesn't tell us what they will do with the alternatives they have for changing or retaining their legal position. What legal positions people generally prefer to have is a separate modeling problem: as stated in section 6.7 the quality of a legal position is reducible to the normative rules, but in the end the question is whether b is interested in $\text{SUPPLYSOLDITEM}_a/\text{PAY}_b$ over $\neg\text{PAY}_b$: if b is, then b improves his position by OFFERACCEPTED_b . In the case of a sale that is hard to predict, since it depends on the item in question.

This is certainly not always the case: people generally like rights, eligibility to benefits, powers, privileges, immunities, and dislike duties, disabilities, liabilities, etc. Although planning is often goal-directed, there is a strong positional aspect (cf. [167]) to it, certainly when positional improvements come cheap. In predictable settings like the one described in section 5.2.1 (cf. [272]) this kind of thing goes without saying: very often we can go a long way by assuming that other people have the same preferences as we do.

Whether establishing a legal fact is a good or bad thing for an agent is generally clear; The legislator leaves clear hints how a concept is to be evaluated by the words he chooses, certainly relative to a setting, and if these didn't already have a clear evaluative interpretation, they will acquire one over time:

$$| \text{TAKING} \sqcap \text{THEFT} | \prec_p | \text{TAKING} \sqcap \neg\text{THEFT} |$$

A normative rule whose case is expressed by one concept, *makes* this concept value-negative or value-positive (cf. [138, 139]). In many other rules concepts also become so over time because of the constellation of rules around them. A legislator would not generally change this evaluative meaning a concept has acquired without inventing a new word to describe it¹².

In relation to scripts, section 4.5.2 distinguished between the physical human being and the *agent* role he plays in a script. The person plays an agent role, the agent fills the *actor* role, and other roles, in an action. The agent is identified by the set of actions that is performed in that role, but in addition certain preferences relevant to his role can be ascribed to him (ascription being the application of a default role when it is consistent to do so). It is the agent, for instance a buyer or seller, with the actions that characterize his role, and the default set of preferences ascribed to him, that explains the small set of normal executions of a script.

The preference theory attributed to the agent can be abstracted to a reified decision rule believed in by the (player of the) agent, just like the rule represented in the source of law, although this a mental concept and not an institutional one: \prec_a relation is a relation between domain $| \exists \text{disallowedBy}.\{a\} |$ and range $| \exists \text{allowedBy}.\{a\} |$, that describes the decisions normally made by agent role a , and works like the normative betterness relation in section 6.7.

Some actions, like the action of buying/selling, involve multiple actors. A completely passive role in the action is filled by the *patient*, who may also have preferences

¹²This is precisely the issue around political correctness, which is itself pejorative (i.e. value-negative): choice of words matters.

attributed to him, but doesn't take a decision to play the passive role: the patient of a murder for instance generally prefers not to be murdered. For planning this is less relevant, but for the legislator this can play a justifying role for certain normative rules. Note that an agent does not need to have explicit expectations about the preferences of others: these are already apparent from the agent role attributed to them.

The action, agent role, and rule are reified and described by OWL concepts. The *script* as intended by Schank (cf. [246]) is a theory, a set of logical propositions connecting action, the agent roles involved in their performance, and the preferences attributed to those agent roles. The process of recognizing the execution of a script is however an ampliative activity, and involves default rules. In the context of law this was explained in terms of constitutiveness.

Some questions remain open after consultation of the relevant literature (here i.a. [138, 241, 267, 100, 102, 101]) on preferences: the nature of the relation between preferences of persons and those of agents, and the nature of the relation between agents/persons and preferences. No name is proposed for the property that links preferences and agents/persons anywhere: the agents simply *have* them. In addition, not everyone will agree that the script attributes *preferences* to the agent: in the context of a specific execution of a script, it seems more appropriate to speak of plans, of goals, of reasons, of intentions, etc. In my view these are more specific roles or epistemic derivatives of the preference, and to fix the operational semantics of these terms we would have to commit to a more specific theory of planning than the one proposed in section 4.3.

As Girard eloquently pointed out (cf. [176] about the legacy of the French historian and philosopher, and [118] for a discussion of empirical evidence for *mimetic theory*), *imitation* is a driving force behind human behaviour: we can therefore expect to share a major part of our preferences because we acquired them from each other. Presumably this is why we can discuss behaviour in terms of prototypical scripts, and recognize the intentions of others.

One potential misunderstanding should however be cleared up. Holding the same preferences is hardly a reason to be natural allies: competition for scarce resources is in fact most easily explained by the fact that people have the *same* preferences. People compete for the same things because they value them equally, and usually for exactly the same reasons. When two drivers collide at an intersection, they do so because they *both* generally prefer not stopping over stopping. When a driver and co-driver disagree on what is the appropriate speed to drive on the highway, they would seem to exhibit different preferences. Whether we have the same or different preferences is not a guide to whether we are likely to get into conflict. Having the same preferences does not make us cooperative: some things are simply better achieved by cooperating and others better by competition.

More relevant is the distinction between I-preferences and we-preferences, akin to the I-intentions and we-intentions distinction briefly mentioned in section 4.2 (which apply to specific actions, and therefore the actor in them, instead of to agents). In [243] we find the same distinction as private preferences vs. communal preferences. The I/we distinction applies to the object of the preference – the agent whose action is evaluated – and not to the holder of (believer in) the preference: I don't consider

whether “groups” can be said to have a preference; What matters is whether they can assume an agent role.

In case of scarcity it is more realistic to have an I-preference *I prefer that I take the best seat* over a we-preference *I prefer that we take the best seat*, for instance. In cases where we are all better off by acting with the same intention the we-preference *I prefer that we all yield for traffic from the right* will generally win from the I preference *I prefer that I have the right of way* since it is clearly more reasonable.

In [243] some additional interesting classifications are found that determine whether social coordination and publicly announced rules are likely to arise or not. [243] also points out that communal preferences in themselves are also not necessarily cooperative in nature.

Sharing the same set of preferences, regardless of whether they are we-preferences or I-preferences, can however go a long way in practical planning. The representation of we-preferences of course doesn't present us with any problems: these are simply the norms or preferences attached to roles in certain actions, and the preferences relate to the action executed in that role.

These can of course be restricted to specific populations as the players of roles: one might for instance have certain scripts in mind regarding purchases over the Internet, but I may distinguish between buying from Dutch web shops – for which I have more elaborate scripts about the way faulty products, or products that don't meet my expectations, are handled – and buying from other web shops¹³.

Private preferences, or *subjective norms*, are default rules indexed to an individual, for instance: (DEFAULT r (known.BUYER \sqcap { $i_{AlexanderBoer}$ })) . . .). These individuals may be the *self*, other actually existing ones, or prototypical ones¹⁴. The case for attaching a special status to the self, with a special i_{self} constant, is interesting (cf. generally [181]), and sometimes pragmatically necessary, although such an identification is of course rather context-specific for a knowledge representation.

How the notion of a script is worked out in a knowledge representation, based on the represented legal normative rules, and subjective and communal normative rules of agents, and background knowledge on actions, in practice often depends on the context of use. In an administrative setting the legal rules will be seen as a set of constraints on a *business process* to be designed. For a planning system scripts are often interpreted as plan operators, and for a design system they are interpreted as design operators. Any attempt at standardization in this direction (e.g. *STRIPS*, *ADL*, *PDDL*, cf. generally [240]) is inextricably linked to a certain type of planning or design algorithm. In case law they are the thing an interesting set of related case positions deviate from ([79, 80, 9] for instance discuss findings with the example script of chasing wild animals – hunting game, fishing – and case positions involving interference with it). In comparative law we may also see the case positions, or interests of stakeholders that are being protected by legislation.

¹³Only 16% of the Dutch, and 7% of Europeans, bought a product abroad in the last year; 62% of the Dutch bought a product from a Dutch web shop (*Press release of EU commissioner Kuneva of consumer affairs, June 20, 2008*). A compelling explanation is unfamiliarity with the exercise of consumer rights abroad.

¹⁴See *prototype* in section 3.4.3.

The various representations encountered are motivated more by pragmatic constraints than by a coherent view of scripts as mental objects. This work does not significantly improve on alternative representations by proposing a representation of these preferences. The preference itself can also be rightly criticized for being cognitively not very plausible. The relation between legal rules and the scripts they invoke is however central to understanding the positional evaluation in terms of for instance Hohfeldian categories, adjudication of cases and use of case law, discussed in section 6.7.2, and comparative law, discussed in section 6.7.3.

6.7.2. Case Law

When we think of case law, we generally think of court cases that involve a serious conflict between a plaintiff and a defendant, in most cases decided either for the plaintiff if some future or past behaviour of the defendant is determined to be disallowed, or for the defendant if the behaviour was not determined to be disallowed. What legal consequences this decision has depends on the type of legal procedure (tort, criminal, administrative, etc). Section 4.8 pointed out that the disallowed-decision of the court normally does nothing more than granting a potestative right to the plaintiff or a third party, generally to do something unpleasant to the defendant.

When considering the daily activities of the court, it is however important to make a distinction between the conflicts between plaintiff and defendant that are of interest here, and notarial activities, where a judge acts as a witness in a transaction between two parties that cannot be characterized as plaintiff and defendant in order to establish legal facts. Even the vast majority of conflicts between private parties handled by the courts are routine cases that can be characterized as simply denying or granting potestative right to a plaintiff: center stage takes the question whether plaintiff can make the court establish a certain legal fact that is advantageous to him.

Obviously this type of case is the most obvious candidate for extra-judicial solutions, like LKBS and mediation: the current trend is to try to move these cases out of the hands of the courts. This is not necessarily as easy as it may appear. The defendant can refuse to cooperate with any extrajudicial procedures if the potestative right to go to court remains in existence, and may do so on purpose to intimidate the plaintiff and delay the resolution of the issue in favour of the plaintiff. The plaintiff may also file a lawsuit with the mere purpose of intimidating the defendant. These scenarios are obviously not likely to make it into case law.

The activity of courts of law can be described in terms of two important principles from a knowledge engineering perspective: Firstly, the courts apply the applicable legal rules to the cases brought to them, and they are themselves bound by legal rules. Secondly, the courts are expected to be predictable, even in the absence of relevant legislation, or if the relevant legislation is ambiguous: if two cases are substantially the same, the courts should come to the same decision in both cases.

Stare decisis, Latin for ‘to stand by that which is decided’, is the general maxim used in common law that when an issue has been settled by a court decision, it forms a precedent which is not afterwards to be deviated from by other courts. The doctrine of *stare decisis* does not completely prevent re-examining and overruling prior decisions,

but it is considered (cf. generally [260]):

”a fundamental jurisprudential policy that prior applicable precedent usually must be followed even though the case, if considered anew, might be decided differently by the current justices. This policy [...] is based on the assumption that certainty, predictability and stability in the law are the major objectives of the legal system; i.e. that parties should be able to regulate their conduct and enter into relationships with reasonable assurance of the governing rules of law.” (Moradi-Shalal v. Fireman’s Fund Ins. Companies (1988) 46 Cal.3d 287, 296.)

A party arguing overruling a precedent faces a difficult task, roughly in proportion to the age of the precedent, the extent of society’s reliance on it, and its conceptual and functional similarity to other related rules of law.

A very similar notion in civil law is found in *jurisprudence constante*, which is basically a weakened formulation of stare decisis: the longer a series of identical decisions in identical cases based on the same legal rule by the courts becomes, the more determinative the decision found in the series becomes for future cases. A conceptual difference between stare decisis and *jurisprudence constante* is that *jurisprudence constante* denies the possibility of courts deciding a case based on precedent alone, without appeal to a legal rule found in legislation, and denies that the court decision potentially creates a rule.

The difference is mainly a justification issue: the precedent case can be cited as an example of why the used interpretation of a legal rule is the correct one, but citing a precedent case does not fulfill the burden of proof of backing one’s decision with a legal rule from a formal source of law (a constraint along the following lines: \exists representedBy. \exists metalex:resultOf. \exists agent.FORMALLEGISLATOR).

It is not hard to see how social mechanisms reinforce the stare decisis/*jurisprudence constante* doctrine: 1) Courts naturally don’t like to be overruled by higher courts, 2) legal professionals are naturally interested in discovering patterns in what the courts do, and in anticipating future court decisions with these patterns, and 3) society at large likes the application of the law to be predictable. The design rationale for legal rules (see section 4.1) is itself based on the stare decisis argument.

One could claim that the stare decisis notion has somewhat tautological character: cases to be decided obviously never are the same; That two cases are the same is as much a result of the choice to come to the same decision as the other way around. If one wants to come to a different decision, doesn’t one simply point to a difference between the present case and the previous one?

But things aren’t that simple: new facts of the old case can no longer be established to compare them to the new case, whereas determining new facts of the new case is still possible. It is not possible to determine whether the new case deviates from the old case by considering an additional fact that is not in the *case description* of the precedent case. It is therefore the case description that binds future court decisions, and not the case. The link between case description and decision is a categorical one, and the precedent case therefore definitely postulates a *rule*, regardless of whether we are allowed to call it a *legal rule*.

It is perfectly reasonable to represent it as such, and in exactly the same way as other, legal, rules. We do however have to keep in mind that the rule postulated is the

totality of the mapping from case description to outcome: there is no doctrine *stare dictis* (to stand by or keep to what was said) or *stare rationibus decidendi* (to keep to the rationes decidendi – the justification – of past cases). A court decision postulates just one rule: the immediate generalization of the case and its outcome.

A useful framework for relating legal rules expressed by legislation to rules in case law is the framework for case law theory construction developed in [18, 19, 79, 80, 9]. In these publications, Trevor Bench-Capon, Giovanni Sartor, and others, introduce a framework for relating case law, which describes case law theory in terms of 1) cases and the outcomes of these cases, 2) factors in terms of which the cases are described, 3) abductively formulated rules linking sets of factors to (postdicted outcomes), and 4) values promoted or demoted by having the presence of a factor weigh in favour or against deciding for the plaintiff. A theory is understood to comprise:

1. a selection of relevant cases from the case background;
2. a selection of factors from the factor background;
3. a set of rules linking sets of factors to outcomes;
4. a set of preferences amongst rules; and
5. and a set of preferences over values promoted by the factors within the rules.

Figure 6.2 explains how cases, factors, and values relate to each other. Note that I use the vocabulary of *revealing* and *motivating* from sections 6.7 and 2.4 here to relate the layers. In the original figure in [9] a distinction is made between *explaining* the case background, and *determining* the new cases. On an operational semantic level both relations between factor-set preferences and case decisions are of the same kind: the preferences between sets of factors motivate the choices, and the background of choices from the past reveals the preferences between sets of factors.

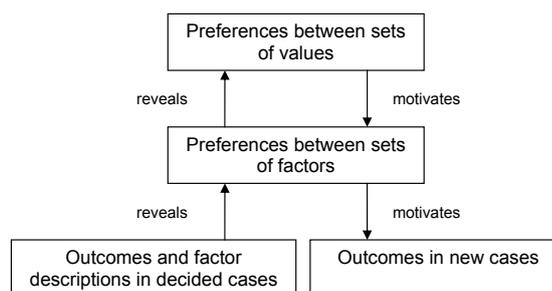


Figure 6.2. Construction and use of case law theory, adapted from [9].

The theory is developed in [18, 19], backed by the results reported in [79, 80], and finally consolidated in [9]. In [80] a tool, named AGATHA (Argument Agent for Theory Automation), is described that generates case law theories from a seed case and a body of relevant precedent cases. It is encouraging that AGATHA in a well-understood domain produces better theories than earlier handcrafted theories (viz. [79]; cf. [80]).

The *factors* are the propositions of interest that indicate a relevant brute or institutional fact, the values are associated with norms of analysis in this work, and the rules are the legal rules we find in legislation.

The outcome is – as is appropriate for tort law – phrased in terms of *deciding for the plaintiff* or *deciding for the defendant*. This is a generalization of the decision alternatives: one of the possibilities is that the case brought to the court is unlawful if decided for the plaintiff. The vocabulary – deciding for plaintiff or defendant – applies equally to criminal law, but seems out of place when the plaintiff is a public prosecutor. The rules linking sets of factors – to be read conjunctively – to outcomes in favour of the plaintiff or defendant, are simply datalog rules linking a *generic case* in terms of [266] to an outcome, for instance one of *allowed*, *disallowed*, *silent*, in which case they would be easily recast as normative rules. Bench-Capons *rule* is the same type of entity, with the same characteristics, as the indicative legal rule in this work. To decide for the plaintiff is to establish the legal fact he proposed.

Factor background information is represented as tuples (f, o, v) , where f is a factor, outcome-favoured o is one of {PLAINTIFF, DEFENDANT}, and v is the value promoted by that outcome. An outcome favouring the plaintiff usually maps to *disallowed* if the main issue to be settled is normative, while an outcome favouring the defendant should map to either *allowed* or *silent* if the rules that the court is supposed to apply do not apply to the case, or the burden of proof for them is not met.

The distinction between the approval or the silence of the law as we find it in [266, 189, 172] is not readily apparent in case law: one could argue that there is a difference in cases decided for the defendant between court decisions that posit positive *allowed* rules to decide for the defendant that have an influence on later decisions and *silent* court decisions that merely note that any grounds for deciding for the plaintiff are simply absent. There are good explanations of why the possibility of a *silent* court decision attracts less attention in private law than in areas like criminal law, where potential murderers sometimes walk out of court because of the law's silence:

1. since the conflict is between two private parties, and involves considerable investment and risk, it is reasonable to expect that the conflict only ended up in court because the involved parties genuinely *disagree* on the outcome, which suggests that there are at least some rules favouring the plaintiff;
2. court decisions that are *silent* rarely have an effect on jurisprudence and are generally less likely to end up in case law studied by legal professionals; and
3. there is no good rule for distinguishing silent decisions from allowed decisions, both against the plaintiff.

Bench-Capon prefers to relate values to single factors only, and not to sets of factors. This choice appears to be motivated by an *ontological commitment*, being that factors are assumed to be naturally *evaluative* concepts, inextricably linked to a certain value (cf. [235]). Factors are similar in function to the relevant *contrasts* of psychological geometry introduced in 6.1.1: the axis of reference is the norm or the value. The evaluative reading applies to any fact, not only those involved in deciding whether something is allowed or disallowed. Any factor that is cited in favour of one side argues that there is a reason to prefer the case of that side, i.e. the conclusion

it is arguing for. In other words, if $\alpha \prec_a \neg\alpha$ is a motivating preference in a decision, then α is a relevant factor.

This choice affects the granularity of modeling. In traffic, for instance, we can comfortably link a legally relevant phrase like *defendant was making a special manoeuvre* to a value in the right context, but not a more neutral one that doesn't suggest evaluative context like *defendant was making a left turn*. By doing this, the conceptualization does not take full advantage of things like *exceptions* on the level of factors: the presence of one factor could be entailed by the presence of another factor that promotes another value and points in the direction of an opposite outcome. The relation between both factors can of course be found by induction from the cases, but the information that there are logical subsumption relationships between the factors is not part of the case law theory. The framework on the other hand does not prohibit expansion with a logical level at a lower level of granularity to explain subsumption relations or possible overlap between the factors, just like the separation of the allowed and disallowed case of a normative rule is only a start to more complicated modeling of the normatively qualified behaviours.

The courts have some freedom in filling in the gaps left open by the legislator with the rules they create. Clearly the courts have the freedom to apply their own theories of brute reality to a case: the legislator does not supply such a theory. One could therefore say that they are free to create new elaborations of indicative rules and requirements, which, if they have an effect as precedent, will function in the future as affirmative rules affirming a more general rule in formal law (cf. *affirmation* in section 6.3).

They however also have a limited power to change institutional reality: since the recognition of conflict and the solution of conflicts by way of *lex specialis* also depend on one's theory of brute reality – it is only relative to such a theory that one can decide that one constellation of facts is more specific than another one (cf. section 6.3.1) – the court's decisions, if they have an effect as precedent, even create new *choice rules* (cf. section 6.6.3). The court can obtain the freedom to *not* apply a legal rule by constructing a case for a conflict or dilemma that gives him the freedom to choose between that rule and another one.

What the courts presumably do when they use this space left open by the legislator – at least when they are not being “activists” – is simply connecting the dots: they ascribe intentions to the legislator, and to the previous courts that created important case law, and try to infer the intended normative order and decide in accordance with it.

The courts in this way become a motivator for legislative change. New legislation does not come out of a vacuum: there are reasons to draft new legislation, and one of these reasons is to periodically *clean up* legislation¹⁵ by codifying developments in case law and other legislation, for instance from a higher legislator (the European Commission to name one), that have come to be considered *opinio iuris*. The need to do so is obviously greater in civil law jurisdictions than in common law jurisdictions, where case law is permitted to take the center stage.

¹⁵In the Netherlands this is generally called a *veegwet* (wipe act), literally intended to dust off an old act by codifying relevant *opinio iuris* into it.

If the legislator disagrees with the intentions ascribed to him, he must explicitly deny the existence of rules as articulated in case law, which will lead to increasing elaboration of legal rules over time. To block application of undesirable jurisprudence rules, the legislator has to define new *requirements* or *applicability rules*. There is no obvious formal method for the legislator to repeal case law, although the intent to block the application of certain case law can be made explicit in legislation or legislative documents accompanying and explaining the intent of legislation. MetaLex at present doesn't cover this eventuality.

Case law can be largely treated in analogy with formal legislation, including the application of the choice principles of section 6.6, but the analogy breaks down when we consider versioning. The legislator usually takes care to restrict the applicability of rules to a specific time interval. Case law can become irrelevant over time because it was 1) overruled by newer and superior case law, 2) rendered inapplicable by additional requirements or applicability rules in formal law, or 3) because it was absorbed into formal law.

6.7.3. Legislating and Comparative Law

A last context to be discussed in this book where intended normative order in relation to scripts plays an important role, is the activity of comparing alternative methods for creating an intended normative order itself.

Comparative argumentation about the law is the province of the academic field of comparative law, but it is important to realize that comparative arguments about legislation are not only made by academics. These arguments are also commonplace in macroeconomics, in politics, in legislative drafting, and finally in the court room in the choice between individual legal rules. It is this use of comparative argument that is relevant to legal knowledge engineering. Comparative arguments are made wherever choices between rules or sets of rules are made, but often they are not recognized as such.

There are three common types of comparison of legislation (cf. [42, 50]):

Comparison of alternatives Proposals for legislation addressing the same problem are compared to judge which one is better according to preconceived norms of analysis;

Vertical comparison in time Versions of the same legislation in time are compared to determine the effects (costs and benefits) of changes of legislation on behaviour, products, etc;

Horizontal comparison between jurisdictions Legislation addressing 'similar' things in different jurisdictions is compared to inform others about the effects (costs and benefits) of moving themselves, their property, products, or services over the borders of a jurisdiction.

The legislative design problem, at least the part that involves constructing and choosing the preferred candidate solution, is itself a reformulation of a comparative law problem.

Montesquieu is regarded as the father of comparative law. His comparative approach is obvious in the following excerpt from chapter XI (entitled *How to compare two*

different Systems of Laws) of book XXIX of *De l'esprit des lois* (viz. [91]):

To determine which of those systems [i.e. the French and English systems] is most agreeable to reason, we must take them each as a whole and compare them in their entirety.

Comparative research in Law usually claims to compare coherent legal *regimes*, including:

- The aspects of social life that law proposes to regulate;
- The extent to which law actually regulates social conduct;
- The attitudes about the proper structure and operation of a legal system; and
- The dedicated legal organizational structures, agents, and processes that exist for legislating, execution of the law, and adjudication.

When we however look at the practical *functions* of comparisons (cf. for instance [174]) – when they are not academic – then we see that different variants of sources of law are indeed the candidates considered and chosen between. A harmonization effort in law, for instance, usually involves a number of comparisons between regimes, with different purposes: They are compared to find a new *common core*; To separate the good from the bad legislation; To repair the latter. Relevant external legislation is involved in legislative drafting to establish its suitability for a *legal transplant* to another jurisdiction. Surely all of these things would not be possible or useful if the legal regime does not in the end boil down to the sources of law: the rules are the main objects being harmonized or transplanted. The observation that a thing that works in one jurisdiction is not at all guaranteed to work in another one is undoubtedly true, but of limited usefulness.

Besides that the legal regime also remains largely the same in some scenarios. If one compares alternative legislative changes to a new situation, the old situation that is being changed is always the same. A migratory regime is designed and the costs of migration from the old to the new situation are also evaluated, to take into account the costs of migration, and so that the individual and collective costs of migration can be minimized and fairly distributed.

It is realistic to compare sources of law of similar societies, certainly in areas like taxation where uniformity of approach has had a clear and obvious economic value for a long time, and a shared vocabulary already *de facto* exists (cf. for instance [106]).

In the last decades the interest in the problem of comparing and harmonizing legislation has been steadily increasing. One reason is the increasing legal convergence between governments in the European Union, and the increasing traffic of people over borders of jurisdictions. This leads to an increased need of administrations to know and understand legislation of friendly governments to be able to assist citizens and reduce negative consequences of movement.

Another reason is the increasing globalization of companies; Products and services are offered in many jurisdictions at the same time, and the product or service has to meet the provisions of all jurisdictions in which it is offered. Different regulations can also lead to differences in competitiveness for the product or service. For a financial product, for instance, it is considered important to qualify for tax deductions that make the product more attractive. For a medical product it is important to know whether it can be sold over the counter without a prescription.

At the same time the Internet also goes through a revolutionary growth phase, and there is an increasing interest in the use of information systems to make the problems created by increasing legal convergence and economic globalization manageable (cf. generally [42]).

The discovery of the requirements of comparing legal systems and sources of law has a direct relevance to theory in legal knowledge engineering: it has the potential to create a more principled, more robust, and language-independent methodology for modeling legislation, because it decouples the model of legislation from the purpose for which it was modeled and makes assumptions about the surrounding context, and knowledge expected from users, explicit. Modeling only one set of regulations for a specific customer often fails to achieve this level of context-independence.

Many recent initiatives have recognized the problem of mapping legal vocabularies to each other and proposed standard ontologies of international legal vocabulary (see [42] and generally [112, 22] for recent references in computer science, or for instance UNIDROIT¹⁶ and Common Core; cf. [70]). The approach of integrating two ontologies representing two different versions of a source of law has been tried by [115, 116], under the heading of *norms dynamics*. Most projects start from the assumption that a taxonomy of legal concepts is the obvious way to explain the relations between concepts from multiple jurisdictions. The problem of comparing legal systems is thus reduced to an *ontology integration* problem; Multiple jurisdiction-specific ontologies must be integrated into one international ontology. General approaches and tools for integration of ontologies are discussed in for instance [226, 173, 210, 73].

There are several significant problems with this approach, which limits its usefulness. Firstly, observe that this book only classifies a minor part of the information contained in the sources of law as *legal ontology*. The subsumption problem for legal rules is still an open problem, as section 6.6 concluded. Even the problem of legal translation poses significant problems (cf. [42] for a worked out example): institutional realities exist *sui generis*¹⁷, and one cannot just claim that *public body* in the UK translates to *bestuursorgaan* in the Netherlands, or the UK's *theft* to Dutch *diefstal* or German *diebstahl*. In the case of theft the ontological mismatch is especially clear: theft is an act of appropriation, not necessarily by taking, while *diefstal* and *diebstahl* are acts of taking with the intent to appropriate, but a successful appropriation is not necessary.

One term may stand in for the other as a synonym in a certain context, but this doesn't make them the same concept¹⁸, as section 2.3 explained.

A recent comparative law view on why legal concepts like *theft* and *diefstal* may be considered to be sufficiently similar to function as a target for comparison is found in [273].

Most importantly, however, the ontology integration view does not account for the purpose of the comparison in a satisfactory way, and that explains why publications

¹⁶<http://www.unidroit.org>

¹⁷In a category of their own; As the only instance of their class.

¹⁸Look for instance into 7246/2003 – C5-0165/2003 – 2003/0811(CNS) for a total of 23 proposed amendments by the Hellenic Republic to clarify the meaning of the phrase *same criminal offense* in the context of application of *ne bis in idem*.

in the field of comparative law do not seem to be related to ontology integration at all. If comparative argument appears related to ontology integration, it is merely because the ontology happens to describe the relevant *factors* (cf. section 6.7.2) or *evaluative concepts*.

Comparative arguments are based on rhetorical comparison between two targets with respect to an axis of reference (cf. section 6.1.1), with the aim of uncovering interesting similarities and differences. This axis of reference is a *norm of analysis* (cf. [145, 146]), and is yet another norm which is based on the *subjunctive betterness* pattern, just like the legal norm, the communal norm, and the subjective norm.

To make clear what norms of analysis are, and how they are used, I will give some informal examples from the field of taxation, in particular taxation of capital income. Capital income can be taxed based on the actual (*ex post*) income data or the predicted (*ex ante*) income data. The Dutch *predicted capital yield tax* introduced in 2001 is an example of an *ex ante* tax on expected yield. Only a fictional risk free yield is taxed; additional gain or loss attributed to specific skills, labor, or information advantage is not taxed. The *capital gain tax* (where gain is taxed at the moment of realization of income) and the *capital growth tax* (where apparent capital gain is taxed at fixed intervals based on an estimate of what income *could* be realized from capital regardless of realization of income) are *ex post* taxes on capital gain. These tax risk free yield (the compensation for delayed consumption) as well as additional gain or loss attributed to specific skills, labor, or information advantage.

The following is a comparative argument based on a norm of analysis:

The consequence of the possibility of delaying capital gain taxation is that capital growth is taxed with varying rates depending on how long investors can afford to delay realization of income. This is inconsistent with horizontal equity.” (Cnossen & Bovenberg in [83])

The norm applied is known in the literature as the *horizontal equity* norm (those with equal ability to pay taxes should be assessed equal tax payments; cf. [252]). The comparatum is implicit: capital yield and capital growth are more equitable than capital gain tax. The author states that it is “inconsistent” with the principle, but that observation is only relevant because of the alternatives that *are* better implementations of horizontal equity.

Another example:

It is beyond dispute: the taxpayer could easily realize the predicted capital yield if he would invest his capital in risk-free investments. But what if he does not? Can we ascribe income to that taxpayer? An analogy from “human capital”: A taxpayer works 20 hours a week, and earns 30,000. The taxpayer could easily earn 60,000 for 40 hours a week. But what if he does not? Can we ascribe that income to that taxpayer and tax it?” (Vording in [83])

The comparatum is again missing: capital gain and capital growth tax are more effective than capital yield tax for some reason. The implicit norm of analysis is supplied by Cnossen & Bovenberg (also in [83]): “The principle of realization of income is based on the observation that tax should not have to be paid if the means are not actually available.”

The analogy with labour strengthens the argument in a slightly deceptive way: The difference between working 20 hours per week or *easily* working 40 hours a week

is hardly the same as the difference between choosing risk-free and risk-prone investments.

Average capital growth is closely related to the amount of risk associated with an investment. Not realizing the risk-free capital yield is usually a consequence of risk-prone investment strategies. Government may not wish to share in the fortunes of investors, or may not wish to *subsidize* risk-prone behaviour.

Additional such arguments are easy to construct: for instance, a fictional yield and an actual yield are easier to tax than an estimated one, because less information is required to determine it, a fictional yield results in a more predictable government budget than the other ones, etc. In all these examples the relevant difference is a rhetorical comparison relative to an axis of reference.

Rhetorical arguments for similarity were less easy to find in [83], but the field of law certainly asks for them, as the following examples show:

1. Art. 38(1)(c) of the *Statute of the International Court of Justice* directs the court to apply “the general principles of law recognized by civilized nations.” This not only entails discovering similarities between legal systems, but also confronts us with the problem of deciding which of its members the UN considers civilized nations and which not.
2. Art. 215 of the *Treaty of Rome* establishing the EEC provides that the non-contractual liability of the Community is to be governed by “the general principles common to the laws of the Member States.” Note that this formulation appears to make the intended similarities a moving target, dependent of when which member state acceded to membership of the EEC. The assumption is that the intended similarities are limited in number, and widely shared in the international community.

The question is who gets to decide whether things are sufficiently similar. In international law it is easier to define the norms that should be shared by member countries than to decide who is the authority that determines whether a member state complies with the defined norms. In harmonization and standardization the problem is which system moves to which other system. Harmonization and standardization are usually relative to another system that specifies common aims. In the EU we see this method in action: the commission defines the guidelines that member states must be consistent with, and it also defines the yardsticks used for determining whether a member state complies with the guidelines.

Interesting differences and interesting similarities are interesting because they indicate a positional difference or similarity in evaluation relative to a norm of analysis.

If the function of law is reduced to normative rules and the *allowed–disallowed* axis, then the problem of comparing two different normative orders would conceptually be relatively clear, although not necessarily easier to realize. In section 6.7 the idea of *envisioning* the normative order was introduced: what categories of cases does a normative system distinguish and how does it order them? The evaluative differences are found in pairs ordered in opposition by the two normative systems. This of course depends on being able to align the ontologies underlying the two normative systems, so that subsumption relations between categories of cases can be computed, and this

will be considerably easier for developments over time in the same jurisdiction, than between jurisdictions.

Institutional ontologies cannot in principle be aligned, just subsumed under the same core ontology, but brute reality in some domains (for instance traffic, taxation, consumer law, etc.) presumably can, certainly in very similar economies and cultures with shared values and concepts. This however creates the modeling problem of *grounding* institutional reality in a model of the legally relevant consensus reality (cf. [94, 63, 92, 93, 265]).

Section 6.3.1 however argued that specificity on a functional level is not the same as ontological subsumption if we combine ontological subsumption with ampliative reasoning. The law is organized around the behavioural *scripts* it attempts to influence: specificity in the context of normative conflicts, and “interesting similarity” in comparative law are not merely found in subsumptive relations between scripts. Some things do not subsume each other but are nevertheless considered sufficiently similar if a certain *abstraction* would be applied to them.

The comparison reveals differences and similarities relative to the evaluation dimension created by legal norms. Some similarities and differences are however more interesting than others: not all of them are worth reporting. Whether they are worth reporting depends on norms of analysis: an evaluative difference is interesting if the norm of analysis applies to it. If the norm of analysis postulates $| C | \prec_{na} | C' |$ while jurisdiction j_1 confirms it, $| C | \prec_{j_1} | C' |$, and j_2 is contrary to it, $| C | \succ_{j_2} | C' |$, the difference between j_1 and j_2 is interesting.

What is in effect compared in a rhetorical comparison is one script relative to two legal systems, with the same set of subjective and communal norms in the background. The fiscal examples above for instance relate mostly to investing: there is a difference between money in one’s hand – tax is to be paid in the form of money – and the estimated monetary value of capital investments in shares, bonds, real estate, etc. One can move money into investments, producing a capital yield, or move it out and keep it at hand. There are tactical considerations connected to the timing of moving it in or out, and it is for instance undesirable if one must sell a house at a certain point in time to pay taxes over its possession.

Regardless of whether the context is comparing alternative sets of rules during legislating, or comparing two different solutions in different jurisdictions, the differences are analyzed from the point of view of certain *stakeholders* and their (subjective and communal) norms (see section 4.5). These stakeholders are not persons but agent roles that participate in certain salient scripts. The normative systems are analyzed by looking at the choices they create in prototypical executions of a script, and whose interests are being served by those executions. The resulting arguments – *this agent role in this setting has a better alternative available relative to n_a in j_1 than in j_2* – are then in some cases generalized to populations – *up to x people may be affected negatively by adopting the rules of j_2* .

The question is where one gets these scripts. In the case of fiscal law these will be supplied by economists, together with their norms of analysis. In many other cases it is obvious to resort to case law.

Court decisions are phrased in terms of factors relevant to the system in which the

decision was taken. The court decision will not contain the information necessary to decide whether the factors relevant to the decision in the other system are present or absent. Using case law from legal system j_1 to explain differences with legal system j_2 results in a representation bias towards system j_1 . The differences that will be found can only be found relative to the contrasts that j_1 creates but j_2 doesn't.

Using case law from the old system to predict the performance of the new system is equally biased. One can only verify that it improves on the old system with respect to important contrasts relevant to the old system.

The same applies to a "migration of persons, goods, and services" scenario. In section 6.7.1 the example of web shopping was discussed. Many people may avoid purchases in other jurisdictions because of a perceived lack of legal protection. EU harmonization of the degree of protection only has a limited effect on people's attitudes, however, since what people note is first and foremost that the relevant scripts that they know about do not apply abroad. The difference in price hardly justifies having to learn something new.

In some cases the problem of comparing normative orders is contextualized to a clear setting that supplies the script. The growing interest in internet portals and web services is a good example: comparing for instance value added tax regimes for the purpose of creating a single calculation web service for different jurisdictions or an integrated portal for spatial planning restrictions from different jurisdictions (cf. [40, 45, 46, 41]) evokes clear scripts. The LKBS of course also embodies a set of scripts.

In this book separating the legal rules from the various assumptions about the quality of the available knowledge, the conceivable alternatives, the preferences of the user, etc., is a reusability issue for the legal rules. At the same time these other legally relevant knowledge sources also are also components of a reusable script when made explicit.

The conception of comparative argument as starting from a selected set of scripts is uncommon in legal knowledge engineering. With the benefit of hindsight the evaluative perspective of [42] *is* the script to be evaluated against. As noted in section 6.7.1, the script, as a typical mental concept, is in knowledge engineering represented by an abstract object whose structure is motivated by practical software engineering concerns, often without a genuine interest in the domain of cognition.

6.8. Conclusions

In chapter 5 I nearly completed my exposition on the representation of the contents of sources of law. The law consists of interrelated institutions which can be described as a system of legal rules and other legal facts. The sources of law should be considered a paper trail left by its operations, which often involve formal acts, for instance for legislating and for dissolving disputes in court.

In this chapter a few modest additions were made: section 6.2 introduced some more specifics on normative rules, and section 6.6.3 introduced a specific kind of applicability rule, the *choice rule*. The addition of these two rules finishes the construction made in this book of the problem-setting-independent and, in this sense, reusable knowledge of sources of law.

To explain the use of the choice rule I had to introduce *normative conflict*, which explains why choice rules are sometimes introduced for pairs of normative rules. The applicability of the choice rule is however not dependent on the presence of a normative conflict. Neither is normative conflict a phenomenon that arises from legal rules as such. Normative rules are defeasible by applicability rules and requirements posing additional constraints on the cases qualified by the normative rule; Since the normative rule only derives *its* violation, normative rules are not normally defeasible relative to other normative rules.

The issue of normative conflict lies beyond the institutional interpretation of legal rules; To explain it, one needs to appeal to a normative order intended by the legislator and show that the normative rules do not result in it for a specific category of cases.

Both the issue of normative conflict and another phenomenon of interest in legal knowledge representation, the contrary-to-duty norm, are best explained in terms of obligation or subjunctive betterness. Some *collections* of normative rules – think for instance of a single criminal code – are expected to result in a coherent account of normative order in terms of a *set* of propositions about a betterness relation partially ordering alternatives. If this account is inconsistent, it exposes normative conflicts. If the explanation of a normative order as a single set of equally important propositions fails for the analyzed collection, an alternative conceptualization of the collection must be found. In literature we for instance find the notion of a partial order on normative rules, but this account is not entirely satisfactory in cases of *intersection conflict*.

Normative conflicts are eventually resolved by explicitly formulated choice rules formulated for the occasion in which the normative conflict arises. The courts are the legitimate authors of choice rules. The mechanism by which a court arrives at a relevant choice rule for a normative conflict is not one that should be integrated into the OWL DL representation of a source of law, for the simple reason that the explanation of the normative conflict may appeal to making a closed world assumption about the set of alternatives available to an agent: the court only considers conceivable, i.e. known, alternatives.

Since the legislator is neither omnipotent nor omniscient in his abilities to bring about normative order in brute reality, his attempts to do so will inevitably fail occasionally. *Intentions* have a habit of coming to the foreground when attempts and failures matter: courts appeal to the intentions of the legislator to explain how legislation should be applied. Perceived failures of the legislator – commonly in the form of perceived normative conflict – create the space for courts to get involved in creating normative order, by formulating choice rules, which are therefore often found in case law.

Normative order obviously plays a broader role in problem solving than merely resolving normative conflict. Resolving normative conflict is not its primary function.

The rules of the institution have an important, instrumental function for those that engage in social interactions governed by those rules that goes beyond their interpretation as institutional qualifications of behaviour. Reciprocal recognition of actions by persons depends on the adoption and attribution of agent roles. The legal institution creates and refines such agent roles through its rules, *and* gives people reasons to adopt and to attribute them.

These agent roles play a central role in planning and plan recognition for the addressees of the law and the people that interact with them. All legal rules tend to have this effect, if they are known and understood.

Normative rules certainly give people a reason not to perform some acts, and a reason to expect others to do so, but the importance of normative rules as a guiding force in behaviour is perhaps overrated in computer science & law. The reason for this attention to normative rules is a simple and obvious one: the account of normative order in terms of obligation is relatively well-understood and obviously fits well on sets of normative rules, and normative rules are a characteristic feature of legal institutions. The normative order, as perceived by its participants, is however not an account of a system of normative rules in terms of obligation.

The Hohfeldian relations introduced in chapter 3 offer a richer vocabulary, by also distinguishing relevant relations based on the *power* to create certain legal facts, but it is only the fact that another agent takes an interest – exhibits preferences towards them that are not directly based in normativity – in these legal facts, that makes the agent recognize a legal relation.

I use an account based on a betterness relation to explain normative conflict – perhaps more accurately the *dilemma* some sets of normative rules pose in planning – within a set of normative rules presumed to express a coherent normative order, while I also suggest that the preferences attached to agent roles can be described both in terms of such a betterness relation and, if so desired, even in terms of normative rules. In this case the agent role takes the place of the norm or collection of norms.

Perhaps it seems strange to do so, but if the normative rules the legislator formalizes give rise to both a reason to make choose certain alternatives over others, and a reason to predict that others will do so, then the normative rules can be used for both a prescriptive and descriptive function.

As MacCormick pointed out (cf. [192] and example 6 in section 4.4) this is in fact the strange thing about many applications of the idea of institutions, beyond its application in law and rule-bound games: the participants in “institutions” often *cannot* in fact verbalize the rules in effect, which is a reason to doubt the very distinction between descriptive rules that merely describe regularities in behaviour and constitutive rules that bring about those regularities.

In the end these different accounts are accounts of goal-directed behaviour linked to *agent roles* assigned to a person.

The final section of this chapter returned on an application context of legal knowledge representation discussed only in the introduction and section 4.5.1.

The values the legislator tries to promote through the intended normative order unsurprisingly also take the form of propositions about a betterness relation, the *norms of analysis*. A weakness in the prevailing view of comparing legislation based on ontology integration is that it completely ignores the legislator’s intention to change an existing normative order. Comparative law in actuality compares intended and actual normative orders *in in the context of relevant social scripts*, and not only the part of the script explained by normative rules.

If knowledge engineers limit themselves to the institutional interpretation of legal rules, no meaningful comparison can be made by a KBS, because no knowledge is

available about which institutional entities and rules perform substantially the same functions in different legal institutions. The interpretation of normative rules in terms of obligation or betterness, without reference to certain specific social scripts that function as context of the comparison, will be of limited help, since, firstly, legislators use other instruments for influencing behaviour besides normative rules, and, secondly, not only the law is a reason to behave in certain ways. Moreover, intentions with respect to normative order are ascribed to the legislator: a better authoritative source than the source of law on the intentions of the legislator does not exist.