



UvA-DARE (Digital Academic Repository)

Legal theory, sources of law and the semantic web

Boer, A.

DOI

[10.3233/978-1-60750-003-2-i](https://doi.org/10.3233/978-1-60750-003-2-i)

Publication date

2009

[Link to publication](#)

Citation for published version (APA):

Boer, A. (2009). *Legal theory, sources of law and the semantic web*. IOS Press. <https://doi.org/10.3233/978-1-60750-003-2-i>

General rights

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, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 4

Agency and the Law

4.1. Introduction

Society uses a wide variety of sources of law, and uses many labels other than legislation to describe sets of written rules: quasi-legislation; administrative rules; codes of practice; guidance; guidance notes; policy guidance; guidelines; circulars; framework documents; outline schemes, and statements of advice. A commonality of these documents is that they are understood to postulate or describe norms that guide or mandate conduct in a given type of situation. This guidance can pertain both to physical behaviour and to decision making processes.

Before one can turn to the question of how to represent legislation and similar documents, one must consider the role these documents play in guiding behaviour.

This chapter deals with the relation between these formalized rules and behaviour. It does not address the problems legislators encounter in managing a large body of such rules, the solutions they have found for these problems, and the consequences these have for knowledge engineers who try to represent the meaning of these rules in logical form: this is the subject of the chapter 5. In this chapter a number of legal theoretic devices are introduced, and positioned as practical reasoning devices in the context of *planning* and *plan recognition* problems: since this context is missing when we consider sources of law by themselves, a representation of the source of law will often omit this information. Chapter 6 reintroduces some of these notions in the context of an *intended normative order*.

One of the major objectives of this book is to consider the problem of representing legal rules while making minimal epistemic commitments to the way these legal rules are used. At the same time we concluded in chapter 2 that we have to account for two very different uses of legal rules:

Planning to perform a task one set oneself; in the case of law to bring about beneficial legal facts, while avoiding detrimental ones; and

Situation and action recognition to infer one's own (legal) position, and what relevant (legal) facts others bring about, are going to bring about, and unsuccessfully attempted to bring about.

Both generic types of intelligent problem solving behaviour are often influenced by legal rules, but the legal rules are of course not the only source of knowledge taken into account.

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 subjects are not of specific interest to the field of law, but the realization of interesting problem solving competence will invariably involve strong commitments, and often assumptions, about the epistemic framework in which legal rules are used. Such commitments to epistemic frameworks are often deeply embedded in legal theory, and in representation techniques based on legal theory.

This chapter surveys a number of important legal theoretical concepts used in legal knowledge engineering, in an attempt to separate epistemic reflection on the use of knowledge of the rules from the reusable rules proper. The law as an *institution* is a

central concept.

A subject that is carefully avoided in most of this chapter is the notion of *reified* legal rules with their properties, and the properties of the documents that contain these rules. Instead this chapter discusses the logical form of constitutive and institutional rules.

4.1.1. The Institutional Perspective

This book takes law to be an institution whose primary purpose is to create normative order by way of formalization – the sources of law specify a *formal, institutionalized normative order* (cf. [192]) – usually wherever spontaneously arising, informal normative order fails to achieve the desired results.

The following rationales are usually associated with the formalization, i.e. the writing down, of rules:

Accumulating Knowledge Accumulated knowledge from conduct in the past is written down in the form of rules to guide conduct in the future. The rules set out the relevant criteria, guide collection of information, decrease the amount of mistakes, and generally allow relatively stupid people to solve complex social problems.

Consistency Writing down rules supposedly encourages consistency, fairness, equality of treatment of persons, groups, and organizations in different places and at different times. rules limit the discretion of the decision maker in treating specific persons, groups, and organizations differently, and therefore reduces bias and corruption. Written rules also make behaviour of others predictable, which reduces conflicts.

Democracy The process of writing down rules allows for greater public involvement than the mere making of a decision. The use of rules is in this sense a precondition for effective democracy on a large scale.

Legitimacy Written rules contribute to the perceived legitimacy of decisions, because of the reasons above, and allow the decision-maker to cite the source to justify the decision.

The rules create normative order, but they are not norms.

Much of the terminology to describe institutions has been strongly influenced by the work of Searle (cf. generally [249, 248]), although the notion of institution and the closely related notion of constitutive rules were introduced earlier. Searle describes the distinction between two types of rule, being *normative* – or regulative – and *constitutive* rules, as follows:

Some rules regulate antecedently existing forms of behaviour. For example, the rules of polite table behaviour regulate eating, but eating exists independently of these rules. Some rules, on the other hand, do not merely regulate an antecedently existing activity called playing chess; they, as it were, create the possibility of or define that activity. The activity of playing chess is constituted by action in accordance with these rules. Chess has no existence apart from these rules. The *institutions* (emphasis mine) of marriage, money, and promising are like the institutions of baseball and chess in that they are systems of such constitutive rules or conventions ([249], p. 131).

This distinction has been taken up by many in computer science & law (viz. [29]), although often with a very limited application: constitutiveness is taken to be a characteristic of a limited type of legal rules or legal acts. In legal theory in general there is a bewildering array of conceptions of constitutiveness (cf. for instance [199]), some of which are irreconcilable, and this book will add yet another one.

Both the regulative and the constitutive rule are to be distinguished from the knowledge representation rules of chapter 3 that merely generalize past observations in the hope that they generalize to future ones. The regulative and the constitutive rule on the contrary only affect events that happen after the rule comes about, and application of the rules by an agent is presumably the force that makes the consequent of the rule happen. The regulative rule guides behaviour, while the constitutive rule guides interpretation of events in the terms of an *institutional reality*.

The regulative function of law is generally taken as the defining one of the legal system in positivist legal theory. This is for instance apparent in Kelsen's notion of a *Grundnorm* (cf. [172]), which is clearly of the regulative type, or in Austin's adagium (cf. [10]) – later developed by Hart (cf. [140]) – of law as the command of the sovereign backed by the *threat of punishment*. Normative rules embody the primary purpose of law.

Section 4.2 of this chapter will kick off with a discussion of constitutive rules and institutional facts, since the regulative function – even if it is in the end the really important one – is secondary to the function of constitutive rules. To describe some uses of constitutive rules we have to spell out the relation between someone's decisions and someone's behaviour. In law, this connection is made through the concepts of intention and action, discussed generally in section 4.3, and applied to law in section 4.4.

Section 4.4.1 introduces a number of systematic relations between the rules of the institution, and the types of knowledge representation rules introduced in section 3.4; This is an application of the idea of coherent ontological strata versus defeasible mappings between strata introduced in chapter 2 to institutions as coherent systems versus an uncontrollable environment.

Law *institutionalizes* normative order by way of institutionalizing recognition and evaluation of conduct. The institution and the normative order it institutionalizes are however two separate things. This point of view will be worked out further in sections 4.2, 4.5, and 4.6. Attention to only one of these two aspects, or the insistence that rules are to be *either* conceptualized from an institutional point of view or a regulative one, will lead to unbalanced knowledge representation.

4.1.2. Norms, Normality, Normative Rules, and Agent Roles

Normative rules, discussed in section 4.6, occur in three main variants in natural language: obligations, permissions, and prohibitions. The use of specific auxiliary verbs is however not at all indicative of whether some proposition is normative or not. Both a legislator and a sociologist can utter, for example, the sentence "Tanks adjacent to the hull are not used to store fuel oil", and in neither case the sentence would be semantically ambiguous. In the first case, the legislator sets forth a rule prescribing

how a tank adjacent to the hull should be used. In the second case, the sociologist verifies a regularity, analyzes the relation between the norm and social reality. A legal normative rule is normative by virtue of its function and the way it came about, not of its propositional content.

Section 4.6 takes the position that normative rules are in principle standard constitutive rules, whose conclusion – allowed or disallowed – however connects to an understanding of the world in terms of better and worse actions.

Clearly, normative rules only address conscious human choices. The eruption of a volcano may very well be undesirable, but it makes no sense to try to tell it ought to behave. It is also not very sensible to prohibit making mistakes. Normative rules are information items intended, by their author, the legislator, to influence other peoples choices. The normative rule describes a choice, and the preference the author holds for one of the involved alternatives, and it imposes that preference on the addressees by way of some social mechanism: it is this social mechanism that results in some scripted behaviours and transactions becoming a *norm*.

This social mechanism is generally speaking of relatively little importance to knowledge engineering. Section 4.5 gives this question more attention than most literature in the field, and also touches on the perspective of the legislator (who will return in section 6.7.3 later).

Rules in the sense discussed, normative or otherwise, address agents by *role*. The agents addressed by the rule are described, and not identified. The rule applies to you because you are (recognized as) a citizen, driver, legislator, civil servant, judge etc.

Not every command is a normative rule. A command that addresses a role filled by one person at the time, for instance the monarch or the prime minister, is still a normative rule. A command that directly addresses *you* cannot possibly be one.

Agent role, discussed in section 4.5.2, is the most useful way in which legislation is organized. Tax law addresses taxpayers, criminal law addresses state and citizen, private law addresses interaction between citizens, labour law addresses interaction between employer and employee, etc. Also the attribution of legal powers or competences, addressed in section 4.7.3, is mediated through the role. Reciprocal recognition of actions by persons depends on the the adoption and attribution of agent roles.

It is the understanding of situations in terms of the rules, which connects a set of preferences and expectations to persons through the agent roles attributed to them.

The deontological interpretation of normative rules in section 4.6 is based in the assumption that people avoid the risk of punishment, i.e. consider options that are allowed better than those that are disallowed. This interpretation can be used to plan your own actions, but also to predict the behaviour of others to your advantage, the theme of section 4.8. The other can also be a collective, for instance everyone, as section 4.8.3 explains.

Not only normative rules create preferences and expectations, however, as section 4.5 explains. All rules will lead to habit formation.

4.1.3. The Problem of Agency

In the world around us, we see that formalized norms are usually adopted, provided they are known. There are several explanations for violation of norms, assuming in the first place that it is based on wrong choices:

Unwillingness The agent prefers the wrong alternative. Conditioning (reward and punishment) is the universal solution for this problem, and according to some influential voices the fundamental function of law.

Ignorance The agent failed to understand the law, failed to conceptualize possible choices, or failed to foresee possible outcomes of a choice. If the agent drives on the wrong side of the road, for instance, without the intent to cause a crash, the problem is apparently that the agent failed to foresee the likely crash that would follow. Law can help solve this problem by commanding people to drive on one side – in this case the side is irrelevant – of the road.

Force Majeure In some cases an agent may be forced to choose contrary to the rules because a legal alternative was not available. This case is usually either not considered a violation, or it is a special case of violation without adverse legal consequences.

Non-compliance with the rules is not always an issue of unwillingness. As section 4.3 points out, describing the behaviour of others presumes that you understand their intentions. Since these intentions are not directly observable, but can only be guessed at based on observable aspects of behaviour and preferences and knowledge that you ascribe to the other, applying the rules to the behaviour of others is a complex reasoning activity.

Many of the common complications arising from this, and the assumptions we must minimally make about agency in order to represent normative rules and apply them in planning and action recognition, are addressed in section 4.7 and section 4.8, which discuss situations and actions, deadlines, states and change, normative positions, powers, jural relationships, and legitimate expectations.

4.2. Institutions and Rules

Institutions are, according to a common definition, “structures and mechanisms of social order and cooperation governing the behaviour of two or more individuals”, for instance law, marriage, money, democracy, the marketplace, church, school, etc.

Law is an institution whose primary purpose is to create normative order by way of formalization – i.e. a *formal, institutionalized normative order* (cf. [192]) – usually wherever spontaneously arising, informal normative order fails to achieve the desired results. A similar characterization is already found the work of Geiger, who speaks of law as a mechanism to create normative order monopolized by a central authority (cf. [119]). The notion of law as an institution, based on so-called constitutive rules, has been a productive concept in legal theory and computer Science & law, in particular to justify the use of logic programming rules to represent law (cf. generally [29, 199]).

An *institution* is a collective intentional or *social* entity, i.e. an entity that exists merely because a collective (i.e. group of natural persons) recognizes and intends its existence. The institution itself is a good example of *intention* applied to the existence of artifacts.¹

The structures of the institution are defined by the institutional facts that make up the institution, and its mechanisms of change are the constitutive rules that specify what *constitutes*, or counts as, an institutional fact. Conversely, the institutional fact has a constitution base (following Hindriks in [148]), which consists of the application of a constitutive rule to the *constituting facts*, which are *brute* facts, yielding an institutional fact.

Definition 5. A *constitutive rule* of an institution derives an institutional fact from one or more other constituting facts, at least one of which is a brute fact.

Definition 6. A *brute fact*, relative to some institution, is a fact that has no constitution base in that institution.

Brute facts are pre-existing and external to the institutional reality constituted by the rules. This does not however mean that they are in any sense “natural”, or non-institutional. The institutional facts of one institution can be the brute facts of another one. If we for instance state that checkmate constitutes winning, checkmate is an institutional fact of chess, and winning is an institutional fact of games. If we consider “winning the chess game” to be part of the institutional language of chess, the relation between checkmate and winning can no longer be considered truly constitutive, unless one expands constitutive to include any terminological axiom, which I will not do. On this point this book follows observations by Hindriks on Searle’s work (in [148]) and not Searle (cf. [249, 248]), who does insist on separating social reality from a pre-existing one. Hindriks identifies these rules as *essential rules* of the institution: the ontology of the institution. An institutional fact that merely rephrases other institutional facts using the institutional ontology shares the constitution base of the institutional facts constituted by brute facts.

Definition 7. An *institutional rule* of an institution derives an institutional fact from one or more other constituting facts, all of which are institutional facts.

The distinction between institutional and brute facts is very similar to the role of the Breuker’s *legal abstract model* (cf. [63]): a layer of “legal” concepts and relations built on top of a large layer of commonsense knowledge ([266], p.57). In Breuker’s original conception this world knowledge could ground different institutional normative orders in a single consensus reality. This is an attractive proposition for the purposes of comparison of two institutional realities that do not intersect at all, as both would presumably be grounded in the same consensus reality, giving us something concrete to compare.

The (institutional or brute) fact is simply a reified statement about something. One may think of the constitutive rule as a simple institutional qualification of something

¹See my earlier observations about the interchangeability of the planning and design problem type in section 2.4.

– the predicate or concept is institutional – or as giving rise to a new institutional thing. This is usually a question of modeling style and requirements, and not a quality inherent in the constitutive rule.

Example 3. In an auction, raising a hand constitutes a bid; This may be represented as $\text{RAISEHAND} \sqsubseteq \text{BID}$ or $\text{RAISEHAND} \sqsubseteq \exists \text{constitutes.BID}$, or a more detailed representation which takes into account the auction setting, the participation of some person in it, the timing of the raising of the hand, and – consequently – the amount of the bid, etc. We can speak of the institutional concept BID , some institutional object o , which is institutional because it is of type BID , and the institutional fact $\text{BID}(o)$.

From an ontological point of view the act of raising one’s hand and the act of bidding a certain amount cannot be one thing, because they do not share essential properties. The raising of a hand is a physical event, while the bid is an institutional, and therefore social, event.

Proposition 2. The *constitutes* (inverse: *constitutedBy*) property; Every institutional thing is constituted by a non-institutional thing:

$$\text{INSTITUTIONALTHING} \sqsubseteq \exists \text{constitutedBy.}\neg \text{INSTITUTIONALTHING}$$

Note that since this *subjective* restriction applies to one and the same institution, and not institutions in general, this is a design pattern for ontology represented in OWL rather than a terminological axiom.

The *constituting act* of raising a hand to make a bid and the supervenient *institutional act* of making a bid are in the example inextricably linked to each other. The same is the case for a constituting act of taking that constitutes a theft. The case is often less clear-cut when we are talking about states, objects, substances, etc. In many cases we see that although the constitutive rule states that some brute state constitutes some institutional state, the act to initiate or terminate that state is labeled constitutive. This is not entirely surprising, since we are dealing with *intentional* things brought about by actions.

A constitutive act in this sense is an act of representation: its *function* is standing for, or representing, something else. Functions of things are not inherent in the thing: they are assigned by (intentional) agents, and the recognition of the function depends on the observer.

This creates an ambiguity, however. We raise our hand at the auction *because* we intend to make a bid: the action has no other rationale. The thief on the other hand has no intent per se of being labeled a thief: he merely takes that risk. To resolve this ambiguity, we have to distinguish the constituting act from the constitutive act. This is the subject of the next section.

4.2.1. Constitutive Acts, Speech Acts, and Intent

In law, a legal act² is an act that creates a legal fact i.a. *because* it was intended by the actor to do so. In analogy, consider a *constitutive act* an act *intended* to constitute

²*Rechtshandeling* in Dutch doctrine, *Rechtsgeschäft* in German doctrine.

an institutional act. A constitutive rule may specify that an act (for instance raising one's hand) constitutes an institutional act *because* it was intended to do so: the constitutiveness of the act is a condition of the constitutive rule. To unambiguously create the institutional fact we have to make a demonstrable declaration of will, and, conversely, to prove the institutional fact one must provide evidence of the intent to produce an institutional fact.

Definition 8. A *constitutive act* is an act intended to constitute an institutional act.

As stressed by for instance Jaap Hage in [132] *intent* has a central role in bringing about institutional facts, but only in its relation with *action*: the institutional fact is brought about by performing a constitutive act, and – normally speaking – an action presumes intent to bring about the product of the action. To describe behaviour in terms of actions is to presume intentionality to produce the products of the action: this is why actions can fail. If a driver turns right we are talking about an action, and if the driver happens to hit a bicyclist while doing it, this is an event, and it is an (unintended) side effect of the action which is not part of the essence of the action itself. Without the action the event could not have happened, but hitting the bicyclist is not the action itself, which was aimed at turning right.

Raising one's hand at an auction constitutes a bid. Raising one's hand to a friend may however also be an attempt to draw attention, and conventionally constitutes a greeting. Raising one's hand to a friend walking in on an auction in the central lobby of a cruise ship you were watching constitutes a greeting, but may be *recognized* by the auction master as a bid, which is an unintended side effect of the action. The auction master's problem is to determine who participates in the auction, and the setting is in this case not conducive to this determination.

Legal theory has a whole battery of concepts to deal with intent, the failure to bring about what was intended, and the fact that not all effects of actions are intended ones: declaration of intent, intent and conditional intent, attempt, recklessness, negligence, mens rea, etc. Still the operative principle behind constitutive rules and institutional facts is that people to a large extent have control over what institutional facts they bring about: this is the whole point of trying to regulate behaviour by way of declaring sanctions in advance.

A *constitutive act* is an act of representation that creates some institutional thing that wasn't there before, by virtue of the intent to create the institutional product of the act alone.

Collective intentionality – Searle's original requirement for performing constitutive acts – does not necessarily presuppose that collectives *are* intentional: it is sufficient that its members can make a distinction between *I*-intentions and *we*-intentions, since the interpretation of institutional reality happens in each individual mind without recourse to a collective one. I intend to undertake some action because I expect that *we* interpret the action as representing something else. Searle specifically defends the thesis that it is unnecessary to explain how *we*-intentions reduce to *I*-intentions. This is an *internalist* point of view, and also the one we would usually take on behalf of (the user of) an expert system.

Collective acceptance, an alternative concept introduced by Hindriks (cf. [148]), derived from Searle's collective intentionality (cf. [248]), can be taken to mean actual acceptance by at least one person besides the actor. It answers a variant on an old philosophical question: when nobody's around, can your actions be truly constitutive, regardless of your *we*-intentions? According to Hindriks this is only the case if they are recorded and later accepted as constitutive by the required audience that accepts the act.

This distinguishes the constitutive act from other acts of representation inherent in speech acts: if one for instance requests someone to do something, the request represents an attempt to make someone do something, regardless of whether it is *accepted* by anyone. Similarly, a promise to do something represents an intention to do something, and only when it has a constitutive character gives rise to a duty to do that thing. Think for instance of a threat: a duty only follows from the threat if there is an audience for whom a threat has the conventional effect of producing a duty, which is nowadays not normally the case. On the contrary: for most people the threat constitutes a crime. The only thing that distinguishes threats from promises is that the thing you promise to do is not appreciated by the target audience.

Although there are some types of speech act typically associated with producing certain conventional effects – declarations, directives, promises, etc. – constitutive acts should not be equated or confused with specific types of speech act.

Secondly, the fact that some act constitutes another act does not mean that the constituting act is also *constitutive*: this is in principle only the case if the actor *himself* intends to perform the institutional act.

The constitutive rule can always be applied by bystanders to someone else's actions. In some cases application of a constitutive rule involves *ascription of intent* to someone else's actions, and therefore also the explicit representation of intent, while in many other cases it does not. The representation of constitutive rules should also include the possibility of representing intentions. The next section proposes a simple representation.

4.3. Intention and Action

Section 2.4 distinguished between the internal and external perspective, and the ex ante and ex post perspective on decision making. The ex ante and internal perspective of the decision maker who considers a *plan*, is very different from the ex post and external perspective of the observer who describes someone else's behaviour in terms of *actions*, i.e. ascribes a plan to others based on their behaviour.

This relation between observable behaviour and plans is mediated by the concept of *intention*. It has two distinct senses, depending on whether we take an internal or an external perspective. That is, we intend to do things and we do things intentionally. The relation between what we intend to do and what we do is too weak to reduce these senses of intention straightforwardly to each other (cf. for instance [58]). This is also recognized in law, which uses certain concepts and fictions to deal with this problem.

Ideally, we want to connect intentional action, and its conceptualization in law, to the planning problem type of section 2.2.4, without however committing to a specific

approach to *automated* planning.

To understand what intention is, it makes sense to focus on its *ex ante* sense: Bratman gives good practical and philosophical reasons for the *methodological priority of ex ante intention over ex post intention* (cf. [58]). *Ex post* ascription of intention can simply be considered an explanation of someone else's behaviour in terms of his (*ex ante*) intentions.

Alternatively, we may consider the *ex post* recognition of intent as the primary one, and simply consider the recognition of your own intentions as a special case of recognition of intentions in general: this interpretation is more in line with Newell's position that knowledge is something that is ascribed to intelligent agents to explain their behaviour (cf. [209] or section 2.2.2).

The distinction is central to the theory of mind (cf. [121]): Bratman's internal, *ex ante* theory is representative of the introspectionist or *simulation theory approach to the mind*, while the Newellian external, *ex post* theory is representative of the *theory-theory approach to the mind*. We are either simply applying an acquired *theory* that we use to understand behaviour in general to ourselves, and have no privileged introspective access to the state of our own mind, or we do have such introspective access, at least partially, and have the ability to imagine ourselves in somebody else's shoes by way of a kind of *simulation*.

For the purposes of this book, the acquired theory approach to intention is the best fit. We conceive of the ontology of the mental domain in terms of metaphors and abstraction, and expect for that reason that the way we conceive of planning is a derivative of the way we interpret behaviour.

Intention is clearly linked with planning and action, from both the simulation theory and acquired theory perspective. Observe that intention cannot be simply explained in terms of preferences, desires, goals, etc. One does not usually say that one has the intention to become a famous football player just because one dreams of being one, and one does not generally intend to have the same cake and eat it even if one desires to do so. Intentions *are* closely and inextricably linked to specific actions: intention is the commitment to a plan one decided to execute or are executing, or to a design one decided to implement or are implementing (cf. generally [27] who have a similar view on intentions).

Cohen and Levesque for instance list the following requirements on a satisfactory representation of intentions in [85]:

1. Contrary to desires and preferences, intentions held by an agent are consistent with each other and with the agent's beliefs.
2. An intention poses a problem for an agent: the agent must have a plan that he believes realizes it under certain conditions, and he must be committed to executing that plan.
3. The agent monitors the success or failure of his attempts to realize his intentions: failure constitutes a new problem.

One has the intention to do something when one is doing it. Whether the intention started just before execution of the plan, when one "committed" to it, or just after it when one "recognizes" it, is for engineering purposes of minor impact.

A description of behaviour in terms of actions is thus grounded in the recognition of behaviour as a *plan execution*. It describes a structure in two different strata: a plan in the mind of the agent that he is attempting to execute, and observable behaviour of the agent that succeeds or fails to execute that plan. It is based on the presumption that the agent, whose reasoning capabilities are limited, *must* commit to a plan and remains committed to it for a longer period of time. The agent only takes decisions once he has completed an attempt to execute a plan, either because the plan failed, or it succeeded, or because something interesting came up.

The acquired theory approach, and law, presumes that we can ascribe plans to others with reasonable accuracy based on their behaviour. Note however that the agent can interleave the execution of multiple plans, and that some plans may involve monitoring things. One may for instance intend to buy something as soon as one gets the chance: plan execution starts immediately, with a monitoring activity, but it will be difficult for other agents to recognize the plan execution. This raises some questions for the acquired theory approach, but not for the relation between plans and actions as long as one is willing to consider passive monitoring actions as part of a plan.

Another case which raises more general questions is when the plan requires us to schedule the start of its execution. If the public transport planner tells me that I have to leave in five minutes to catch the bus, I already sort of commit to it by a hasty visit to the toilet. But if I used it the evening before, then I may even have forgotten my plan the next morning.

It is in these cases that one may feel that one had an intention long before one started execution, and these are of course the hardest cases for criminal courts to crack when they have to judge whether an attempt to execute a plan was initiated. These complications are however rare.

Normally speaking, if we want to represent statements about intentionality in relation to actions, we ascribe a plan: intended are all those events not interpreted as failures of the plan.

One cannot however simply say, that the agent “has a plan”. The term plan is used in too many different senses in knowledge engineering literature, which mostly considers the meaning of the terminology under consideration in relation to automated planning. There is a difference between the plan being executed by the agent, which is what we are interested in, plans under consideration by the agent, and “primitive” plans the agent knows of that are normally executable and that the agent uses as components for the composition of plans. Since mainstream planning literature calls all of these plans (cf. i.a. [58, 25]), or alternatively tasks in the case of *hierarchical task network* (HTN) planning (cf. [107]), it is not very helpful in differentiating between a plan being executed, and the plans manipulated in the mental process that resulted in an executable plan.

Here we distinguish between a plan, which is simply the mental object we manipulate when we are planning, and the task, which does seem to suggest commitment and is the thing we eventually execute. Tasks are executed, and the detailed specification of the task presumably requires planning before, and sometimes during, execution. Plans are more generally the structures being manipulated in planning.

The following breakdown, loosely based on the description given in [85], captures

the uses of intention-related terms in this book:

problem (objective, goal) is desired change

task is executed to solve a problem (achieve an objective, goal)

competence is to know a method for performing a task, which may involve decomposing it into subtasks

plan is a specification of a method for performing a task

action is executing a task, by causing events (occurrences of changes), resulting in success or failure

intention of an action is the task one executes: it simply refers to execution

The term competence is added because it is relevant to the legal theoretical concepts of (legal) competence or power.

A similar breakdown should in principle also be possible for designs (cf. section 2.2.4), which should be treated in analogy with plans. Also explicitly addressing designs in this chapter would however add little.

Most of the things in the list above belong to the mental realm, except for action, which bridges the gap between observable behaviour in the realm of the senses and the mental representation that gives rise to it:

Proposition 3. To describe behaviour as *action*, is to interpret it as the execution of a task:

$$\text{ACTION} \sqsubseteq \exists \text{executes.TASK}$$

Further elaboration on the structure of a task is not necessary for the purposes of this chapter, and quite hard to do without committing to a specific approach to automated planning. The action is an occurrent: it happens against the canvas of space and time. The action causes events, which are intentional if they are part of a successful performance of the task. Since the task specification is something that is mostly used in natural language to contrast actual performance with it, it makes sense to presume that there is some structural similarity between the action description and the task description.

The classification of intention as a relation between action and task may be contested. Some people may prefer to attach intentions to agents instead of actions, and to the (intended) result or *goal* of the task instead of to the task. These can however be interpreted as cases of metonymy (cf. with section 2.3). It is obvious to connect the intention to the action because of the temporal concurrence of holding the intention and executing the action. The concept action already presumes the deliberating agent. To say that an agent has an intention means nothing when separated from the action.

The preoccupation with goals is a similar case of metonymy. When one blows up a car with a bomb in the hope of killing its driver, one *intends* to place a bomb in a car, and to terminate the existence of a car, of a bomb, and of the life of a driver: all of these follow from the specified task, and will be recognized as intentional in law although the law may make additional distinctions. The term *goal* should be limited to the desired death of the driver that was an input to the planning process. The action has an intention, not a goal³.

³Note that for instance [14] also attacks the common conception that goal-directed and intentional

It is perhaps arguable that a more detailed representation of intention would demand its reification and connect it to action, task, goal, *and* agent.

The notion of action as task execution plays a central role in the representation of intent in *constitutive acts*, and in particular legal acts, as will be demonstrated in the next section.

The relation between actions and tasks however also plays a role in explaining certain other aspects of law. The most complicated ingredient for a fully developed theory of intentionality is perhaps the competence to perform a task: presumably this is based on some organization of knowledge which predisposes the agent to believing the performance of certain tasks will generally be successful. Moreover other agents are also able to ascribe competence to others in the context of task execution.

Examples of such “primitive” tasks that can be performed, or recognized, without further ado according to planning literature are such complex procedures as buying a quart of milk, painting a table, taking a plane to Melbourne, etc. To explain where these primitives come from I will appeal to Schank and Abelson’s *scripts* (cf. [246]), a concept developed for automated story understanding. Section 4.5 connects these scripts to *normality*, and discusses the relations between normality and normativity.

In law, even buying a quart of milk can however become a complicated affair, because it involves a transaction between two agents free to deviate from the script. If just action is already complex, transaction adds the complication of their interaction. On the level of transactions we find a new vocabulary – associated with Hohfeld (viz. [156]) who ironically thought of them as *fundamental* legal concepts – to gain traction over our problem. This new abstraction level is the subject of section 4.8.2.

4.4. Law as an Institution

Law is an institution, and therefore can be analyzed in terms of constitutive rules and institutional facts, backed up by an institutional ontology. When specifically applied to law we can speak of *legal rules* and *legal facts*, backed up by a *legal ontology*. The *legal act*, however, is an institutional act only because it was intended by the actor to be one: we cannot properly make the distinction between the constituting act and the constitutive act. Confusingly, theft is therefore merely a legal qualification of a natural act and not a legal act, but that some act constitutes theft *is* a legal fact, if we follow prevailing doctrine.

Proposition 4. An act that constitutes a *legal act* is constitutive: it is intended to constitute the legal act.

Example 4. The notion that one must intend to create a constitutional fact is captured by one of the following uses of the *constitutes* and *executes* property, applied to

are the same thing from another angle, pointing to a substantial body of research on *unintended but goal-directed behaviour*. This complication is however not taken into account here, since [14] considers goal-directed behaviour that would not be described as intentional.

an action:

$$\begin{aligned} & \text{RAISEHAND} \sqcap \exists \text{executes.BID} \sqsubseteq \exists \text{constitutes.BID} \\ & \text{RAISEHAND} \sqcap \exists \text{executes.}(\text{RAISEHAND} \sqcap \exists \text{constitutes.BID}) \\ & \qquad \qquad \qquad \sqsubseteq \exists \text{constitutes.BID} \end{aligned}$$

A RAISEHAND action, interpreted as the execution of a bid, constitutes a bid, or, better, a RAISEHAND action, interpreted as the execution of a RAISEHAND constituting a bid, constitutes a bid. Note that the term RAISEHAND is applied to both an action and a task, as is generally done in natural language. It is the use of the *executes* property that distinguishes whether we are talking about the real thing or its mental representation.

To position the role of intent in institutional reality I will distinguish (roughly following Sartor's account in [244]⁴) between legal rules that:

1. merely state that some event or situation constitutes an institutional fact,
2. state that some act intended to constitute some institutional fact constitutes that institutional fact,
3. are *intended* (by the legislator) to enable people to create institutional facts, and
4. are *intended* (by the legislator) to enable people to create institutional facts in order to pursue their own interests.

The distinction between the second and third type is tenuous: in some cases the legal rule only confers the benefit of legal recognition to actions that would also take place without it (for instance buying and selling), while in other cases the legislator creates a rule solely for the purpose of creating a recognizable way to achieve a certain legal effect (for instance a permit application procedure for gaining permission for constructing a shed in your garden). The legislator sometimes explicitly has to *create* a recognizable way of achieving certain novel legal effects.

Each category adds another constraint, and is as such a proper subset of the previous one. The classification is however purely by function: in essence they are all the same rules. Only the second type adds a discernable component: a use of the *executes* property. If we have to ascribe task execution to someone else, the *executes* constraint is however not going to add any information, since we will infer intention from the more readily observable aspects of the behaviour under scrutiny.

If the legislator intends to enable people to create some institutional fact, one usually speaks of (attribution of a) *power or competence*, and if the power or competence is attributed to help people pursue their *own* interests one speaks of a *potestative right* in Sartor's terminology in [244]. The link between pursuing one's own interests and the introduction of the concept of a right will be treated in more detail on section 4.8.

Pursuing one's own interests is however a problematic concept. It is easier to approach matters from the angle of pursuing someone else's interests. Civil servants for

⁴Note that this account however lacks the component of ascribed intention: intention is perhaps implicitly available in the distinction between productive and behavioural characterizations (cf. section 4.7.1) made there.

instance often have powers – linked to their role of civil servant – that they can only use as part of a predefined administrative task realizing an objective of the administrative body, and outside it they don't. Within the context of the task they have the power. In KBS this notion is rarely important, as we are usually dealing with routine tasks with a predefined structure.

Sartor discusses these rule variants as types of *normative conditionals*, and discusses the constitutive or “counts-as” rule as a subtype of them. Although I do not adopt this conceptualization, I do concur with his observation that all of them are simply typical variations of a single underlying pattern: that of a rule connecting antecedent to institutional fact. Sartor gives a more limited interpretation to constitutive and “counts-as” rules, which are in his opinion non-deontic of character. This distinction is problematic: the unlawful taking of a good that belongs to somebody else clearly *constitutes* theft, but the same formulation also – as a side effect – prohibits the unlawful taking of a good that belongs to somebody else if we understand the concept of theft to be *evaluative* in character, in this case being unambiguously negative.

Intuitively one would like to arrive at the conclusion that criminal law prohibits theft, but at least the Dutch, German, Italian, and English formulations do not explicitly *command* anyone not to commit theft. In section refsect:h4-norm and 4.6 I will discuss the relation between the constitutive and normative character of rules: here we only come to the conclusion that neither theft, nor the act that constitutes theft is a legal act.

Constitutive acts – in general – can be informal or formal, and by this we usually mean: backed by text. Text is taken to mean any representation preserved in a form whose existence is independent of both sender and receiver. Formal acts play a central role in law. Law uses formal constitutive acts to formalize constitutive rules, enact constitutive rules, and to repeal constitutive rules. We can therefore also speak of *formal rules*, constitutive rules created by a formal act, and in the case of the law these are fortunately explicit and unambiguously prior to their effects.

Definition 9. A *formal act* is an act of representation, preserved in a form whose existence is independent of both sender and receiver, of an institutional fact, with the intent to be constitutive of the represented institutional fact.

Example 5. In a literate society, the legislator often chooses for formal legal acts, since these make it easier to recognize intent (i.e. task performance). Instead of bidding by raising hands, auctions may also accept formal written or electronic bids:

$$\exists \text{result}.(\exists \text{represents.BID} \sqcap \exists \text{receiver.AUCTIONMASTER}) \sqcap \\ \exists \text{executes.BID} \sqsubseteq \exists \text{constitutes.BID}$$

This means that an action that results in a message representing a bid to an auction master, with the intent to perform a bid, constitutes a bid. The result property is not the recommended way of representing what is brought about by actions, but strictly serves as a simple example.

Proposition 5. A *legislative act* is a formal act, that creates institutional rules, constitutive rules, and/or institutional facts required for correct interpretation or functioning of institutional rules or constitutive rules.

The legislative act is a strong argument for the institutional interpretation of law. A powerful criticism of the notion of the constitutive rule is to point out the suspect causal mechanism: if the rule backs the institutional fact, then obviously the rules exist prior to the institution, and if their results are recognized by an audience, then that audience should be able to verbalize the rules in effect.

Example 6. MacCormick in [192] discusses the example of queueing – a behaviour that a large part of the world population engages in – to show that the distinction between institutions and non-institutional conventions or normative orders are hard to draw using the concepts of constitutive rules. It is not obvious that the participants in a queue will be able to verbalize the rules in effect. What is clear, however, is that queueing behaviour will be increasingly formalized – in the form of a visible and recognizable queue – as more people start to participate in it and queue-jumping starts to occur.

While we can doubt the construct validity of the constitutive rule in general, we can rest assured that the analysis of institutions should at least work for the law already in operation, at least to the extent that it formalizes its constitutive rules. Only constitutional law itself, which gives the rules for creating and changing the rules, usually came about in a rather messy fashion (cf. for instance [260]).

Law of course is not a monolithic system. There are many institutions that assume legislative power, and there is no guarantee that these institutions cooperate in harmony in creating their legal ontologies, legal facts, and constitutive rules, or even that they recognize each other. A legal institution can selectively recognize the institutional facts of another institution by way of constitutive rules, which may be formal or informal. It can also adopt the ontology of another institution without recognizing its institutional facts altogether: they merely use the same terminology, but ignore each other's institutional facts completely. Clearly marking off the boundaries between different institutions, or deciding what it is that is being recognized, is not always trivial.

Proposition 6. A legal institution can recognize the institutional facts of another institution by way of constitutive rules.

Example 7. A simple example is the recognition of driver's licenses. The law in many countries regulates who is allowed to drive, in order to allow only people who are competent to drive on public roads. When the legal system finds someone competent to drive it issues a drivers license. This (constitutive) act is intended to create the institutional fact that this person is licensed to drive, and it is formal: it results in a physical representation of this institutional fact in the form of the drivers license. The drivers license is *prima facie* evidence for the license to drive, but the license to drive may be retracted (constitutive act), in which case the drivers license becomes void if this happens in his absence.

A country may recognize the institutional fact of another country by way of a constitutive rule, or it may simply directly recognize the foreign physical driver's license as proof of competence by way of a constitutive rule. The two mechanisms have different results in relation to voided licenses, but to find out that the driver's license is voided the two countries obviously need to exchange institutional information.

$$\begin{array}{c} \text{BELGIUM:DRIVINGCOMPETENCE} \\ \sqsubseteq \exists \text{constitutes.NETHERLANDS:DRIVINGCOMPETENCE} \end{array}$$

$$\begin{array}{c} \text{DRIVERSLICENSE} \sqcap \exists \text{represents.BELGIUM:DRIVINGCOMPETENCE} \sqsubseteq \\ \exists \text{constitutes.NETHERLANDS:DRIVINGCOMPETENCE} \end{array}$$

In the first example it is an institutional proposition, belonging to another institution, which is recognized. In the second example a physical object is recognized.

Institutions may use very similar, or even exactly the same, institutional ontology without sharing constitutive rules. The similarity is in this case very shallow, as there is no link at all between both institutional realities. The institutions are in effect sharing the same abstract models.

4.4.1. Constitutive Rules and Knowledge Representation

At this point we have the basic ingredients for representing legal rules as constitutive and institutional rules: the `constitutes` and `executes` properties, relating disjoint ontological strata.

It is however incorrect to say that the domain and range of the `constitutes` property are disjoint in general; They are disjoint from the perspective of any one specific institution. Institutions do however recognize each other's institutional facts as brute facts. The institutional interpretation is an abstract model, to be implemented for any specific institution i by creating a specific property `i:constitutes` that does relate disjoint domains.

There is also another important point to be made here: Although the logical form of institutional rules and constitutive rules is being discussed in this chapter, we have not fully committed to the notion that legal rules *are* institutional rules and constitutive rules. As we will see in section 5.2.1, they are not.

The simple examples used sofar in this section have all used OWL DL axioms instead of the defeasible rules introduced in section 3.4.2. The question is however whether the use of constitutive rules should be represented with OWL's monotonic \sqsubseteq operator or as a kind of ampliative reasoning. *Institutional* rules define the institution's ontology and are properly modeled by the monotonic \sqsubseteq operator. *Constitutive* rules however perform a mapping from one ontological stratum to another ontological stratum and this mapping may be defeasible.

Each of these strata have their own space of possible models as described by their ontology, and there is no guarantee that constitutive rules always perform an unambiguous mapping from one stratum into the other one. It is in fact not sensible to

assume that they do, and we need to resort to the defeasible rules of section 3.4.2 if they don't.

Constitutive rules may come in the $C_b \sqsubseteq \exists \text{law:constitutes}.C_i$ *indicator* and the $\exists \text{law:constitutes}.C_i \sqsubseteq C_b$ *requirement* varieties, or the combination of both, where C_b is a concept from brute reality, and C_i a concept from institutional reality. To reduce maintenance we will assume that it is the $C_b \sqsubseteq \exists \text{law:constitutes}.C_i$ that is defeasible, and should be modeled by a rule of the following form:

$$\begin{aligned} &(\text{DEFAULT } r \text{ (known}.C_b) \text{ (free.}\exists \text{law:constitutes}.C_i) \\ &\quad \text{(assume.}\exists \text{law:constitutes}.C_i)) \end{aligned}$$

A matching brute reality and legal reality is not guaranteed by the model set described by the source of law: it describes the part of brute reality that specifies the interface to the institution, not the part of it that doesn't. The purpose of the rules is not to describe the ontology of the brute reality the institution is deposited on top of. The ontological coherence of institutional reality and brute reality take precedence over the applicability of constitutive rules.

Proposition 7. A *legal fact* can only come into existence if it is *consistent* with the settled facts, and if there is some constitutive rule that *indicates* it.

Consider the following detailed, but simplified, real world example, loosely based on [244], who models a similar structure in two different rules:

Example 8. Let us say that 1) the taking of something constitutes theft, and that 2) theft must have been performed with the intention to appropriate. Interpret this as a *set* of two rules, in no particular order.

The separation into two rules presumably comes from Italian criminal law; The German and Dutch version both have a “with the intent to” sentence fragment *embedded* as a condition in their variant of rule 1, which changes matters. The issue is not whether these rules accurately interpret Italian law, or any other law, or the intended reading in [244].

The freedom of interpretation increases if rule 2 reads: *The* theft must have been performed with the intention to appropriate. To come to this interpretation for rule 1 and 2, *the* in rule 2 must refer to a theft as in rule 1. See section 5.3.4 for this alternative reading, which depends on the structure of the writing that expresses the rules. In this case rule 1 and 2 are really one rule, just as would be the case if 2 was embedded in 1 as a sentence fragment. Another variation on the same theme is the following formulation: *The* theft of rule 1 must have been performed with the intention to appropriate. If the two rules are however structurally separate entities, and there is no *the* connecting them on the linguistic level, they should not have to be joined in a rule or OWL class.

In the rules as I phrased them the (observable) taking *indicates* in my view the possible existence of a theft, while the intent to appropriate (which happens to be not observable) is a *requirement* to classifying something as theft.

The sense of the verb *must* as used here hints at a necessary condition, or the “epistemic obligation” to apply the rule. [244] gives another interpretation of the

verb *must* used in this sense. The verb is in my view however a red herring: the interpretation of the sentence does not change if it states that theft is performed with the intention to appropriate. What matters is whether or not the sentence can be interpreted as being embedded in rule 1.

Observe that this example illustrates that although we are usually “deriving” institutional reality (the theft) from brute reality (the taking and the intention with which it is done) arrows may well point the wrong way in rules:

$$\begin{aligned} \text{TAKING} &\sqsubseteq \exists \text{law:constitutes.LAW:THEFT} \\ \exists \text{law:constitutes.LAW:THEFT} &\sqsubseteq \exists \text{executes.APPROPRIATION} \end{aligned}$$

The first rule is a prima facie sufficient condition for theft, and the second one a necessary one (see figure 4.1).

This interpretation is as much saying that takings always happen with the intent to appropriate. If you take something from the shelf in the supermarket, you do not have the intent to appropriate it: there is no acceptable mapping from brute reality to legal reality if we interpret the constitutive rules as OWL axioms. This is however a perfectly reasonable set of rules.

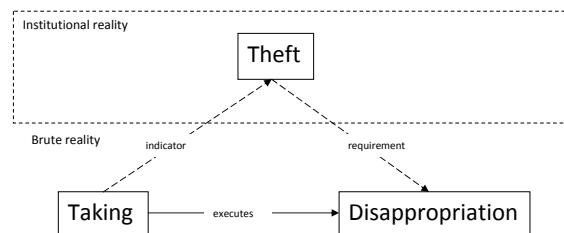


Figure 4.1. The act of taking functions as an indicator of theft, among other possible indicators. The intent to appropriate is however a *necessary* condition. The total of rules describing the relation between a theft and the brute facts underlying it define its constituting base.

Consider what happens if in the future we would have to add 1b) illegal copying (also) constitutes theft⁵:

$$\text{COPYING} \sqsubseteq \exists \text{law:constitutes.LAW:THEFT}$$

This creates one of the major maintenance problems often encountered in the representation of sources of law. Rule 2 should in this phrasing have an effect on thefts indicated by rule 1b, and therefore cannot be embedded in rule 1. Moreover if rule 2 reads *the* theft [...] the article *the* suddenly resolves to rules 1b if it is inserted between rule 1 and 2.

⁵This is an ontologically absurd extension that will never happen, despite the clamoring in the media about copyright violation as theft, but the descriptive fidelity of the example is not the issue. The example is about the two rules as given here, approached without assumptions about how things should work.

An embedding consists of combining the rules into one by making the necessary condition part of the sufficient condition:

$$\text{TAKING} \sqcap \exists \text{executes.APPROPRIATION} \sqsubseteq \exists \text{law:constitutes.LAW:THEFT}$$

The appropriation is in this case not a necessary condition. This is a very common approach, and the solution chosen by [244], who clearly has an embedding in mind: theft is the conclusion of the rule. The rule now settles for an *unknown* intent to appropriate in OWL DL, since we may simply assert that something constitutes a theft without restriction. The problem is also that the rule has to be changed to accommodate rule 3 about COPYING.

From a database-oriented point of view one may also argue, quite plausibly, that constitutive rules only map from legal reality to the brute one, and take the position that all constitutive rules therefore specify necessary conditions:

$$\begin{aligned} \exists \text{law:constitutes.LAW:THEFT} &\sqsubseteq \text{TAKING} \\ \exists \text{law:constitutes.LAW:THEFT} &\sqsubseteq \exists \text{executes.APPROPRIATION} \end{aligned}$$

If we however add rule 3 theft would require taking a good *and* copying information instead of *or*, which can't be the intended interpretation, so we still have to modify an OWL axiom to accommodate a new rule.

Reformulating the rules to achieve specific effects as in the example is not a good idea from a maintenance point of view. Rule 1 and 2 should only be taken together in knowledge representation if one is of the opinion that they *express* only one rule. In that case, which may be indicated by the use of the article *the*, rule 1 and 2, when read separately, are clearly meaningless. Since there is no need for isomorphism with rule 1 and rule 2 respectively, there is also no need for a special type of rule insertion an extra condition in other rules, as [244] argues.

Both reformulations above try to answer the question which thesis is defeated, taking into account ontological stratification: the first one presumes a settled brute reality and explains it in terms of legal facts, while the second one assumes a context in which one is explaining a settled legal fact in terms of a matching brute reality.

These approaches are both correct interpretations of the use of stratification: one takes one stratum as the independent variable and the other one as the dependent variable. But the source of law should not be directly represented in these ways: the theory put forward by the source of law is after all perfectly consistent in itself.

The set of rules constructed here does not make that possible. The solution for this set of rules, based on the assumption that legal facts only come into existence if they are consistent with the settled (brute *and* institutional) facts, is to guard the derivation of the legal fact with the requirement that this derivation should be consistent:

$$\begin{aligned} (\text{DEFAULT } r_1 (\text{known.TAKING})(\text{free.}\exists \text{law:constitutes.LAW:THEFT}) \\ (\text{assume.}\exists \text{law:constitutes.LAW:THEFT})) \\ \exists \text{law:constitutes.LAW:THEFT} &\sqsubseteq \exists \text{executes.APPROPRIATION} \end{aligned}$$

The derivation of the institutional fact of a theft is based on the *ampliative* inference from an observed taking and the non-ampliative inference to the intent of the taking. The two rules in other words describe a typical *generate & test* approach (cf. section 2.4.1). Note that logically speaking, if one makes the open world assumption, it is possible to assert that something is a theft without proposing that that something is a taking.

The reason we don't do that is because we view the law as a category of systems that can only be manipulated through the interface of socially recognized constitutive rules. For legal facts this means that it is relevant which rule was applied (which is addressed in section 5.2.2).

In rare cases we may also want to use constraints.

For purposes of application in a KBS one also needs to decide whether one assigns a burden of proof to the user with this set of rules. In this case one might for instance want a separate argument (as defined in 3.3) for the ascription of intent instead of allowing intent to be justified by default by the taking itself, since the intent is one of the constituting facts:

$$\begin{aligned} &(\text{CONSTRAINT } r_2 \text{ (known.}\exists\text{law:constitutes.LAW:THEFT)} \\ &\quad \text{(fails.}\exists\text{executes.APPROPRIATION)}) \end{aligned}$$

Even though $\exists\text{executes.APPROPRIATION}$ is a dispositional belief in a knowledge base using the necessary condition stated above, there is no assertion that refers to the task being executed by a constant. Note that also this constraint does not break the isomorphism requirement: it clearly refers to rule 2 only.

These constraints are not directly relevant for the reusable representation of the source of law: one does not automatically assume the burden of proof just because one applies the rule to explain people's behaviour. Burdens of proof must be met by formal justifications, not by reasoning in general. Datalog-style constraints are, as opposed to OWL's *tbox* axioms, defeasible. The falsification of a constraint is commonly used to drive interaction with a user.

Proposition 8. A *burden of proof* can be represented with a constraint.

The representation of burdens of proof by constraints is sufficient for simple Datalog-based KBS as sketched in section 3.4.1, where one can simply pick a set of constraints on the amount of elaboration necessary for acceptable arguments that one would like to enforce in the (static) KBS problem setting.

If one wants to make the applicability of burdens of proof *conditional* on features of the problem solving setting, a more advanced representation is necessary. Explicit reasoning on the allocation of burdens of proof is for instance often required in legal argumentation (cf. for instance generally [229]).

4.5. Norms, Normality and Normativity

In the preceding sections the institutional interpretation of legal knowledge was developed. There is however another important, in many accounts even more important,

aspect to the law. The law is often said to formalize *norms*. In this section the notion of norm is investigated, from a cognitive perspective, before we turn to the normative rule of section 4.6. This section also sets the background for chapter 6, which will address the *normative order*.

We commonly make a distinction between *normality* and *normativity*. Normality refers to the norm from a descriptive perspective: it merely describes an observed regularity in the behaviour of instruments or agents. Normativity only applies to the behaviour of agents, and communicates something in addition to the mere regularity of behaviour: the behaviour is motivated by the recognition of the norm by those whose behaviour is consistent with it.

Normality and normativity are different things, and obviously should be treated as such. We learned long ago from Hume in [160] that “ought” does not follow from “is”. This relates to the discussion in section 2.2.2 about the truth value of norms: no constellations of facts can ever prove the truth of a norm. The norm can be communicated between agents, and it can be recognized and adopted, but it is never the spontaneous result of logical reasoning from facts.

There are however a number of ways in which normality and normativity, different sides of a same coin, interact. In the commonsense conception of justice this relation is clear, and fairly well-understood. Montaigne already pointed this out in [90]:

I am prepared to forgive our own people for having no other model or rule of perfection but their own manners and behaviour, for it is a common failing not only of the mob but of virtually all men to set their sights within the limitations of the customs into which they were born.

The very word morality derives from *mos* or *mores* (in plural) *maiorum*, the ways of our ancestors. That custom is the basis for what is moral seems to involve a shift from what is to what ought to be.

Normativity as a source of behaviour is the traditional province of Computer Science & Law and Legal Knowledge Engineering. Normality is the province of descriptive sciences, and in Knowledge Engineering is implicit in notions like *defaults* (cf. generally [175]) in reasoning, world knowledge and *common sense knowledge* (cf. [63]), or Schank’s scripts (cf. [246]).

Not all norms have something to do with behaviour of intentional agents. The *norm* is identified by its *epistemic role* in problem solving and not something that exclusively *belongs* to the vocabulary of the legal domain, or to that of morality. A norm is a standard of performance, a measurement scale. It is used to predict performance and to diagnose performance contrary to expectations, regardless of domain (cf. [266]).

The norm is conceptualized differently in different contexts: When assessing readings from a broken instrument – for instance some circuit board – a norm is for instance the distribution of scores obtained from a correctly functioning norm group, or some margin around the specified ideal functional mapping from input to output that describes the expected and intended behaviour of the instrument. The expected behaviour of the instrument is also intended: the instrument was designed by an intentional agents with some intended purpose in mind. This process shares little with law except the epistemological roles (assessment, norms, case, qualification) in the problem solving process.

In sociology, a *social norm* is a pattern of behaviour expected of an agent within a particular society in a given situation. In this context the norm is interpreted as directly reflecting some preference of the involved agent; Given that an agent is aware of a number of behavioural alternatives and made a choice between them, we can infer that the agent revealed a *preference* since choice and preference are interdefinable. If we observe a pattern of agents revealing the same preference and suspect that there is some social mechanism that explains how that shared preference arose, we may infer that there is a social norm that explains that preference. What makes the behaviour normal is simply the fact that it is *expected*.

There is a presumption, for instance in predicting the consequences of legislation, that people will normally do what is normative. In other words, it is normal to comply with the norms.

In legal theory we also find mention of custom as a “source of law”. The unlawfulness of taking a good (theft) is for instance judged against a cultural context. Is it normal to take something from a shelf in a supermarket? Is it normal to take your neighbour’s ladder to get your cat out of a nearby tree? The assumption is generally that what is normal is lawful, and legislation prohibiting what is normal will be less easily accepted.

Not recognizing this can lead to serious overestimation by legislators of their power to change society. A legislator that likes to be in control should limit himself to instructing people to do what they are willing to do.

Normality and normativity are just two ways of explaining patterns in behaviour: something is normal according to the observer because the observed pattern happens to reveal it to the observer, and it is normative if it motivates (or “causes”) the choices that lead to the pattern. Normality appears as a social fact and a social value, because the norm is used as a standard of reference against which deviations from normality are evaluated. The normal has both factual and normative force. Associated with the former is custom which is revealed through observation as a matter of fact. Associated with normativity are morality, ethics, and the law.

To explain how normality and normativity relate, and what effects this relation has that are relevant for representing and evaluating legislation, I will introduce the concept of *status quo bias*. It is central to the formation of common sense standards of justice or fairness, and can for instance be taken into account in comparison and evaluation of legislation on behalf of the legislator to judge the effectiveness and efficiency of legislation.

This concept is related to a version of economic theory called *prospect theory* (as opposed to *utility theory*, which is almost universally considered a standard for rational decision-making), credited on Kahneman et al. in [164]. His account, based on results from experimental psychology, revolves around the *reference transaction*, and its role as a yardstick for what is fair. That is, what we become accustomed to (the status quo) attains a normative status against which deviations to our disadvantage are considered unjust.

Status quo bias refers to the finding that in games an option is more desirable if it is the status quo for no other reason than that it is so designated (cf. [165]). Status quo bias is often discussed in relation to the endowment effect and loss aversion (cf.

[166]). Status quo bias differs from the other two in that it does not (even) depend on framing changes in terms of (possible) losses and gains.

The endowment effect refers to the differential weight placed on the value of an alternative depending on whether one “possesses” the alternative and is faced with its loss or whether one does not possess it and has the potential to gain it. Losing a given alternative, which is part of one’s endowment, is felt to be a greater loss than the corresponding chance of gaining it when it is not part of one’s endowment. Losses are more heavily weighted than foregone gains.

Furthermore, a certain degree of inertia is introduced into a choice process since things that are already included in the individuals endowment will be more highly valued than those not held in the endowment. Kahneman and Tversky introduce the term loss aversion to capture the relative steepness of the loss-portion of this value function relative to that on the gain side of it.

Kahneman et al. in [164] developed the idea of a *reference transaction*: a relevant precedent that is characterized by a reference price or wage, and by a positive reference profit. In the course of life, certain practices and patterns establish themselves and attain the status of normality. If someone disrupts this normal course of affairs, then the disruption will be considered unfair if:

- it is to the advantage of the other and to your disadvantage, and
- if there is insufficient justification for the disruption.

The involved parties are entitled to the terms of the reference transaction. The reference transaction, as a custom, is part of one’s endowment.

Example 9. Consider the following example from [164]: A hardware store has been selling snow shovels for \$15. The morning after a large snowstorm, the store raises the price to \$20. The reference transaction, given by precedent, is marked by a price of \$15. The price increase transgresses the reference transaction. In a telephone survey, 82% of the 107 respondents deemed the price increase to be unfair.

Kahneman et al. (cf. [164]) claim that individuals consider the reference transaction to be fair *because* it is normal: individuals do not have any anterior criterion of justice which informs their judgments of fairness. Kahneman et al. even suggest that any stable state of affairs tends to become accepted as fair eventually. The reference transaction consists in what people have become accustomed to but it also becomes a normative yardstick for assessing the fairness of deviations from itself. This is perhaps philosophically suspect from Hume’s point of view at least, but has also been recognized as an important factor in the sociology of law (cf. for instance [119]).

Kahneman’s work in psychology has had quite an impact on economics, eventually resulting in a Nobel Prize for economics in 2002, because it attacks the fundamentals of utility theory as a predictor of actual human behaviour in a systematic and empirically well-founded way.

The notion of a reference transaction is naturally connected to the *script* in knowledge representation, and my earlier tentative specification of what competence is: to know a method, gleaned from episodic memory, to perform a task. If the snow shovel is made significantly more expensive, or the quart of milk is no longer for sale, etc.,

this is unfair because it takes away something from us: the potential to do certain things following a predictable script.

The champion of scripts in knowledge engineering is Roger Schank (see [246]). The key idea involved in scripts is that our knowledge of concepts, events, and situations is organized around expectations of key features of those situations. Many choices are made in the context of a script: you do the same thing every time as long as the salient features of the scripted situations don't change.

In Schank's theory, memory is episodic, i.e. organized around personal experiences rather than semantic categories. Generalized episodes are called scripts. Schank's ideas were developed in the context of story understanding, but have also played a key role in the development of hierarchical and case-based planning algorithms (the elusive store of *skeletal plans* and *primitive plans*; viz. [107]). Scripts are also organized around the *social roles* (buyer, employee, customer, student, citizen, father, etc.) one plays in life (see for instance [198] for more on social roles).

Proposition 9. Whether behaviour is *normal* is assessed against scripts associated with the social role one assumes.

It only makes sense that institutional settings also get internalized as scripts, as reference transactions, i.e. become the normal way of doing things, and a guide for predicting the behaviour of others.

The legal *normative rule* is a constitutive rule with the purpose of creating normative order. It does so by way of positive or negative qualifications of behaviour. The unlawful taking of a good that belongs to somebody else *constitutes* a theft, theft negatively qualifies the taking, and gives license (through other rules) to other agents to do things to the thief that are generally speaking not in the thief's interest. This regulative function of a fraction of legal rules is however not the only way in which the law promotes normative order.

Both normative rules and other constitutive rules of the institution of law tend to affect what is perceived as normal. Legislation, in so far as it induces certain types of behaviour, has an impact on habit formation and hence on the formation of the standards of justice of the future. If the legislator sets a standard for legally recognized sales transactions, actual sales transactions will tend to conform to the prescribed model as it gradually becomes the new reference for sales transactions.

Normative rules qualifying undesirable behaviour in sales will be phrased in terms of a deviation of these reference transaction: in the interest of predictability it is therefore a good idea to follow that script, and if someone deviates from the script *to your disadvantage and their advantage*, it is only natural to presume that what he does must be illegal.

Proposition 10. All law, not just its normative part, promotes normative order insofar as it has an impact on habit formation and hence on the formation of the standards of justice of the future.

Law has a *de facto* effect on prevailing social standards through habit formation and the formation of standards of justice. This is however not all. The legislator obviously also has *intentions* with the normative order that he tries to realize through legislative action.

4.5.1. Legislative Action

As an observation about legislative action the reference transaction is very relevant. Firstly, it predicts, as noted, that the addressees of changed legislation will generally speaking cling more fiercely to existing reference transactions than utility theory would predict.

Secondly, it gives a basis for understanding *norms of analysis* as understood by Hettich and Winer: the norms used to evaluate proposals for legislation (cf. [146, 145]), which will be addressed again in section 6.7.3.

We for instance expect the following from the combination of Schank's scripted behaviour in planning and the formation of standards of justice around Kahneman's reference transactions:

- Stakeholders will usually evaluate a change in legislation by considering roles – buyer, employee, employer, etc. – they typically or regularly play in reference transactions, and not their future position as a rational agent that could change roles. The buyer of a snow shovel will not contemplate becoming a seller of snow shovels. Political lobby organizations that try to influence legislation obviously do so from the conception of a role the members they represent play.
- The legislator will also usually take into account the reference transactions created by the existing legislation and evaluate the new one as a deviation from it. It is possible to extrapolate roles in these reference transactions to populations filling the role, certainly in macro-economic modeling of taxation and social security⁶. The estimated number of people playing each role will often be considered constant in the legislator's evaluation, except when “social movement” from role to role is the purpose or a salient aspect of the change.
- It is easier to change non-salient aspects of the reference transaction than salient ones. Making people pay more often by making the transaction occur more often for instance works better than making them pay more in the transaction.
- Peacock [221] makes a more ominous prediction about the legislative process: the reversal of policy measures is more difficult than their introduction, particularly if the losers of a reversal are clearly identifiable. Consequently, according to [221], a government should consider this effect of its policy measures before implementing them. If a policy is intended as a short-term measure, it may, when it comes to reversing it, have become so customary that reversal is very difficult or nearly impossible.

The appropriate normative yardsticks for evaluation of legislation measure deviations from abstract reference transactions. The norms and concepts by which a new version of legislation is going to be judged are probably largely based in the existing version. Changes made will be Pareto improvements – at least one individual better off, without making any other individual worse off – but in terms of *social roles* identified in the relevant reference transactions and not of individuals, which would be an impossibly hard problem. In [263], Tanghe also makes this distinction in political

⁶Dutch readers will be familiar with the continually updated *koopkrachtplaatje* (roughly *purchasing power table*) and the population segment they are in.

discourse focusing on inequality and its relation to injustice: this concept in its most widely accepted sense revolves around equality of ability to assume social roles, Tocqueville's *equality of conditions*, but this form of equality has arbitrary results when extrapolated to specific populations of one's choice.

These social roles, and the expectations about them, play a central role in organizing the knowledge we use in planning.

4.5.2. Representing Normal Behaviour

Kahneman's reference transactions directly relate to the perceived competence to perform mundane or "primitive" tasks. The effect of constitutive rules is to create a script through which one can achieve certain legal effects, i.e. recognized by others. If these effects are believed to be beneficial by the agent, the agent will generally speaking use the script in planning. Conversely, other agents will ascribe plans to the agent based on the same scripts and assumptions about his beliefs and desires, and recognize the legal effects he brings about. Normal behaviour is simply expected behaviour.

If legal effects of constitutive rules are believed to be detrimental by the agent, the agent will use the script in planning to recognize bad plans. The most important rule of this kind is obviously the normative rule. Agents will however also evade for instance liability to taxation, or legal effects that create costly duties, liabilities, or disabilities.

Example 10. For instance: if you need a building permit for a shed higher than 2.5m, this is an argument *against designing such a shed*. Another nice example comes from the CLIME project (cf. [284]): There is a maximum size constraint for ships allowed entry into the Panama canal called PANAMAX. This rule is however irrelevant for operators of ships: ship designers all over the world use it as a hard constraint for ship design.

We apply the notion of "normal" behaviour both to objects and to other people (and perhaps, following [95], to ourselves). Certain categories of objects, tools, machines, artifacts generally, have a *function* or instrumental role. We have designed them for the purpose of exhibiting certain predictable and useful behaviours when we interact with them in certain ways, and this allows us to perform certain tasks. We also classify them by that function. If their behaviour deviates from the expected one, usually in a detrimental way, they are broken and need to be repaired or replaced.

When interacting with other people we do exactly the same. To buy a quart of milk, you have to find someone who can fill the agent role of *seller* of milk, etc. We judge deviations from expected behaviour in the same way, except that the remedies are of a different character. [198, 153] point out that social roles are identified by the set of actions that can be performed in that role. The social role is also a separate object whose existence depends on the "brute" object in the stratum below. This is for instance shown by the application of terms like "good" to roles like seller, student, or cook: the properties apply to the agent in the role, but not to the agent separately (cf. [139]). Legal roles are simply social roles in a legal institution. Roles are functions when applied to objects, and agents when applied to persons as the active and efficient

cause of events. Agent roles that depend on recognition, and are therefore institutional, are social roles, and if they depend on recognition of the law they are legal.

These sets of actions connected to agent roles can be composed into scripts with interacting participants, who each take a role. To recognize a series of events as an instance of the script, is to expect that the other agents will act according to their role. If they deviate from the script, a new explanation of the situation is needed. Executing a task is to play your role in a suitable script you selected to solve a problem or achieve a goal.

To adopt the agent role in a script one must take a number of normal decisions, expressing a set of preferences, which are attributed by others to the agent observed to be involved in the script. Deviation from the scripted conditions leads to a re-evaluation of one's behaviour.

For the representation of sources of law, this information about the planning process has only an indirect relevance. As stated in section 4.3 we do not commit to a specific account of planning, and therefore have no model of what this information structure looks like that relates attributed social roles to expectations about the intentions of the filler of the social role.

Since our knowledge about normal behaviour relates to the agent role rather than the person, we should however explicitly distinguish the person who plays an agent role from the agent role itself. To be an agent (role), is to be played by a person and to be the actor in an action:

$$\text{AGENT} \equiv \exists \text{playedBy.PERSON} \sqcap \exists \text{actorIn.ACTION}$$

The higher order “set of actions” *belonging* to an agent role is simply indicated by the following type of axioms (the actor in a theft is a thief, and vice versa):

$$\exists \text{actorIn.LAW:THEFT} \sqsubseteq \text{THIEF}$$

The notion of agent roles is essential to the interpretation of competence, including legal competence, and normative order. The term *agent* in this book, and specifically in section 4.7, does not equate to person.

Interestingly, however, the normative rule is often explained without any appeal to agent roles.

4.6. Legal Normative Rules

There is a simple criterion to distinguish *normative* constitutive rules from other constitutive rules: the normative rule can either be violated or it can cancel the violation of another normative rule. The normative rule can be defined in terms of the institutional legal fact it can create: the *violation* of a normative rule, or the cancellation of the violation of a normative rule. The first one could be naively represented as:

$$\text{EX:SOMETHING} \sqsubseteq \exists \text{law:constitutes.LAW:VIOLATION}$$

As already indicated in section 4.1, normative rules come in the three major flavours of obligation, prohibition, and permission. The significance of the normative rule of

the obligation or prohibition type in legislation is in its effect as a constitutive rule: by stating that one ought to do or bring about x , it states (as a matter of terminology) that not doing or bringing about x is a violation.

It is the institutional qualification violation, which is inherently *evaluative* in character, that opens up the interpretation of the act as wrong in the deontological sense. It is this interpretation that makes *deontic logic* an obvious tool for representation of law. The effect of a permission is to cancel what would otherwise be a violation: it only means something in the presence of other, conflicting obligations or prohibitions.

Proposition 11. A *normative rule* is a constitutive rule that derives its violation, or the cancellation of a violation, from one or more constituting facts.

The institutional ontology of law incorporates inherently evaluative concepts like violation, obligation, crime, theft, etc. because it is intended to evoke the deontological interpretation, just like the institutional ontology of chess appeals to concepts like winning and losing because it is intended to evoke competition, and a manufacturer of hammers calls its products hammers because it intends to evoke an evaluation of their usefulness for delivering blows, which would not be conveyed by calling the hammers paper weights.

The purpose of law is to institutionalize normative order.

The concept of violation also evokes the concept of *punishment*. Not because violations are always followed by punishments, but because punishments *are by definition preceded by* violations. The preceding violation is the terminological *conditio sine qua non* of punishment⁷ Punishments are obviously not just disagreeable consequences of one's actions, but intentional and predictable counteractions by others intended to suppress violations.

Moreover, as already suggested in section 4.2, the following should also be considered (inter alia) a normative rule: "The taking of a good, that wholly or partially belongs to another, with the intent to unlawfully appropriate it, constitutes theft and will be punished with a prison sentence of at most four years". It is not the verb used or any other structural regularity in the expression that identifies the normative rule, but its inherent evaluativeness. It works as a standard constitutive rule for determining a legal fact of theft, negatively qualifies it by using a term – *theft* – that's inherently negatively charged⁸, and concludes with an institutional rule capping the length of prison sentences for an act that constitutes theft⁹.

The reference to *punishment* for theft makes this immediately recognizable as a normative rule. Not stealing is affirmed as a norm by the normative rule.

The classification of a rule as normative immediately puts a rich vocabulary of normative positions at our disposal. As already pointed out, normative rules are most commonly phrased as obligations, prohibitions, or permissions, but the surface form

⁷[141]: "Legal responsibility is the liability of a person to be punished, forced to compensate, or otherwise to be subjected to a sanction by the law."

⁸See for instance [139] for more about the relation between concepts and value statements.

⁹To represent this last rule transparently, one would first need to fix an interpretation of the more general rules of criminal procedure.

of the rule is – contrary to the examples given of constitutive rules – not determinative of its interpretation.

The ontology of normative positions is shared among legal institutions, and is indeed an identifying feature of legal institutions. This merits special attention for this ontology. Since this ontology is however based on a abstraction level beyond that of (institutional) legal rules created and communicated by sources of law, the formal account of normative rules is treated separately in chapter 6. This ontology is not found *in* the sources of law, but is rather background knowledge to the interpretation of the (intended or expected) effect of the source of law on behaviour.

In discourse we encounter the use of normative notions like obligation both in a generic sense, as an identification of a type of surface form of rules, and as a normative position in context relativized to specific objects. Consider the following two examples:

1. If (seller) offers to sell some (good) to (buyer), and (buyer) agrees to buy the good then (seller) has an obligation to sell the (good) to (buyer).
2. I have an obligation towards John, to sell him the painting I offered him for sale *if* he agrees to buy the painting.

The first one is a rule. The second one is partially filled in script based on the rule that describes a legal position: it indicates that some of the conditions for deriving a violation of the rule have already been met, and that if John agrees to buy the painting, which is outside my control, not selling the painting to him would constitute a violation. One is in a “state of obligation”. Clearly, the obligation and the rule are separate entities: the rule apparently gives rise to states of obligation, which are also institutional facts:

Proposition 12. A *normative rule* can be used to derive a state of obligation to do something, or the cancellation of such a state of obligation, from one or more facts.

The state of obligation is clearly ontologically related to the institutional fact of violation. Obligation is traditionally represented by an operator O , “it ought to be that”.

Depending on the kind of logic and the type of problem considered, knowledge engineers choose for different conceptualizations of normative rules.

There are several simple rule-based approaches that try to approximate norms in Datalog rules with notions like violation or obligation (cf. for instance [162]) wherever necessary.

A principled approach to representing normative rules as rules that derive an institutional fact from brute facts is the functional approach of Valente in [266]. It depends on a value function v mapping from descriptions to the ordered set $\{allowed, silent, disallowed\}$, where violation obviously maps to disallowed, and allowed cancels a disallowed in certain circumstances:

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

$c \in \mid \phi \mid$ means: c is a member of the extension of ϕ (i.e. $\mid \phi \mid$ is the set of worlds w such that $M, w \models \phi$). This representation is clearly committed to considering normative rules as constitutive rules that derive their own violation (here disallowed), or the cancellation of the violation (here allowed) of another norm. Silence is the *absence* of an institutional fact.

Its strength is in dealing with permission and handling conflict between norms (cf. for instance [286], and section 6.2). A weakness is that it cannot be used to detect a *contrary-to-duty obligation* (cf. section 6.4), which is one of the tests we will use in chapter 6. [266] is satisfied with the observation that this approach does not lead to the contrary-to-duty paradox.

Valente's account strongly influences the view of normative rules in this book, in which we will adopt the notion of allowed and disallowed as institutional facts. Moreover, the institutional fact only relates to the specific normative rule. This lead us to a refinement of proposition 11.

Definition 10. A *normative rule* n is a constitutive rule that derives $\textit{disallows}(n, i)$ or $\textit{allows}(n, i)$ from constituting facts about i .

Logical accounts of *obligation* explain certain properties of a family of concepts revolving around *subjunctive betterness* (cf. [89, 195]): they mainly express that certain entities *would be better* than others. It is this account that is relevant to planning and to plan recognition from observed behaviour.

Very well-developed in computer science & law literature is the modal *deontic logic* approach, which turns obligation, prohibition, and permission into modal operators and works from the observation that obligation and prohibition are interdefinable ($O\alpha \equiv F\neg\alpha$), and obligation implies permission ($O\alpha \rightarrow P\alpha$). What deontic logics are especially good for is capturing the uses of obligation in a positional sense, as explained above.

The modal interpretation of O is a nonempty model set of deontic *alternatives* in such a way that what ought to be is the case in its deontic alternatives. These alternatives ideal in the sense of being subjunctively best worlds. Thus if $O\alpha$ is true of an object, it means that it would be better if α were true. It does not mean that it is obligatory for anyone to do something which might make α true.

The dyadic O operator is similar. $O(\alpha \mid \beta)$ is true of an object if the condition given by β determines some non-empty class of deontic alternatives in which α is true. It means that given that β , it would be better if α were true. It does not follow from this that given that β it is obligatory for anyone to do something which would bring about α .

Orthodox deontic logics do not deal with the circumstance that the only way to make α true involves forbidden actions, that making α true is impossible, or that although bringing α about would certainly be better, it is also not *my* obligation but somebody else's¹⁰.

In the literature this discussion is known as ought-to-be vs. ought-to-do (cf. [87]): they are widely *assumed* to be mere reformulations of the same thing, even though it is well known from practical applications that they are not. The assumption that if it ought to be that α (given β) is equivalent to saying that it would be better if α were true (given β) is built into all orthodox deontic logics.

Deontic logics don't represent violations directly. *Legal* obligations do not express the goodness of something but the given that the decision to perform or not to perform the action referred to as obligatory has consequences (being violation or no violation, and the reaction to violation). Interpretation of the violation of the obligation cannot be separated from a theory of decision making and action that answers questions such as what was within the power of the decision maker, what did the decision maker intend, what did he foresee, and what did he try to do? Deontic concepts should be explained in terms of betterness, but also of, firstly, habit formation, and of choice, of beliefs, of decisions, of violations, of sanctions, etc.

Since deontic logics appear to capture the same properties of betterness as preference logics do, but without reference to the social mechanism of norms, there have also been attempts to relate the two concepts (cf. for instance [138, 268, 262, 43, 44]). *Preference logic* (cf. [267, 100, 102, 101]) captures the use of preference (here strongly \prec , and weakly \preceq) and indifference statements (here $=$) between propositions. Against preference logic the same criticism applies: preferences should be explained in terms of choices, intention, beliefs, action, and not just in terms of other preferences.

This connection to decision making and action will be discussed in section 4.7.

The notion of *subjunctive betterness* (as used by [89]) is however a very important aspect of normativity. Logical theories of subjunctive betterness can be investigated on their own, without any further presumption on how norms work, and a number of theoretical issues in AI & Law can be addressed by considering the norm as a statement of subjunctive betterness alone. Both deontic and preference logics take the same stance towards the concepts they formalize, i.e. both express the axiological concept of subjunctive betterness, and both are incomplete formalizations of the thing they claim to formalize. *Subjunctive* refers to the grammatical mood of the type of expressions covered by expressions of obligation and preference. The subjunctive mood is used for subjective expressions, indicating wish, command, possibility, counterfactuals, etc.

In the case of deontic logics this means that an obligation "given β it ought to be that α " is considered a paraphrase of "given β , α would be better than $\neg\alpha$ ".

Certain features are characteristic of many deontic logics.

Let $|\phi|$ be the set of worlds w such that $M, w \models \phi$. Central to deontology is the notion of choice:

Proposition 13. Deontic choice $O(\alpha | \beta)$: if an agent has the choice between a set of alternatives $|\alpha \sqcap \beta|$ and $|\neg\alpha \sqcap \beta|$ then the agent *should* choose an alternative from

¹⁰This clearly separates legal obligation from ends-oriented accounts of morality.

$| \alpha \sqcap \beta |$.

Clearly one set is better than the other: $| \alpha \sqcap \beta | \succ | \neg \alpha \sqcap \beta |$ if we interpret preference as a relation between sets of worlds described by a proposition, here an OWL Class as defined in section 3.4. This is however not possible in OWL DL: in chapter 6 I will give an account that models axioms constraining betterness assertions $\succeq (i_1, i_2)$ about pairs of individuals i_1, i_2 , that indirectly also covers $\succ (i_1, i_2)$.

In model-theoretic accounts of deontic and preference logics *worlds* are the traditional objects. In OWL DL we are dealing with individuals, but the account is also based on Kripke semantics.

Given the interdefinability of obligation and prohibition, the interpretation of prohibition in terms of sets of worlds follows just like obligation from the principle of deontic choice:

$$\begin{aligned} O(\alpha|\beta) &: | \beta \sqcap \alpha | \succ | \beta \sqcap \neg \alpha | \\ F(\alpha|\beta) &: | \beta \sqcap \neg \alpha | \succ | \beta \sqcap \alpha | \end{aligned}$$

One of the attractive features of the representation in the form of betterness between sets of worlds is that it produces triangles between a context of applicability β , and good ($| \alpha \sqcap \beta |$) and bad ($| \neg \alpha \sqcap \beta |$) alternatives that form a complete partition of the context of applicability. It naturally fits in a graphical representation of taxonomies, and knowledge acquisition methods like the *repertory grid* that combine development of a taxonomy and sorting along evaluative dimensions (cf. for instance [113, 110]).

Beware of interpreting specifically α as an action, and β as a situation: the alternatives may concern both descriptions of actions and situations, as long as situations can be conceived of as productive characterizations in the sense that social and legal norms only speak about situations controlled by human action.

The following are desirable characteristics of a deontic logic of obligations and permissions *as coherent normative positions* (generally based on [195, 268]):

Proposition 14. What is obligatory is permitted: $O(\alpha | \beta) \rightarrow P(\alpha | \beta)$

Proposition 15. The impossible and the meaningless are not obligatory: $\neg O(\alpha | \alpha)$ and $\neg O(\neg \alpha | \alpha)$ are axioms.

Proposition 16. There are no conflicting obligations. The obligations $O(\alpha | \beta)$ and $O(\neg \alpha | \beta)$ are inconsistent: $\neg(O(\alpha | \beta) \wedge O(\neg \alpha | \beta))$ is an axiom. Idem for $O(\alpha | \beta)$ and $P(\neg \alpha | \beta)$.

Proposition 17. . The ordering $| \neg \alpha \sqcap \beta | \prec | \neg \alpha \sqcap \neg \beta | \prec | \alpha \sqcap \neg \beta | \prec | \alpha \sqcap \beta |$ satisfies the propositions $O(\alpha | \top)$, $O(\beta | \alpha)$, $O(\neg \beta | \neg \alpha)$.

Since law *will* create conflicting obligations, a proper deontic logic which models normative positions is not necessarily also an appropriate tool for representing legal rules. For a deontic logic for legal purposes there are additional design requirements not addressed in detail here. Most importantly we have to deal with conflicting obligations.

In chapter 6 the representation of permission is derived from the following interpretation:

$$P(\alpha|\beta) : | \beta \sqcap \alpha | \succeq | \beta \sqcap \neg\alpha |$$

The translation of a permission to a statement of weak preference instead of indifference ($| \beta \sqcap \alpha | = | \beta \sqcap \neg\alpha |$), as has been argued by for instance [212, 239], is a point of contention. Indifference is incompatible with the intuitions of deontic logicians (cf. [2]); If something is obliged, then it should also be allowed. The asymmetric \succeq statement leaves room for a prohibition or obligation (\prec) and retains the information that the represented permission explicitly allowed a case and not its opposite, on which it was silent.

It is perhaps extended exposure to alethic modal logics and law that causes the intuitions. There is empirical evidence that children generally attribute an attitude of indifference to others who express a permission (cf. [170]).

A norm system that only contains $P(\alpha|\beta)$ is not a well-formed norm system at all since it serves no purpose in guiding and evaluating behaviour. Operator P serves no real purpose in a deontic reasoning system that does not allow for conflicts between norms in guiding and evaluating behaviour. The operator P cannot be understood in any other way than a superfluous utterance stating indifference towards $| \beta \sqcap \alpha |$ and $| \beta \sqcap \neg\alpha |$ if it is evaluated without other normative expressions as context. The P only becomes relevant if:

- it conflicts with an obligation or prohibition existing in the context of discourse, and
- it is used to cancel the conflicting obligation or prohibition.

The explicitly stated permission clearly has another function than the dispositional permission inferred from the absence of knowledge of a prohibition (the *silent* of Valente's interpretation). Since permissions are usually uttered with the explicit intention of amending a specific obligation, and are often found nearby in the same legislative text, it makes sense to 'localize' them to some extent by adding the asymmetry. The asymmetry is also necessary for using the representation in combination with common deontic reasoning systems.

Some broad permissions can perhaps be interpreted as symmetrical. It is for instance generally accepted that freedom of expression i.a. includes a strong permission to keep your opinion to yourself. We might add a fourth dyadic operator for freedom of choice (liberty) with some limited applications in the representation of legislation:

$$L(\alpha|\beta) : | \beta \sqcap \alpha | = | \beta \sqcap \neg\alpha |$$

Obviously the same effect can be reached by asserting two permissions.

Putting the axiological interpretation in terms of betterness, and the institutional interpretation in terms of constitutiveness of legal facts (Valente's allowed or disallowed) together results in the structure described in figure 4.2. This is the target for representation of normative rules in section 6.2, after we have laid a foundation for the representation of legal rules in terms of institutional and constitutive rules in chapter 5.

A legal norm is an a priori obligation or prohibition (i.e. not a positional one), it applies to a certain case, allows a certain case – the allowed case – and disallows a

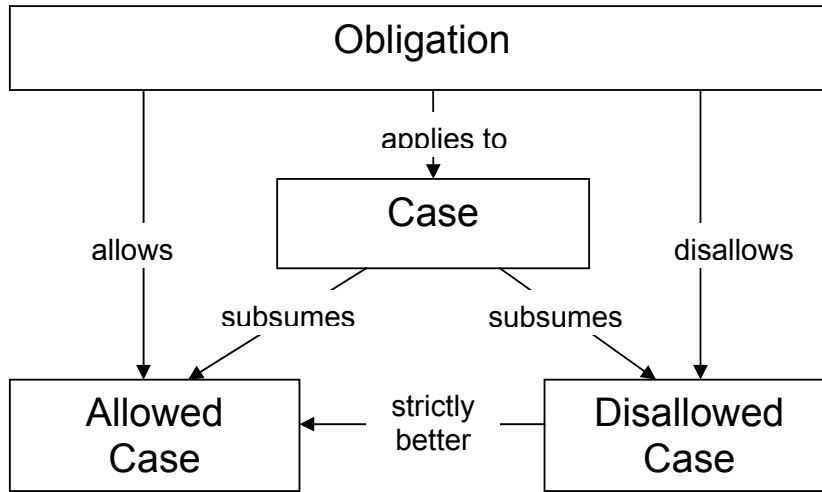


Figure 4.2. An entity-relationship diagram describing the salient structure of obligations and prohibitions.

certain case – the disallowed case. Instead of saying that the norm disallows the case, we can also say that the case violates the norm.

The allowed and disallowed case are both subsumed by the case to which the norm applies. Besides that they by definition form a complete partition of the case to which the norm applies, i.e. all cases to which the norm applies are either allowed cases or disallowed cases. This is true of the obligation and the prohibition: they are simply two different ways to put the same thing into words: the obligation explicitly mentions the allowed case, and leaves the disallowed one to common sense inference, while the prohibition makes the disallowed case explicit.

The nature of the betterness relation between allowed and disallowed case is from the perspective of the addressee justifiable from the consequences of violation of the legal norm, while the legal norm is itself an expression of a preference of the legislator.

In terms of constitutiveness and betterness norm $n = O(\alpha \mid \beta)$ can be interpreted syntactically on the following basis in OWL, as will be shown in section 6.2 later in this book:

1. the three relevant *cases* correspond with OWL classes, β , $\alpha \sqcap \beta$, and $\alpha \sqcap \neg\beta$;
2. normative rule n is an individual, which disallows certain individuals, modeled by OWL assertions $\text{disallows}(n, i)$, etc.;
3. the normative rule is however also a nominal concept $N \equiv \{n\}$ of which we can assert terminological OWL axioms such as that of all things disallowed by n the proposition $\beta \sqcap \neg\alpha$ is true: $\{n\} \sqsubseteq \forall \text{disallows}.\beta \sqcap \neg\alpha$;
4. normative rule n orders certain individuals i_1, i_2 by betterness, modeled by OWL assertions $\succ_n(i_1, i_2)$, etc.; and
5. we can assert terminological OWL axioms about this betterness interpretation

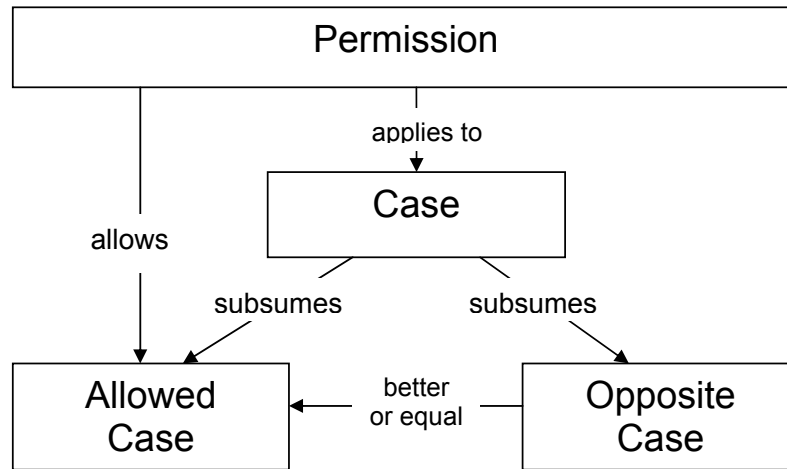


Figure 4.3. An entity-relationship diagram describing the salient structure of a permission. Obligation and prohibition are subsumed by permission.

of n , such as that in situations in which the proposition $\alpha \sqcap \beta$ is true all better or equal situations are situations in which $\alpha \sqcup \neg\beta$ is true, or $\alpha \sqcap \beta \sqsubseteq \forall \preceq_n \alpha \sqcup \neg\beta$.

The permission, shown in figure 4.3, is only slightly different. The permission allows something, but it doesn't disallow anything. The logical complement of the mandated case is here simply called the opposite, following existing practice by some authors (cf. [266, 156]).

A desideratum for the OWL representation in section 6.2 is that the obligation is subsumed by the permission. In addition, one can easily imagine a more generic subsuming concept that just “applies to a case” for constitutive rules in general, and we will indeed develop such a notion in section 5.2.2.

The analysis of normative rules cannot, however, stop here. Obviously, the function of normative rules cannot be considered without appealing to the notion of *agency*. The normative rule does not merely express relative goodness of certain states of affairs, but also directs agents (not) to bring them about.

The betterness interpretation as developed here orders worlds, or in OWL terms individuals. The question is however what concept describes the domains of the \prec relation. Is it a situation, an action, or a plan?

4.7. Agents and Action

Agency is not a subject that should be considered part of AI & Law: most accounts of action and planning are written outside the field of AI & Law, and are often designed with automated planners – that construct a plan to achieve a certain goal state from some initial state for a domain with certain properties – in mind (cf. generally [25,

29, 107, 58, 85, 4, 178, 3, 137, 213, 261]). From a Semantic Web perspective it makes sense to combine a generic mechanism for deontic reasoning with a theory of action, events, or plans designed for some other purpose.

These theories can vary considerably in how they deal with time, with change, etc. Still, we cannot satisfy ourselves with an account of deontic reasoning that abstracts the thing being qualified to some opaque proposition. Doing so gives rise to odd analyses of what norms mean for human action.

We have to make at least some assumptions about agency and change, and relate these to normative positions like being in a state of obligation to do something. An important concept in this respect is the distinction in legal theory between ought-to-do and ought-to-be representation, discussed in the next section.

In a nutshell, this chapter tries to account for two major uses of legal rules:

Planning to perform a task one sets oneself; to bring about beneficial legal facts, while avoiding detrimental ones; and

Situation and action recognition to infer one's own legal position, and what legal facts others bring about, are going to bring about, and unsuccessfully attempted to bring about.

Planning involves generating and comparing alternative plans, while situation and action recognition involve generating and comparing explanations. The planning perspective also prominently involves time and change.

The purpose of this section is to point out some constraints that the conceptualization of intelligent behaviour implicit in the rules themselves impose on the ways in which we can conceptualize agency and change. It is not intended to set a golden standard for the conceptualization of planning and change itself.

4.7.1. To Be or To Do

A central subject in the representation of obligations is the *ought-to-be* versus *ought-to-do* debate (cf. [74]): does the obligation prohibit a state of affairs or the bringing about of a certain type of change? What role does intention play in this? If obligation cannot be reduced to mere subjunctive betterness, as for instance [74, 89] argue, we have to account for what an obligation tells us to do.

Important for knowledge engineering is a derived issue: is there a logical relationship between these two possible formulations that allows us to translate between them?

The key question for this chapter is however: to what kind of objects are the constitutiveness and betterness interpretations applied? Situations, actions, plans, or something else?

[265, 184, 152] argue for a distinction between the following subproblems of enforcing norms:

1. the norm proper, which simply derives $\{allowed, disallowed, silent\}$ from a situation description,
2. the causal connection between the conduct of a person and the qualified situation,

3. the fault implied by the conduct of that person, and
4. the attribution of responsibility, usually but not always to that person¹¹.

The discussion in [152] gives the best overview of the mechanisms involved, and stresses that causation itself is the outcome of ampliative inference explaining the occurrence of an event by a process explained by an earlier or simultaneous event. This causal model is an abstraction of underlying processes. Agent causation simply extends this to mental processes.

Since the end result is an explanation in terms of *agent causation*, or the lack of it, and the legislation itself can be understood in terms of agents “bringing about” i.e. causing certain occurrences, this mechanism of explanation is of minor concern here: legislation typically addresses the explanation, and not the underlying processes.

There are two markedly different “styles” of legislative drafting that are clearly both encountered in legislation: the productive style, which focuses on unwanted results, and the behavioural style, which focuses on ease of recognition of unwanted behaviour. Note that the productive style does obviously attribute it to behaviour: the ought-to-be style describes the goodness of the product of our actions. It deals with states of affairs generally caused by humans, but omits specification of their *role* in it. The ought-to-do style directly describes the action. As pointed out in section 4.1, action recognition presumes recognition of the scripted role the agent assumes and ascription of intent.

Generally speaking, in settings involving interaction between a human and a machine, or a design of one, it is easier to give a productive characterization of what the legislator doesn’t want: e.g. *the carrying capacity in kg of the elevator should be at least 100 kg times the number of people you can reasonably fit into the volume of the elevator*. It is of course possible to give a behavioural characterization of the same constraints, but this approach can be clumsy and unintuitive.

In these settings we are usually dealing with what is often called a *classical environment* that is fully observable, deterministic, with a finite number of discrete alternatives at any point, and where the system or design being manipulated is only changed by the planning agent. If the system or design violates a rule, it is its owner, operator, or designer that did it.

In environments where one or more of these features are not present there is a less predictable interaction with the environment. In this kind of setting we have failed attempts, unforeseen products of acts, and ambiguity about who caused something. Because the intention of the law is to a large extent to influence people’s behaviour, it is more practical from a compliance and enforcement perspective to give behavioural characterizations of behaviour and to point directly at the seller, the driver, etc. Behavioural characterizations are generally speaking easier for the addressee to learn and adopt, and violations of norms with behavioural characterizations are often easier to recognize and prosecute.

Because the abstract goals of the legislator often have a productive nature (for instance reducing the number of people that die in traffic, or reducing conflicts over the possession of goods, or raising money for some purpose), behavioural norms often

¹¹For instance liability of a parent for the conduct of a child.

feel like approximations of what the legislator really wants of the addressees of the legislation. Behavioural characterization increases precision in enforcement and compliance, by making the things to which it applies easier to foresee and recognize, but at the expense of precision in describing the policy goals of the legislator.

Productive characterization creates problems for the addressee if the addressee does not foresee the consequences of his actions reliably in an environment that is only partially observable and stochastic. This is for instance often the case if other human beings are involved. If the addressee perceives a lack of control over the situation, enforcement of legal norms has little effect.

The full mechanism in [152] appears to be conceptualized *explicitly* in the sources of law only in cases of grave importance, for instance responsibility for someone else's death. Here the legislator not only addresses unintentionally causing death, but also failed attempts at causing it.

The fundamental difference between an ought-to-do characterization and an ought-to-be characterization is in the omission in the ought-to-be norm of an action description and an agent identified by role (e.g. the driver coming from the right, the thief who is taking a good that wholly or partially belongs to another). The problem with the ought-to-be characterization is therefore mainly in identifying the agent who *caused* the state of affairs or failed to prevent it from occurring, while in ought-to-do characterizations this is merely a matter of action recognition.

While it is the state of affairs that is considered to be disallowed (following [152]), it is only an agent who will be, for his actions or his failure to act, held responsible for its occurrence. For the agent this means that when he foresees that his action may end up in an prohibited state of affairs, the problem is whether responsibility for the state of affairs will be attributed to him or to someone else.

This is not trivial. It is not necessarily the case that the one held responsible for an undesirable situation should also be the proximate cause of the situation in which the violation arises. You can't for instance simply reduce ought-to-do to ought-to-be. An example: Imagine someone offers exclusively to you a painting for sale, with the condition that the offer is valid until next Tuesday. You have a claim-right to buy the painting, in the sense that the counterparty has a duty, towards you, to sell you the painting *if* you accept the offer before next Tuesday: the behaviour prohibited consists of not selling it to you while you have accepted the offer. The counterparty then sells the painting to someone else, before Tuesday. After he sold it, you accept the sale. You are now in the situation prohibited by the obligation, and your conduct is the "proximate" cause of the situation by expressing your intention to buy the painting, but not the legally relevant cause. You are not responsible, because the duty wasn't yours but the seller's: the fault is with the seller.

In this case, for you, the obligation does *not* mean anything in terms of the betterness interpretation. While it is possible to claim that it is the occurrence of the situation that is allowed, its interpretation in terms of obligation cannot be based on situations.

Proposition 18. The mere fact that a *situation* appears prohibited does not necessarily mean that a plan that brings it about is a bad choice for a specific agent. It depends on whether you will be held responsible for it. If however the agent conceives

of the plan as an alternative in a planning problem, and the plan is based on the assumption that other agents will behave *normally*, i.e. following the prevailing norms, then it is very likely that choosing to execute that plan is a bad choice.

Even if we consider the attribution of responsibility to agents as unproblematic, and subscribe to a simple deterministic model of states and transitions, the relation from behaviour to product is not always immediately obvious. d’Altan et al. (viz. [87]) for instance offer the following hypotheses for the relation between ought-to-do characterizations and ought-to-be characterizations:

1. A state is obligatory if and only if it is the result of an obligatory action.
2. A state is obligatory if and only if all the actions that lead to the state are obligatory.
3. A state is obligatory if and only if it is forbidden to undo it.
4. A state is obligatory if and only if all the actions that are necessary¹² to bring about the state are obligatory.

None of these alternatives is without counter examples, as [87] point out. It is certainly not obvious that all ought-to-do norms can be translated to ought-to-be ones; The other way around is more plausible.

There is another argument in favour of actions as the subject of obligations.

In addition to behavioural and productive characterization, the *attempt* forms a third category: characterization of display of undesirable intent in itself, regardless of the form in which it is executed successfully. This category is problematic, and therefore usually avoided if possible. There is a strong taboo on “ruling over the consciences” of citizens, and a practical reason to avoid it is obviously that it requires mind reading on the part of the court.

Still legislation with a strong moral tone, in particular criminal law, does involve a certain degree of “mind reading” when dealing with things like attempts (to kill someone), premeditation, or the distinction between intentional and unintentional killing. Because the court cannot actually read minds, it has to infer apparent intention from actions. To do this, it has to judge what the agent foresaw as the consequences of his actions, and this requires ascribing knowledge to agents. This isn’t really helpful, as it just moves the mind reading problem to another aspect of the operation of minds.

The solution is unsurprisingly found in the ambivalent nature of norms: it is possible to set standards for what reasonable people *should* know. There is also a related legal concept: the knowledge of the *man on the Clapham Omnibus* (cf. [203]).

How to realize this in KBS is hardly a problem: it is usually taken for granted that the dispositional belief set of some agent (cf. section 3.3) is the logical closure of its individual beliefs with some universal body of shared knowledge. The idea that we share a body of knowledge is an assumption behind ontology modeling. It is indeed a necessity that people who understand each other share a body of knowledge. In addition to shared ontological knowledge, people must also be assumed to share knowledge about what normally happens, as alluded to in section 4.5.2.

¹²I.e. performance of the actions necessarily leads to the state, and non-performance of the actions necessarily prevents the state.

The attempt is an unsuccessful execution of a task. The attempt is therefore an action according to the principles decided in in section 4.3, while the situation that results from it can be left completely unspecified.

One can however also apparently be held responsible for one's non-action. Of this scenario we know two variants:

1. One has an obligation to act and doesn't during some period; and
2. something that is disallowed happened because one did not act to prevent it (*negative causation*, cf. [184, 152]).

The first case is a standard feature of any combination one makes of a deontic logic and causation (addressed in 4.7.2).

The second case only happens if one is presumed to have been aware of the possible undesirable occurrence and didn't attempt to prevent it, i.e. demonstrated a lack of intention to prevent the occurrence through one's choices. Assuming that agents are always executing some task they set themselves, the undesirable occurrence was "on the radar" during the planning of that task and therefore part of the conceptualization of the situation for which one made a plan. Letting the undesirable occurrence happen is therefore part of the task specification and therefore intentional: here too one is in the end responsible for what one did do.

The *action* is what links undesirable changes *to* some situation, undesirable changes *from* some situation, undesirable continuation of a situation, and unsuccessful attempts directed towards these. The subject of the norm is therefore naturally the *action*.

Proposition 19. The subject of a normative rule is an *action*.

4.7.2. Agent Causation and Time

A familiar framework for explaining how stuff is brought about is *event calculus* (cf. section 3.5.3). By acting, an agent either initiates or terminates certain *fluents*, for instance a:

1. *state* of an object; for instance the state of being married, or the state of obligation of the counterparty to deliver certain goods,
2. *situation*, i.e. a configuration of objects; for instance two people being married to each other, or
3. perdurant *object*; for instance a marriage, or a private company limited by guarantee.

Sartor in [244] (table 2) lists a number of *normative conditional* patterns based on the distinction between initiation (action initiates fluent), termination (action terminates fluent), and *emergence* (fluent entails fluent) where the consequent is always a legal fact. These facts are further classified into normative positions (e.g. my obligation to terminate the traffic violation initiated by the traffic participant overtaking me), qualifications (e.g. being a married person), and the existence of legal things constituted by something else (e.g. the marriage that exists between me and my wife).

All of these are obviously constitutive rules. The difference between qualification (married) and existence (a marriage) in this sense is primarily a difference of descriptive

vocabulary and not of substance: reification of relations in a knowledge representation and nominalization of verbs or adjectives in natural language are motivated by technical and presentation considerations, not ontological ones.

Initiation and termination are generally encountered in real life as pairs – a *change* operating on the same substrate, which can be an object or some configuration of objects (a situation, scenario, plan, system, etc.):

$$\begin{aligned} \text{CHANGE} &\sqsubseteq \exists \text{initiates.STATE} \sqcap \exists \text{terminates.STATE} \\ &\sqcap \exists \text{substrate.OWL:THING} \sqcap \exists \text{extension.PERIOD} \\ \text{STATE} &\sqsubseteq \exists \text{substrate.OWL:THING} \sqcap \exists \text{extension.PERIOD} \end{aligned}$$

The change terminates a state and initiates a state of the same thing. State (for “stative”) is taken to be the generic type subsuming the previous categorization of ways in which fluents are described.

Observe that the combination of the deontic operator ‘it is obligatory that’ (O) from section 4.6 with a simple intentional action operator ‘agent i brings about’ (E_i) would give rise to four atomic types of obligations (cf. generally [250] for a discussion):

1. it is obligatory that i brings about C ;
2. it is obligatory that i brings about not C ;
3. it is obligatory that i does not bring about C ; and
4. it is obligatory that i does not bring about not C .

As noted earlier, action presumes the recognition of plan execution by i . Let’s for the moment set aside the problem of how i is going to (know whether i can) bring about C , and how one determines whether i did, didn’t, attempted it, etc. The difference between ‘ i brings about not C ’ and ‘ i does not bring about C ’ requires some further analysis.

Obviously bringing about not C is only meaningful if C is the case: we can reformulate this to *terminating* C (whenever it may arise), while not bringing about C is about *not initiating* C , and C is a state, some occurrence that occupies a time interval or *period* on the canvas of time, that is initiated by some change and terminated by another change. Similarly, we can talk about *not terminating* C (whenever it may arise), and about *initiating* C .

The agent can be held responsible for the fluents he initiates and terminates: this is the easy case. He is also responsible for the fluents he was able to initiate or terminate but didn’t. In the Netherlands it is for instance permitted for a pair of bicyclists to ride next to each other. If a third traffic participant overtakes them, this traffic participant initiates a traffic violation, against the obligation to keep to the right. At the same time jurisprudence shows that each one of the three traffic participants can be held responsible for the traffic violation, since each of them is able to terminate the traffic violation. Current jurisprudence apparently interprets the obligation as pertaining *both* to initiation and to termination, but this does not follow from the phrasing of the involved obligation and permission which simply addresses the stative pair of bicyclists riding next to each other (in other words, an “ought-to-be” characterization).

This should be kept in mind when interpreting ought-to-be descriptions: does it pertain to initiation, to termination, or to both.

Normative rules are usually about bringing some change about. There is a general pattern explaining how legislators will typically phrase normative rules:

- O : some change \succ the continuation of the sketched situation;
 F : continuation of the sketched situation \succ some change; and (less markedly)
 P : some change \succeq continuation of the sketched situation.

Norms are only concerned with initiating and terminating C , and do not have any bearing on any actions that neither initiate nor terminate C . None of these formulations tells me anything about bringing about unrelated fluent D acting on another substrate.

Only in the case of directives to *not* initiate or terminate C we can determine directly whether a possible violation has taken place: it takes place if we *do* initiate/terminate C , i.e. when a change occurs. Also if the obligation specifies that C must be terminated *before* D is initiated, violation takes place if D is initiated while C is still the case. Directives to initiate or terminate C are however assessed against *deadlines* (cf. generally [97]), and only the passing of the deadline gives rise to violation. The obligation to initiate C may for instance well be irreconcilable with a course of action that consists of first initiating D and only then C , but this does not have anything to do with the position of the initiation of D relative to the initiation of C in time.

The implicit deadline leads to a bit of a problem if we consider this in the context of things allowed and disallowed by a norm. Take as an example the obligation that i terminates C ; This means the following in terms of normative qualifications:

1. i terminating C is allowed;
2. the continuation of C by i (beyond some unspecified deadline) is disallowed.

The representation of case 2 is obviously a bit problematic, because it addresses the non-action of ‘doing nothing about C ’. Exact identification of the intended logical complement of terminating C is hard.

Let it for instance be prohibited to drive with a vehicle without functioning headlights between dusk and dawn. In practical terms this is supposed to mean that 1) one has to check whether one’s lights function before one starts driving, and 2) that one has to switch the lights on before dusk. One may discover a headlight is broken while on the road after dusk sets in¹³. One is now in the state which is disallowed, without consciously having decided to initiate it, and one is obliged to terminate it.

Leaving the road to repair one’s headlights is usually going to involve some more driving on the road, to find a suitable place, to buy the light bulb at a gas station, or to go to a garage. Since no deadline conditions are given, intuitively, only action betraying the *intention* to continue the violation would *clearly* constitute a violation. If one gets pulled over by the police one can get fined, but it is for instance unclear *how many times* one can get fined for substantially the same fact: the fact that one is fined for not terminating C does not mean that the obligation disappears. It would however be unreasonable and unfair for the police patrol to stay around for another

¹³Since lights tend to break when switched on, the earlier check doesn’t help.

opportunity to give a fine and they generally don't. It is also reasonable to not give a fine if the driver can repair the light directly. In other scenarios other factors determine enforcement decisions: enforcement decisions depend on enforcement policy and, in the absence of such policy, private considerations.

Proposition 20. It is not possible to determine whether an obligation to bring about a change has been violated, unless one sets a deadline. Obligations to bring about a change *before* some other change present no problems. Reasonable enforcement policies are based on ascription of intention to behaviour while the obligation exists.

In a logic of agency which treats actions as instantaneous changes of state occurring at discrete time steps (i.e. Hoare logic, dynamic modal logic), it is natural to consider the next state as the implicit deadline. Combined with the presumption that, given full knowledge of the initial state and a known menu of possible actions, the state following execution of an action should be predictable with absolute certainty, we can axiomatize *obligation* completely in terms of agency, like the proposals in [87]. We should keep in mind, however, that the verity of these axioms depends on the conceptualization of the planning domain, and not on the properties of the concept *obligation*: it doesn't shed any light on the nature of obligation.

Representing a normative rule involves some common sense judgment on the part of the knowledge engineer. How complicated a representation needs to be depends i.a. on the following:

1. Does the normative rule demand that one displays the intent to comply, i.e. does it take into account the possibility of failure of attempts?
2. Does the rule require continuation or change?
3. If it describes a change, is it phrased as an initiation or a termination?
4. If it requires a change, is there an indication of a deadline or a *before* condition?

We can identify a number of ingredients that should be covered, although we cannot solve the deadline problem. For the representation of normative rules, apply the following principles:

1. Normative rules apply to action;
2. Action causes changes;
3. Action, whether conceived of from the perspective of the actor or recognized by a bystander, is situated, i.e. takes place in a situation; and
4. A situation consists of a limited number of participants, it is not a (state of the) world in the logical sense (cf. section 3.4), but a conceptualization of the context in which the action takes place.

The direct connection of *situation* to action is inspired by [261, 15], and the representation of discourse context in general in computational linguistics (for instance [227]). It deviates significantly from AI planning literature, which is largely based on the assumption that 1) world and situation are the same thing, and 2) that situation + action results in a new "situation". This works well for planning, but is not terribly realistic.

Instead an agent reconceptualizes the situation in the context of planning the next action: since this is done in the knowledge of one's previous action, and one may be

acting still in the same role, it is not surprising if the new conceptualization borrows participants from the old one. Since we have no particular application in mind, it is wise to take the conceptualization in the source of law at face value.

Actions are situated, performed by an actor, and cause changes:

$$\text{ACTION} \sqsubseteq \\ \exists \text{situation.OCCURRENCE} \sqcap \exists \text{causes.CHANGE} \sqcap \exists \text{actor.AGENT}$$

As always, this is the simplest possible conceptualization, ignoring issues like decomposition of both actions and situations. The obligation to repair a broken headlight – without a clear enforcement policy – covers the following three concepts:

$$\exists \text{situation.BROKENHEADLIGHT} \quad (4.1)$$

$$\exists \text{situation.BROKENHEADLIGHT} \sqcap \forall \text{causes.}\neg \text{HEADLIGHTREPAIRED} \quad (4.2)$$

$$\exists \text{situation.BROKENHEADLIGHT} \sqcap \exists \text{causes.HEADLIGHTREPAIRED} \quad (4.3)$$

The obligation applies to 1, disallows 2, and allows 3. Disallowed is any action performed in a situation which involves a broken headlight and does not involve fixing it. To find an acceptable resolution to deadline issues we would have to address decomposition of actions, for instance:

1. Disallowed is any action performed in a situation which involves a broken headlight and does not involve fixing it, and is not part of an action that involves fixing it.
2. Disallowed is any action performed in a situation which involves a broken headlight and does not involve fixing it, and is not an execution of a task that involves fixing it, or of a subtask of a task that involves fixing it.

These kinds of refinement are either based in common sense or in jurisprudence. The scenario of negative causation, although superficially similar, is different. In this case a *future* change is implied by one's conceptualization of the situation; An action likely to be deemed wrong is for instance:

$$\exists \text{situation.IMPENDINGDISASTER} \sqcap \forall \text{causes.}\neg \text{PREVENTDISASTER}$$

It may seem strange to include expectations about the future as part of the situation one acts in, certainly if one conceives of situations as is usually done in AI planning, but it is no stranger than including expectations about the past: if we see someone coming towards us with a blood-covered chainsaw, our explanation of this event will cover both the past (to explain where the blood comes from) and the immediate future¹⁴. The notion of normative position, which can also be conceived of as being part of the situation in which one acts, is also often essentially about *future* possible consequences of one's actions.

¹⁴The explanatory power of a scientific theory of the past, like evolution theory, is for instance in its prediction of future observations about the past, i.e. missing links yet to be found.

4.7.3. Positions and Power

Interesting are the positional uses of obligation, called *normative positions*. The most obvious of these is the state VIOLATION initiated by an action that is disallowed:

$$\exists \text{disallowedBy.NORM} \sqsubseteq \exists \text{causes}.\exists \text{initiates.VIOLATION}$$

Other familiar ones are systematically related to corresponding normative rules, but are secondary to the reading of normative rules as being constitutive of the qualifications $\{\textit{allowed}, \textit{disallowed}\}$. Given the existence of a norm “given α, β it ought to be that γ is initiated” and me being in a context that entails α , I have a contextualized obligation to initiate γ if β .

The relation between this type of concept and the original norm is similar to “being in a position in which I have to get ingredient a ” when I am preparing a dish that requires ingredients a, b, c and I have only b and c . If I believe I cannot obtain ingredient a , my position would be “not being able to make the dish”. This is a planning concept (and a fluent that exists only in my mind, or that of others) that has no direct bearing on the recipe.

The same applies to norms: the contextualized obligation is not the original norm, but a state relative to the norm that only acquires special meaning by adding a theory of what I am able to bring about and what I am trying to do. It is possible but not very useful to automatically derive positional obligations. In scripts, as planning concepts, they have their uses as a shorthand for a compound action, for instance an action in a situation α that brings about the initiation of β but not of γ is not allowed by n :

$$\begin{aligned} \exists \text{situation}.\alpha \sqcap \exists \text{causes}.\exists \text{initiates}.\beta \sqcap \forall \text{causes}.\forall \text{initiates}.\neg \gamma \\ \sqsubseteq \exists \text{disallowedBy}.\{n\} \end{aligned}$$

While normative rules apply to action, positional concepts also often follow the state constitutes state pattern, for instance:

$$\text{FIREALARMSIRENWAILS} \sqsubseteq \exists \text{constitutes.FIREALARM}$$

Positional concepts play a role mainly as a description of the result of the application of a constitutive rule. They also occur as conditions to constitutive and institutional rules. A special type of condition is the (legal) power or competence. In common law jurisdictions the term power appears to be preferred (cf. Bentham, Hohfeld, and Hart), while civil law jurisdictions commonly call it competence (cf. [69]).

The *power* to bring certain things about refers to a certain legal qualification in the constituting base that applies to the agent that gives the ability to bring about a legal fact. “Only parliament declares war” for instance means that being the parliament is a necessary condition for the initiation of the state of war. It is distinguished from other similar potentialities (“Only people who have money can buy stuff”) by the fact that the qualification is a fact within the institution.

The agent is able, competent, or it is within his power, or he has the potential, to bring certain things about. Almost every agent is for instance able to open and close an unlocked door, but most agents are not able to open a locked one. This is *ability*, and was already indirectly addressed in section 4.5.2: to have the ability to do something in a certain situation is to meet the conditions for filling an agent role. Some agents are able to marry a pair of other agents, but most are not. Since marriage is an institutional status (nowadays), this is (legal) *power* or *competence*.

The abilities and powers of a person are limited to the agent roles for which he meets the necessary conditions *in the situation*. Agent roles are *stative*, and in that sense part of the situation in which an action is performed: it is however common to distinguish states that specifically pertain to the agent (e.g. police officer) from those that do not (e.g. dark), which are rather conceived of as conditions for executing the task, and relatively shortlived ones (e.g. unarmed) from longlived ones (e.g. police officer).

Some authors have tried to explain power-conferring rules as a variety of normative rules (notably Von Wright, Bentham, Cornides, Kanger, Lindahl, and Kelsen), while others explain them, as I do here, in terms of constitutive rules constraining institutional action (Hart, Ross, Searle, and Bulygin in [69], which is also the source used here for the classification of authors by the position they take on this subject).

As pointed out in 4.4, we usually speak of a power if the legislator intended to create a way for a certain class of agents to bring about the legal fact. Attribution of power is captured by standard constitutive rules, if these have legal conditions that pertain to the agent role, i.e. for instance:

$$\text{MARRY} \sqsubseteq \exists \text{actor.HASPOWER TOMARRY} \sqcap \exists \text{causes}.\exists \text{initiates.MARRIED}$$

The marry action is performed by an agent with the power to marry and causes the initiation of the state married.

4.8. Interaction Between Agents

The mere publication of a formal representation of set of constitutive rules does not make a functioning legal system. Implicit in the act of legislating is the implied threat to *react* if violations take place. The *Functional Ontology of Law* in [265] therefore posits a special category of legal knowledge that covers this function: *reactive knowledge*. The institution however does not depend on an ability to “punish” itself: it reacts to violations by either obliging, permitting, or empowering others to react. Essential to the functioning of the institution is that there is a critical mass of agents who are willing to follow these directives or use these means provided to punish others.

Tort law for instance depends on payment of damages to function as a sanction of the injuring party and an incentive to report the violation for the injured party, and police officers follow directives because they are paid to do so with taxes collected (using the law) i.a. for the benefit of having police officers who react to violation of criminal law directives.

The following is for instance a somewhat simplistic constraint¹⁵ on the power to arrest:

ARREST \sqsubseteq \exists situation.OFFENCE \sqcap \exists actor.POLICEOFFICER

Arrests can be performed in a situation of offence, and by a police officer.

The institution works if it is possible to arrange normative order in such a way that reacting to one agent's violation of the law is in the interest of another. Very often a role is given to the "victim" of the violation, but it is important to note that no appeal to *restorative justice* (i.e. the notion of norm violation as an act against another individual, who should receive some kind of restitution) is needed to explain why the victim usually takes on this role. Very successful and efficient (but morally unappealing) enforcement systems have for instance been based on the principle that the agent who brings the case (and often also the suspect) in for adjudication receives property of the agent violating the law after he is executed or enslaved. The legislator counts on self-interest and commitment to the public interest.

Reaction of course subsumes not just punishment, but also the mere reporting of the case, gathering evidence, optionally apprehending suspect(s)¹⁶, adjudication, and any other supporting activities. It is customary, for obvious reasons, to separate the role of adjudicator (who authoritatively interprets the law and applies it to the case) from the reaction *per se*.

As pointed out before the whole system depends on nothing but constitutive rules, but these rules are arranged in such a way that the addressed population of agents makes enforcement work by organizing punishment of each other, and collecting the necessary funds for employing professionals to make the system run smoothly.

A simple example of this mechanism can also be seen in international paralegal frameworks like ship *classification societies*: ship owners voluntarily pay a subscription fee to a classification society, which in turn uses the collected funds to draft rules and send surveyors to ships to enforce compliance with these rules. The rarely used sanction consists of suspension or loss of "class", which is a problem for the ship owner because port authorities generally demand classification with a reputable classification society. authoritative interpretation of the rules in final instance is left to some court.

[265] is obviously right in noting that there is a functional distinction between norms and other rules that regulate normal interactions and those that regulate reaction sanctioned by the legal system. Contrary to some of the other distinctions made there (for instance between world knowledge and normative knowledge), reactive rules are however in no way distinguishable from other categories of rules by their content. It is their intended function in decision making which identifies them. Another functional category in [265] to which this observation applies is *metalegal knowledge*, which contains the rules of adjudication.

Proposition 21. The judge, police officer, parliament, etc. just follow the rules of the institution. There is no distinguishing *formal* criterion that sets apart reactive rules from other rules. The distinction is purely functional, and often not easy to make.

¹⁵Offence is simplistic.

¹⁶This step only becomes important if the foreseeable punishment justifies running away.

4.8.1. Transactions and Interests

The previous section strongly suggests that the *interests of others* play a large role in deciding what you can do, and that the law is usually set up in such a way that your violation of an obligation may give rise to another person's permission or power to do something to you which is in their, and contrary to your, interest. An example: your obligation as a driver to yield at intersections for drivers from the right, is correlative to their right of way. This may mean two things:

1. Your violating against the obligation may be a precondition for some action the other driver can take against you. Typically, the violation of a traffic rule would make you for instance liable to paying for the damage if an accident happens. This is a specific position which is associated to him being the victim of your action, and it is in his interests to use this position.
2. More important in daily life, the other driver may proceed with the *legitimate expectation* that you will yield. The other driver will typically prefer driving on over stopping, and will expect you to stop in order to comply with your obligation.

These two options need not be correlated although they often are. A police officer can *react* to the traffic violation by fining you, by virtue of being a police officer, without being the victim of your action. The victim of your action is not in a legal position to react to you if no accident happens.

Most normative positions are two- or more-sided in the sense that there is or are possible victims of your noncompliance. But how do we decide that some bystander is a victim of your noncompliance, or a beneficiary of your compliance?

A number of implicit interests are involved in the example, that can be represented in analogy with the subjunctive betterness interpretation of norms:

1. The other driver usually prefers not stopping over stopping, and therefore you behaving normally over you not behaving normally.
2. Both drivers prefer no accident over an accident.
3. The other driver usually prefers you paying for the damage over you not paying for the damage in the case of an accident.
4. You prefer not paying for damage over paying for damage.

The legislator is making certain assumptions about the preferences of typical bystanders depending on their agent role *agent*. In legislation these usually remain implicit.

Imagine an agent who is solving decision problems all the time in a situation involving other agents who might do unpredictable things based on their own decisions. In his decisions he has to take into account his legal position. In the first place, this means that there are certain things that he cannot efficiently bring about, because they are disallowed. But since these norms also apply to the other agents, there are also certain things he *can* efficiently bring about by taking into account their obligations, and assuming they behave normally. Crossing intersections while somebody else is coming from the left for instance, in the expectation that they will yield, or sending

someone the product they ordered, in the expectation that they will make a payment upon receiving it.

Proposition 22. The law does not only constrain action: It also creates abilities.

The true real life beneficiary of an obligation is the one who has a preference for someone else meeting their obligations. The other driver on the intersection is therefore typically the beneficiary of the obligation to yield. This is to be distinguished from another function of explicit assignment of the role of beneficiary by the legislator: In traffic rules the other driver will not be explicitly recognized as the holder of right correlative to obligation, and in criminal law somebody may be explicitly recognized as beneficiary without actually having an interest in it: think for instance of euthanasia as murder. In many cases this fiction is used to attribute *potestative right* to the beneficiary to take some action following noncompliance, while in other cases – like violent crime – the notion of a victim is purely fictional, and intended only for the justification of the law itself.

Proposition 23. The legal institution ascribes preferences to persons depending on the agent role they adopt.

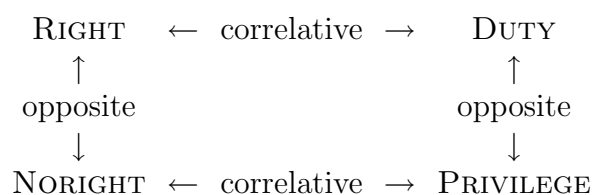
As pointed out in section 4.4 and [244] in general terms, and explained in detail in [243], there is firstly a difference between things one does in one's own interest, and things one does in the public or communal interest, and secondly a difference between powers one can exercise in one's own interest and powers one can only exercise in the public interest.

This discussion of obligation and correlative right extends to Hohfeld's jural relationships, of which duty-right is just one form. Hohfeld's jural relationships can only be understood by appealing to the preferences and intentions that typically come with adopting a certain agent role.

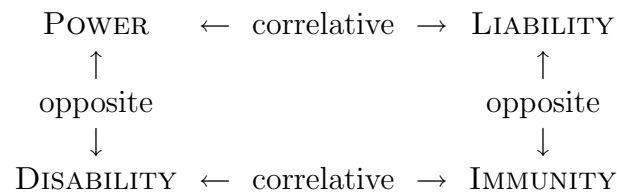
4.8.2. Hohfeld's Conception of Rights and Powers

The principal aim of Hohfeld's work (in [156]) was to clarify *jural relationships* between parties. Hohfeld presents us with an analytical scheme which distinguishes four different categories of jural relationships between parties and makes a number of analytical distinctions between various legal positions.

Hohfeld asserts that there are eight such entities: right, privilege, power, and immunity along with their respective correlates of duty, no-right, liability, and disability. In addition, each fundamental conception is a *jural opposite* to another: privilege, right, power, and immunity are the respective jural opposites of duty, no-right, liability, and disability. These form two squares. The first one is the *deontic square*:



The second one is the *potestative square*:



Hohfeld's system allows for a more fine-grained distinction between patterns we observe in legislation, and generally turns out to be more useful, in particular in private law, for recognizing and classifying patterns in legislative language than the basic deontic categories and constitutive rules in general. It is useful in this context because it covers nearly the whole terrain covered in this chapter, and has been used before in the legal knowledge engineering field (cf. for instance [5]).

Hohfeld conceptualization is not without its critics, however. Hohfeld's names are sometimes slightly confusing. The name privilege for instance suggests more than it actually means. The privilege can also be called plainly permission, right, or liberty. Also right has at times be renamed to claim or claim-right to distinguish it from other uses of the term right. Since section 4.7.3 distinguishes powers or competences from abilities, disability is also an unfortunate name. Ability/disability for "brute" ability, and competence/incompetence (or power/impotence) for institutional ability make more sense in the context of this book.

Also liability is likely to confuse people and has been criticized. It is for instance called subjection in [244]. Hohfeld's liability means that you are exposed to the exercise of a power: you (have to) accept exercise of the power to change legal reality. In private law we encounter a similar concept of the same name, but with a more specific meaning. But rejection of a name like liability for this reason makes it impossible to develop a jurisdiction-independent vocabulary for legal reasoning.

Another problem with Hohfeld's categories is according to some (cf. [135]) that they fail Ockham's test and are therefore not fundamental: there are multiple ways in which his 8 fundamental conceptions can be reduced to each other. This however also applies to the three deontic categories. Knowledge engineers should certainly be willing to take advantage of attempts to interrelate a number of well known concepts from the field of law. There is no need for them to be *fundamental* in Halpin's sense.

Another problem is with the notion of opposite. The opposite pairs seem to have subtly different meanings, which are not quite captured by simply negating the opposite. This is yet another indication that these patterns are hardly *fundamental* legal patterns.

Hohfeld's relationships distinguish between normative positions and other non-normative legal positions, between the competence and incompetence to play a certain agent role, and therefore to cause a certain change of position, and between the obligation to cause a certain change of position or the absence of such an obligation, and most importantly, between the one who acts and the one who predicts the actions of another.

Although we can in principle apply this set of distinctions to any pair of agents, Hohfeld's presumption is that the agent observing and predicting the actions of the

other has an interest in what the other does, and ascribes interests to the other. In essence we are dealing with the ability of one agent to infer 1) that another agent has the ability or inability to change his situation in relevant ways and 2) that the other agent has an interest in changing or not changing it. Hohfeld's concepts of right, duty, privilege, and no-right relate to the interests someone has in changing the situation, presuming they behave normally, and the concepts of power, liability, disability, and immunity relate to the ability to change the situation.

The explanation again has to appeal to scripts (cf. section 4.5): an earlier example given of a "primitive plan" is buying a quart of milk. We realize that the execution of this script depends on someone else's willingness to execute "selling a quart of milk" part, but we still tend to believe that this will be achievable without problems, even for a known reference price. The most obvious reason that we are actively monitoring and predicting someone else's actions that change our situation is because we are executing a script that *involves that other person* as an actor.

Proposition 24. *Jural relationships* only exist between pairs of agents involved in the execution of the same *script*.

Example 11. This has to be made concrete with an example. The initiative to act has to come from both parties, and both parties recognize the same relevant participants of the situation. As an example we will use a sales transaction initiated by an irrevocable offer to sell by y . Person y offers for sale to x a painting for the price of \$500, to be accepted before next Tuesday.

Assume for the purpose of this example, prevailing law is that person y now has a duty to sell, under the stated conditions. Person x therefore has a right to buy under the stated conditions.

We can abstract away the amount (for the price of \$500), the object to be sold, and the timeframe since these do not add to the legal complexity of the case. A not time-limited offer to sell for an unspecified price would however be more difficult to handle. The SALE script involves two actors, the SELLER and the BUYER, who each have the ability to perform two constitutive actions, in no particular order:

$$\text{OFFERFORSALE} \sqsubseteq \exists \text{actor.SELLER} \sqcap \forall \text{actor.SELLER} \quad (4.4)$$

$$\text{SUPPLYSOLDITEM} \sqsubseteq \exists \text{actor.SELLER} \sqcap \forall \text{actor.SELLER} \quad (4.5)$$

$$\text{ACCEPTOFFER} \sqsubseteq \exists \text{actor.BUYER} \sqcap \forall \text{actor.BUYER} \quad (4.6)$$

$$\text{PAY} \sqsubseteq \exists \text{actor.BUYER} \sqcap \forall \text{actor.BUYER} \quad (4.7)$$

From the perspective of the buyer we are dealing with a buy action, and from the perspective of the seller with a sell action.

With this scenario in mind we run through Hohfeld's relations.

Definition 11. A *right* of x towards y wrt. z is equivalent to a *duty* of y towards x wrt. z .

The duty is obviously very similar to the obligation of deontic logic. Party SELLER sets the script into motion by an OFFERFORSALE which is recognized by BUYER, creating an institutional state, or position, OFFEREDFORSALE.

If BUYER proceeds with ACCEPTOFFER, which requires situation OFFEREDFORSALE, he creates for himself, besides the institutional state OFFERACCEPTED, an obligation to pay represented as in section 4.7.3, which is also SELLER's right to receive payment from BUYER. Moreover situation OFFEREDFORSALE and OFFERACCEPTED also creates SELLER's duty to supply the sold item, which is also BUYER's right to the same.

Informally we could also say that the acceptance creates SELLER's duty to "sell the painting", and it is not uncommon to hear it phrased in this way. This obligation is violated if SELLER does not sell the painting. Note however that this duty is also violated if BUYER does not complete his part of the transaction, which is giving the promised amount of money. BUYER may for instance find out that he has less money than he thought and try to back out of the sale. SELLER can however also refuse to take money, and BUYER can refuse to take the painting: the obligations do only apply to the part of the transaction which the party is able to perform.

This kind of statement therefore means in effect that x has a duty to y to perform z with the assistance of y only *if* and as long as y prefers z to happen. In practice this rarely leads to problems if we are dealing with the kind of obligations that only confer a *potestative right* (usually to go to civil court) on the other party. If a third party (e.g. a police officer) would have such a potestative right, the result would be bizarre.

If you have been offered the painting, and the offer is open to Tuesday, you can say now that you will not buy the painting and by saying so relinquish your right to buy it. In general this is however not the case. This *power* cannot be presumed.

Definition 12. A *privilege* of x towards y wrt. z implies *no right* of y towards x wrt. z .

Privilege suggests an exceptional situation. Privilege suggests permission since its jural opposite is the duty. If, for instance, the BUYER in the script is a minor, he has *no* duty to complete the sales transaction. As already noted in section 4.6, the explicit permission functions as an exception to a more general obligation.

Hohfeld considered the no-right as opposite of the right. I, and others before me, see no benefit in introducing a category for the mere purpose of completing the square. A no-right is simply the denial, the negation, of a right. Person y does not have a right towards x with respect to z , *either* because of the presence of a privilege to that effect or simply because of the absence of a duty to that effect. The SELLER for instance has no right to sell the painting before his offer for sale is accepted by anyone.

Readers might object that the sale is a defined legal act, and that it also takes place if BUYER or SELLER fails to complete his part of the bargain. In other words, as soon as BUYER has accepted the sale takes place. There is no way to get out of the sale. Running away with the painting does not change the fact that the sale did take place: it simply becomes disappropriation of the painting.

The rules regulating sales transactions are in actual law *not* normative, and the example is not at all a plausible representation.

The second square deals with this distinction. It is analogous to the first one, but now we distinguish an institutional LAW:SALE that took place:

OFFEREDFORSALE \sqcap OFFERACCEPTED \sqsubseteq \exists constitutes.LAW:SALE

Definition 13. A *power* of x towards y wrt. z is equivalent to a *liability* of y towards x wrt. z .

In short, a power is one's ability to alter legal relations. BUYER has the power to enter into a LAW:SALE with SELLER of the painting when he has offered it for sale. Thus, BUYER has the power to bind SELLER, *and* himself, to the legal consequences of the LAW:SALE. SELLER, thus, has a liability, which is correlative to power, in that he is liable to having his legal relations altered.

Definition 14. An *immunity* of x towards y wrt. z implies a *disability* of y towards x wrt. z .

If x has an immunity against y with respect to the sale of the painting, it means that y has no power to change x 's legal position with respect to the sale of the painting. Contrary to the permission the immunity is not generally understood as coming in a weak – implicit – and strong – explicit – variety. Immunity is always stated explicitly. The legislator may have stated that minors are immune to sales transactions. If BUYER is a minor, then accepting the offer does not change the legal position to LAW:SALE. A disability to bring about a legal effect may be because the counterparty is immune to it, but more commonly because some other necessary ingredient is missing.

While liability and right are simply other perspectives on power and duty, privilege and immunity are simple *one of* the possible reasons for no right or no ability.

Power and immunity are central to regulating the relations between state and citizen. For instance, if the constitution states that the state has no power to place me under a duty to ask prior permission for expressing my opinions in writing, then I have an immunity and the state a disability. In common parlance this is also a right. So *right* in common parlance may translate, depending on context, to a:

1. simple permission;
2. immunity;
3. potestative right, or;
4. right correlative to a duty.

The scenario discussed in this section, and its relation with scripts, will be continued in section 6.7.1.

4.8.3. Everyone and Delegation

So far we dealt with cases between a clear party x and y . As a general rule, the parties will have a right to go to a civil court in the illegal situations and the court will search

a remedy or reparation to be made by the party responsible for the violation to the party wronged by the violation.

Most norms do not clearly identify both sides of a transaction, however. When x runs a red light while having a duty not to do so, for instance, x wrongs everyone else present in that traffic situation. Other traffic participants have an abstract right towards x with respect to stopping for the red light, but will generally only be in the position to go to court to find a remedy or reparation if there is a concrete wrong. The right correlative to duty does not imply any potestative right. Still x wronged “everyone” in the abstract, and therefore the state exercises the reaction to the violation of a duty on behalf of the community as a whole. This “delegated” power is usually exercised in criminal law. Person x will be fined by the state if caught running a red light.

Hohfeld’s analysis has often been criticized for not dealing with these problems in identifying the involved parties, and sometimes reference is made to “universally quantified” others (cf. [156]). A categorical distinction between natural persons, private and public legal personalities, the abstract state in general, and “everyone” or the community seems a more promising approach. The problem is one of identifying the various implicit delegations of the right to search remedy or reparation from (for instance deceased) natural persons and from “everyone” to the state and its constituent public legal personalities. Since modern democracies are based on a conceptual model of *delegation* of power by ‘everyone’ to parliament, we can safely assume that government has the power to delegate these rights to its own parts.

Proposition 25. There are *collective entities* that can play the role of agent in the delegation of powers, but have no other ability besides that. Entities of this type for instance occur in constitutional law.

Delegation of powers is quite central to how administrative law conceives of the legal position of the state and its parts in many legal systems, and therefore also to the legislative process itself. It is considered a central feature of the conceptual organization of legislation in [51], and is discussed at some points in chapter 5, specifically section 5.2.5.

4.9. Conclusions

Law consists of institutions, and these institutions can be conceptualized as systems with a well-defined interface with an environment. The structures of the institution are defined by the institutional facts that make up the institution, and its mechanisms of change are the constitutive rules that specify what *constitutes*, or counts as, an institutional fact. Conversely, the institutional fact has a constitution base, which consists of the application of a constitutive rule to the *constituting facts*, which are *brute* facts, yielding an institutional fact.

Brute facts are pre-existing and external to the institutional reality constituted by the rules. They form the relevant environment. The institutional facts of one institution can be the brute facts of another one.

This notion of the law as a set of socially recognized systems with an interface opens up the possibility of maintaining a one-to-one correspondence between the knowledge components of chapter 3 and the institution. This has as a consequence that defeasibility can only occur on the interface. As explained in previous chapters, this is justifiable as a form of ontological stratification.

Let C_b be a concept from brute reality, and C_i a concept from institutional reality. *Institutional rules* are terminological axioms, of the form $C_i \sqsubseteq C_b$ or $C_i \equiv C_b$, about the institution and its structure.

Constitutive rules take the form of an indicator or a requirement, and sometimes both forms exist for the same facts. An indicator is a rule of the following form:

$$\begin{aligned} &(\text{DEFAULT } r \text{ (known.}C_b) \text{ (free.}\exists\text{law:constitutes.}C_i) \\ &\quad \text{(assume.}\exists\text{law:constitutes.}C_i)) \end{aligned}$$

A requirement is a rule of the following form:

$$\exists\text{law:constitutes.}C_i \sqsubseteq C_b$$

The combination of both results in a peculiarly asymmetric kind of equivalence statement.

In addition I have used the integrity constraint to check the integrity of arguments for a certain proposition. The constraint can be used to model the burden of proof of the user in a specific KBS application in a simple way, and has the following form:

$$(\text{CONSTRAINT } r \text{ (known.}\exists\text{law:constitutes.}C_i))$$

Terminological axioms ($C_b \sqsubseteq C_i$), constraints, or default rules about brute reality do not in principle belong to the legal institution, and do not occur in the source of law: the representation of a rule in the source of law confirms institutional status for at least one concept in the rule.

The identification of institutions with ontological strata of course has consequences for the identification of institutions. Since I use my own criterium for telling apart different institutions (apparent defeasibility is an argument for splitting an institution into multiple institutions), we cannot maintain other such criteria (for instance that each legislator manages its own institutional reality).

Normative rules are also constitutive rules that derive the institutional fact either the $\text{disallows}(n, i)$ or $\text{allows}(n, i)$ from constituting facts about i .

The institutional interpretation is a coherent description of the meaning of law that can be used to model knowledge about the sources of law. Note however that institutional and constitutive rules *are not* types of legal rule: the relation is slightly more complicated, as chapter 5 will show.

The institutional interpretation however tells us little about the functions of law for its users. It applies to chess games as well as to law. The purpose of chess games is winning, but the purpose of law is to formalize a normative order. Law makes social interaction predictable by giving people reasons to do certain things and to refrain from doing certain things.

To explain these functions, we have to appeal to planning and plan recognition. These areas do not however form part of the field of law, and neither do the sources of law formulate a particularly coherent ontology of these, in essence, mental and therefore not directly observable activities.

In some cases such an explanation is straightforward. The analysis of normative rules in terms of normative positions and obligation, i.e. deontic logic, is such a straightforward abstract theory of behaviour, based on the expectation that people generally avoid the circumstances in which they are liable to be punished.

To explain the normalizing effect of other rules one must ascribe intentions and preferences to agents: People sometimes intentionally try to bring about or avoid certain legal facts. Since cognitive resources for planning are limited, and people share a lot of their knowledge, intentions are both predictable and can be recognized on the basis of observing behaviour.

The legislator sometimes directly appeals to the idea that behaviour, and the changes brought about by behaviour, are the result of a plan. Some rules for instance only have an institutional effect if intended. Normative rules can address not only the changes brought about by somebody's behaviour, but also the intention behind the behaviour. The subject of a normative rule is therefore naturally the action, which links observable behaviour to a plan. The legislator may also occasionally use concepts in rules – like the right correlative to a duty – that make presumptions about the preferences and expectations of a counterparty (that is not himself *addressed* by the rule but identified by his social or legal agent role) in a transaction.

In this chapter we introduced some minimal vocabulary for action as plan execution, for the changes brought about by plans, for agent roles, and for the relation between situation and action. Of particular importance is the notion of execution: the expression $\exists \text{executes}.P$ means that 1) something is an action, and 2) that the action is an attempt at performing P , which is a task – i.e. a plan being executed. Although tasks and actions are different things, the description of both typically uses the same vocabulary.

Since the subject of normative rules is an action, an analysis in terms of normative positions and obligation based on possible worlds must evidently be based on *alternative actions*: the true object of comparison must then obviously be alternative plans that could be, or could have been, executed.

While this chapter discussed the phenomenon of law in separation from the formal sources of law, chapter 5 will focus on the relation between the sources of law, as the physical result of formal constitutive acts on a legal institution, and specific identifiable *legal rules*, about which we express knowledge in the form of the knowledge representation rules of chapter 3. The function of law, formalizing normative order, will return in chapter 6.

This book is based on the proposition that the law and normative order can be completely separated. Chapter 6 will however discuss some instances where this becomes problematic because the user of legal rules must ascribe an intended normative order to the legislator to apply the rules. It is however equally problematic to want to completely include the normative order in a knowledge representation, and certainly in a knowledge representation of sources of law.