TITLE

Giovanni Sartor

Abstract.

Contents

CHAPTER 1

Defeasible Reasoning as Argumentation

This chapter provides an analysis of defeasible legal reasoning as argumentation. It first provides a general account of the idea of defeasibility and introduces the idea of nonmonotonic reasoning. It then focuses on defeasible argumentation, considering how defeasible arguments can be constructed and how they can be defeated by rebutting and undercutting counterarguments. The dialectical interactions of defeasible arguments are further explored by focusing on reinstatement and reasoning about priorities. The idea of legal systems as the basis for argumentation frameworks is then investigated.

Keywords: defeasible reasoning, conclusive reasoning, legal reasoning, legal problem-solving

1.1. The Idea of Defeasibility

In a very broad sense, the idea of defeasibility may be applied to any process that responds to its normal inputs with certain outcomes (the default results), but which delivers different outcomes when such inputs are augmented with further, exceptional or abnormal elements.

The computer scientist and theorist of complexity John Holland argues that complex systems—such as a cell, an animal, or an ecosystem—can be characterised "in terms of a set of signal-processing rules called classifier rules" (Holland 2012,28). Each such rule represents a mechanism which "accepts certain signals as inputs (specified by the condition part of the rule) and then processes the signals to produce outgoing signals (the action part of the rule)." He observes that complex systems need to address different situations, requiring different responses, which are triggered by rules having different levels of generality. In many cases the most efficient way to cover multiple different contingencies consists in constructing "a hierarchy of rules, called a default hierarchy, in which general rules cover the most common situations and more specific rules cover exceptions" Holland et al. 1989). Thus, a general rule would provide the normal response to a certain input, but more specific rules would override the general rule in exceptional situations, in which a different response is needed. The emergence of default hierarchies may be favoured by evolution, since such hierarchies may contribute to the fitness of the systems using them.

> Default hierarchies offer several advantages to systems that learn or adapt:
>
> - A default hierarchy has many fewer rules than a set of rules in which each rule is designed to respond to a fully specified situation.

- A higher-level rule [...] is easier to discover (because there are fewer alternatives) and it is typically tested more often (because the rule's condition is more frequently satisfied.
- The hierarchy can be developed level by level as experience accumulates (Holland 2012,122).

This perspective can be applied to different domains at different levels of abstraction. For instance, at the cellular level rule mechanisms specify the catalytic and anti-catalytic processes that induce and inhibit chemical reactions. At the DNA level, rule mechanisms are provided by genes (and parts of them). Each gene delivers the protein matching the sequence of the gene's bases and it may be regulated by other genes that send signals that under particular conditions repress (turn off) or induce (turn on) the functioning of the gene rule at issue. Animal behaviour is also largely governed by systems of reflex rules defining reactions: to heat or cold; or to the sight, smell, or taste of food; or to the perception of incoming dangers; and so on. Such reflexes can be innate or learned by experience, i.e., by conditioning and reinforcement. They may interact in complex patterns: some reflexes are stronger than others, so that they determine the response in cases of conflict, and some reflexes may have an inhibitory impact on other reflexes, blocking them under particular situations. In humans, reflexes are integrated with deliberative processes and means-end reasoning, but still they govern a large part of human behaviour.

As Holland et al. (1989,38) argue, not only instinctive reflexes but also mental models can be based on sets of prioritised default rules:

> The rules that constitute a category do not provide a definition of the category. Instead they provide a set of expectations that are taken to be true only so long as they are not contradicted by more specific information. In the absence of additional information these "default" expectations provide the best available sketch of the current situation. Rules and rule clusters can be organized into default hierarchies, that is, hierarchies ordered by default expectations based on subordinate/superordinate relations among concepts. For example, knowing that something is an animal produces certain default expectations about it, but these can be overridden by more specific expectations produced by evidence that the animal is a bird.

In conclusion, a defeasible process can be characterized a mechanism which responds to its normal inputs with certain default outcomes, but that may fails to respond in this way when the input is accompanied by certain additional exceptional elements.

1.2. Defeasibility in Reasoning and Nonmonotonic Inference

Though defeasibility also applies to reactive agents, it acquires its fullest meaning in cognitive agents: defeasible cognition consists in achieving certain cognitive states (beliefs, intentions, etc.) when provided with certain normal cognitive inputs (perceptions, beliefs, intentions), but refraining from adopting these states, or abandoning them, when the normal inputs are accompanied by further elements. More specifically, the idea takes on a more precise content when referred to reasoning, i.e., to inference or argumentation. A defeasible reasoning process (an inference

or argument pattern) responds to typical input premises with certain default conclusions, but fails to deliver those conclusions when the typical input premises are accompanied by further premises, indicating exceptional circumstances.

The most cited example of a default inference concerns Tweety the penguin. Let us assume that we are told that Tweety is a bird. Given this information and knowing that birds usually fly, we would normally conclude that Tweety flies. Assume, however, that we are later told that Tweety is a penguin. Given this additional information and knowing that penguins are birds which do not fly, we should refrain from endorsing the conclusion that Tweety flies. In fact, we now know that he is a special kind of bird, namely, a penguin, to which the default rule does not apply.

As this example shows, the addition of premises in a defeasible reasoning may lead to the withdrawal of a conclusion. This aspect of defeasible reasoning is conceptualised through the distinction between monotonic and nonmonotonic reasoning. In general, we say that an inference method is *monotonic* when it behaves as follows: any conclusion that can be obtained from an initial set of premises can still be obtained whenever the original set is expanded with additional premises. More precisely, all conclusions that are derived through monotonic inferences from a premise set S_1 can also be derived from any larger (more inclusive) premises set S_2 ($S_1 \subseteq S_2$).

Correspondingly, an inference method is *nonmonotonic* when it behaves as follows: a conclusion that can be obtained from an initial set of premises may no longer be obtainable when the original set is expanded with additional premises. More precisely, conclusions that are derived through nonmonotonic inferences from a premise set S_1 may no longer be derivable from a larger (more inclusive) set of premises S_2.

Deduction is monotonic: as long as we accept all premises of a deductive inference, we must continue to accept its conclusion. Therefore, we also say that deductive inference is *conclusive* : as long as we maintain the premises, any additional information will not affect the conclusion.

By contrast, defeasible inferences are nonmonotonic: we may reject the conclusion of a defeasible inference while maintaining all of its premises. This may indeed happen when further premises are provided that substantiate exceptions to the defeasible inference. In defeasible reasoning " if the premises hold, the conclusion also holds tentatively, in the absence of information to the contrary" (Walton 2008 160). Thus, defeasible inference relies on absence of information as well as its presence, often mediated by rules of the general form: given P , conclude Q unless there is information to the contrary. (Horty 2001,337).

Defeasible reasoning is not only a matter of practice but also one of rational justification, as stated in the following definition:

> Reasoning is *defeasible* when the corresponding argument is rationally compelling but not deductively valid. The truth of the premises of a good defeasible argument provide support for the conclusion, even though it is possible for the premises to be true and the conclusion false. In other words, the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information. (Koons 2009).

As we shall see in what follows, in many situations we are entitled or justified to derive default conclusions and to maintain those conclusions until we come to appreciate that circumstances obtain under which such conclusions should not be retained

1.3. Conclusive and defeasible arguments

Different approaches to defeasible (nonmonotonic) reasoning and its formalisation have been developed (see Ginzberg 1987,Horty 2001,Prakken and Vreeswijk 2001). Here I shall approach defeasible reasoning as argumentation, namely, as the derivation of a provisionally justified conclusions through the dialectical opposition of competing arguments (on argumentation, see Walton 2013). This is indeed the perspective that better fits the argumentative and dialectical nature of legal reasoning, as it emerges in analysis, advocacy, and decision-making.

A *valid argument* can be said to consist of three elements: a set of premises, a conclusion, and a support relation between premises and conclusion. In a *deductively valid argument,* the premises provide *conclusive* support for the conclusion: if we accept the premises we must necessarily accept the conclusion. In a *defeasibly valid argument* , the premises only provide *presumptive* support for the conclusion: if we accept the premises we should also accept the conclusion, but only so long as we do not have prevailing arguments to the contrary. We can extend the notion of an argument to unsupported claims: such a claim can be viewed as argument only consisting in the assertion of a conclusion. The unsupported claim of a proposition will be sufficient to substantiate it, when the truth of the proposition is evident or is anyway agreed upon.

Defaults usually have a general form and consequently have to be mapped onto or instantiated to the specific propositions to which they are applied. For instance, to apply the general default "pet dogs are presumably unaggressive," i.e., in a conditional form, "if something is a pet dot, then presumably it is nonaggressive," to Fido, we must specify or "instantiate" the default to the case of Fido, i.e., generate the following specification: "if Fido is a pet dog, then presumably Fido is not aggressive." This specification, in combination with the premise that Fido is a pet dog, leads us to the presumable conclusion that Fido is not aggressive, through defeasible *modus ponens.* In the examples that follow, I will omit the specification step, presenting the conclusion as directly resulting from the general default and the specific conditions matching its antecedent. In fact, a general default can be seen as the set of all of its specific instances, which include the one applied to the case at hand.

I shall use a diagrammatic representation for arguments, as exemplified below, where the boxes include premises or conclusions, and combinations of premises are linked to the conclusion they conjunctively support. In the diagram of Figure 1, we can see a deductive argument (A) supporting the conclusion that Fido, being a dog, is a mammal and a defeasible argument (B) supporting the conclusion that Fido, being a *pet* dog, is presumably not aggressive. I have represented the premises both in natural language and in the usual formalism of predicate logic, and have labelled the connection between premises and conclusion by the letters C and D to distinguish conclusive from defeasible arguments.

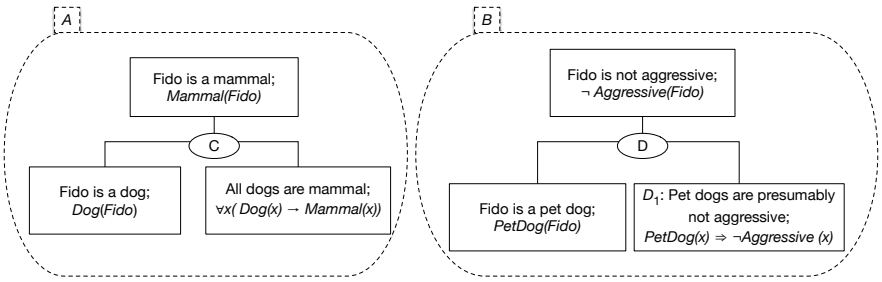

FIGURE 1. Conclusive and defeasible arguments

Arguments in natural language usually have an *enthymematic* form, meaning that they may omit some of the premises that are needed to support their conclusions. Here I shall present all arguments in their complete form, that is, as including all premises that are sufficient to conclusively or defeasibly establish their conclusion.

In particular, I assume that each defeasible argument includes (a) a set of antecedent conditions, and (b) a defeasible conditional, called a *default* , according to which the (conjunction of the) conditions presumably determines the argument's conclusion. I represent defaults in the form "if P_1 and ... and $P_n$then presumably Q", in formula $P_1 \wedge \cdots \wedge P_n \Rightarrow Q$, where the arrow $\Rightarrow$ denotes defeasible conditionality (I will use the arrows $\Rightarrow$, , and $\twoheadrightarrow$ to denot defeasible, material and strict conditional respectively, see Section). Thus a single-step defeasible argument has the following form:

1 $P_1, \cdots, P_n$ (the antecedent conditions), and
2 if P_1 and ... and P_n then presumably Q (the default, in formula: $P_1 \wedge \cdots \wedge P_n \Rightarrow Q$).

therefore

3 Q.

This inference is called *defeasible modus ponens* to distinguish it from the conclusive modus ponens inference of deductive logic. We can represent a defeasible argument by providing the set of its premises (conditions and default): $\{P_1, \cdots, P_n, P_1 \wedge \cdots \wedge P_n \Rightarrow Q\}$, the conclusion of the argument being conclusion of the default. Given a defeasible modus ponens inference (argument) $\mathcal{A} = \{P_1, \cdots, P_n, P_1 \wedge \cdots \wedge P_n \Rightarrow Q\}$, I will say that the conjunction of the $P_1, \cdots, P_n$ conditions is the *reason* for (concluding that) Q and that the default $P_1 \wedge \cdots \wedge P_n \Rightarrow Q$ is the *warrant* for Q. I will also say that Q is *warranted* by that default.

For instance, given argument B in Figure 1, we can say that the fact that Fido is a pet dog is the reason for concluding that he is not aggressive and that that this conclusion is warranted by the default that pet dogs are not aggressive. As example of conjunctive reason, consider the argument in Figure 2, where the conjunction of the two premises P_1 and P_2 provides the reason for the conclusion warranted by the default D. Note that, I freely use symbols $P_1, \cdots, P_n$ as names for propositions and $D_1, \cdots, D_n$ as names for defaults, whenever needed.

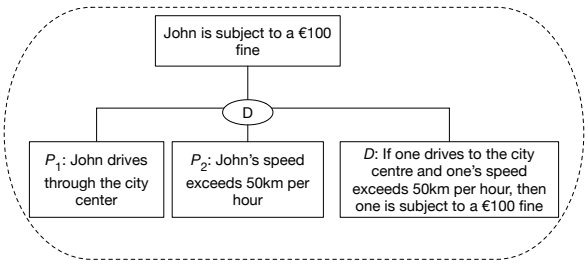

FIGURE 2. Linked argument

The notion of a defeasible argument can be generalised to cover multi-step defeasible arguments, which consist of set of the arguments providing the conditions of a top default, plus that default. For instance, if $\{P,\ P \Rightarrow Q\}$ is a defeasible argument, so is also $\{\{P,\ P \Rightarrow Q\},\ Q \Rightarrow R\}$: (for an example of multistep defeasible argument, see Figure 16, for a formal definition of the general notion of an argument, possibly including both defeasible and deductive steps, see (Prakken 2010,Section 3.2)).

1.4. Linked arguments and convergent argument structures

Besides the distinction between defeasible and conclusive arguments, a second categorisation of arguments is relevant to our purposes, namely, the distinction between linked arguments and convergent argument structures (see Walton 2006,139 ff., Hitchcock 2017,Ch. 2).

A linked argument is an argument that includes, beside a conditional warrant, more than one premises. None of these premises is sufficient to trigger on its own the conjunctive antecedent of the conditional warrant. Therefore, in isolation, each of them fails to provide any (presumptive or conclusive) support to the conclusion of that warrant. For instance, assume the following premises (P_1) John drives through the city centre, (P_2) his speed exceeds 50km per hour, and (D) if one drives through the city centre, and his or her speed exceeds 50km per hour, then one is subject to a 100 euros fine. Only the joint combination of premises P_1 and P_2 triggers (presumably) the conclusion that John is subject to a 100 Euros fine (Q). The resulting argument is depicted in Figure 2.

A convergent argument structure is a combination of multiple arguments, each leading to the same conclusion. Often, but not always a convergent argument structure provides a stronger support to the common conclusion of its component arguments than each of these arguments would do in isolation (see Prakken 2005, Bench-Capon and Prakken 2006). In Figure 3 you can see how two witness testimonies originate separate arguments A and B which merge into the convergent argument C , which provides a stronger support to the common conclusion of A and B .

Figure 4 shows a combination of independent arguments pointing to the same practical conclusion (a conclusion concerning what should be done): being asked the way by driver John, I should not direct him in a wrong direction (tell him that he will get to destination by going to the left), given that my false statement would both be a lie and harm John, impeding him from reaching its destination.

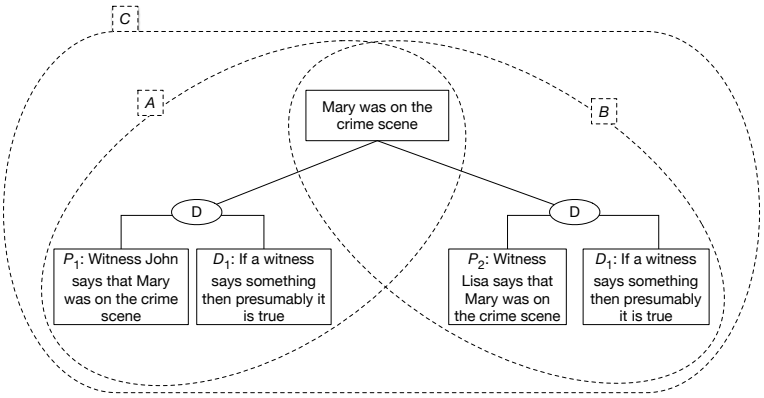

FIGURE 3. Convergent factual argument

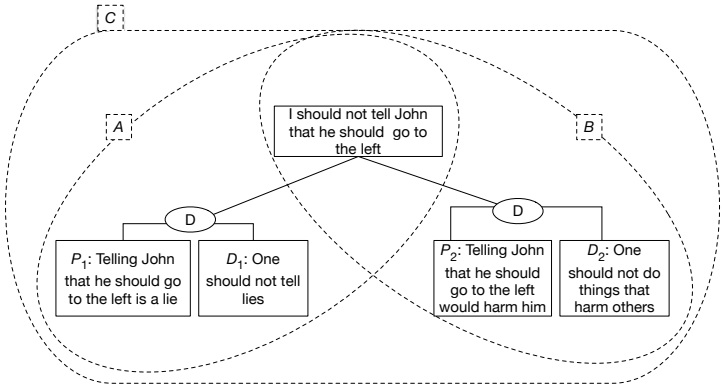

FIGURE 4. Convergent practical argument

Arguments A and B refer to two parallel principles: the duty to be truthful and the duty not to harm others (a foremost requirement of the law according to Justinian's Digest, D 1.1.1).

The distinction between linked arguments and convergent argument structures enables us to distinguish two important concepts: the concept of a contributory condition and the concept of a contributory reason.

A *contributory condition* for a conclusion is a necessary element of a (presumably or conclusively) sufficient condition for that conclusion. This concept applies

to each element of to each element of a conjunctive warrant supporting that conclusion. Given a warrant "if P_1 and ... and P_n then (presumably) Q", each $P_1, \ldots P_n$ is a contributory condition for that warrant.

A *contributory reason* for a conclusion, is a presumably sufficient condition for that conclusion. This concept only applies to the whole antecedent of a warrant supporting that conclusion. If P is a contributory condition for Q, then there must exist a warrant "if P then (presumably) Q, where P may be a single proposition or a conjunction of propositions.

Thus premises P_1 and P_2 in Figure 2 are contributory conditions, but fail to qualify as contributory reasons: neither of them neither of them can separately trigger the conclusion of the argument: both are needed to satisfy the conjunctive antecedent of the argument's warrant. Therefore, neither of them can be properly characterised as a reason for that conclusion. On the other hand, each of premises P_1 and P_2 in Figure 3 and Figure 4 does qualify as a (contributory) reason for their common conclusion, since each of them (together with the corresponding default warrant) fully supports that conclusions, besides contributing to provide a stronger joint support to that conclusion.

In the legal domain, the idea of a contributory reasons applies to the domain of principles, understood as optimisation requirements (Alexy 2002, Ch. 4) or value-norms (Sartor 2013,Section D). The fact that a choice advances a principle (a legal value) is a contributory reason for adopting that choice (of for its constitutional legitimacy). When the same choice advances multiple principles, this originates multiple convergent arguments —the advancement of each principle being a contributory reason— that join to provide a stronger support to that choice. Similarly, the fact that a choice negatively affects the realisation of a principle is a contributory reason for not adopting the choice or against its legitimacy. When multiple principles are negatively affected this originates multiple convergent arguments against that choice.

The idea of a contributory reason also applies to the antecedent of legal rules. As I shall argue in the following, the antecedent of a legal rules usually only provides a presumably sufficient condition for the conclusion of that rule. For instance, a driver exceeding a speed limit may not be subject to sanction in case his behaviour is justified by self-defence (he was trying to escape from a killer) or state of necessity (he was transporting a person to the hospital for an emergency). Rule-warranted arguments and principle-warranted arguments, while sharing the same basic logical structure, present some relevant differences. Firstly, rule-warranted arguments may "exclude" (undercut, in our terminology, see Section 5), rather than oppose (rebut), certain contrary arguments warranted by principles (if we follow the idea of Raz 1985, also adopted by Hage 1997). Secondly, convergent rule-based argument structures usually do not provide a stronger support to their conclusion than the constituting arguments do. For instance, assume that a person has committed a violation that triggers his or her liability both in contract and in torts. This provides for a converging argument structure for the liability of this person. However, this convergent argument structure arguably does not provide a stronger support to the liability conclusion than the strongest of the two separate arguments for that conclusion. I have preferred to speak of a convergent argument structure, rather than or a convergent argument, to denote the combination of arguments leading to

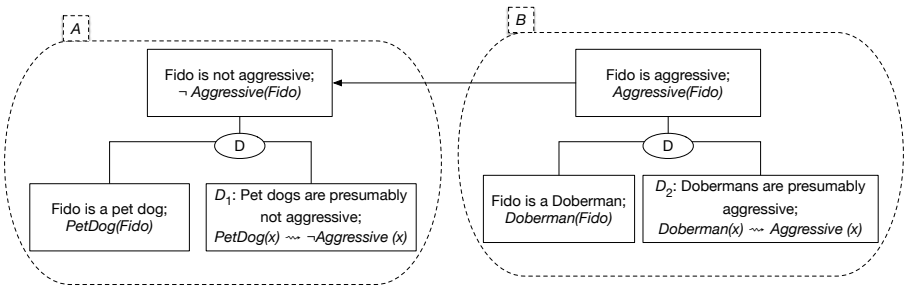

FIGURE 5. Rebutting attack

the same conclusions, to maintain the concept of an argument I introduced above, that requires a single warrant linking premises and conclusion.

1.5. Attacks against Arguments: Rebutting and Undercutting

An argument can be attacked in any of three ways: by attacking its premises, by attacking its conclusions, or by attacking the support relation between premises and conclusions. Conclusive arguments can only be attacked by challenging their premises, since, if the premises are accepted, then the conclusion must also be accepted. So, for instance, if we accept that Fido is a dog and that all dogs are mammals, we must also accept that Fido is a mammal (as soon as we are aware of the logical connection between premises and conclusion). In fact, it may also be possible to attack the conclusion of the argument—i.e., to deny that Fido, who is a robot in the likeness of a dog, is a mammal—but then we must also reject the premise that Fido is a dog (we exclude that dog-like robots count as dogs), or alternatively, we can deny that all dogs are mammals (we also include dog-like robots among "dogs").

By contrast, a defeasible argument can also be attacked by denying its conclusion, even if its premises are not questioned. For instance, let us assume that Fido is not only a pet dog but also a Doberman, and that Dobermans are presumably (normally) aggressive. Then, as shown in Figure 5 we can build an argument that attacks the previous argument by contradicting its conclusions (attack is expressed by the jagged arrow).

Clearly, we cannot endorse both arguments A and B at the same time (their conclusions are contradictory), and so we must either choose between them or remain uncertain as to which one we should choose. When two arguments conflict in such a way that the (final or intermediate) conclusions of one of them contradicts a (final or intermediate) conclusion of the other, we have a *rebutting conflict* between two arguments. To determine the outcome of a rebutting conflict we must consider the comparative strength of the two arguments. If one argument is stronger than the other (at the juncture at which the conflict takes place), then it prevails, i.e., it defeats its opponent without being defeated by it. In this case, the prevailing argument is said to *strictly defeat* its opponent. If neither of the conflicting arguments is stronger than the other, they each *weakly defeat* the other, i.e., their conflict remains undecided (for a logical analysis of these notions, see Prakken and Sartor 1997; Prakken 2010). To compare arguments, we adopt the so-called "last-link"

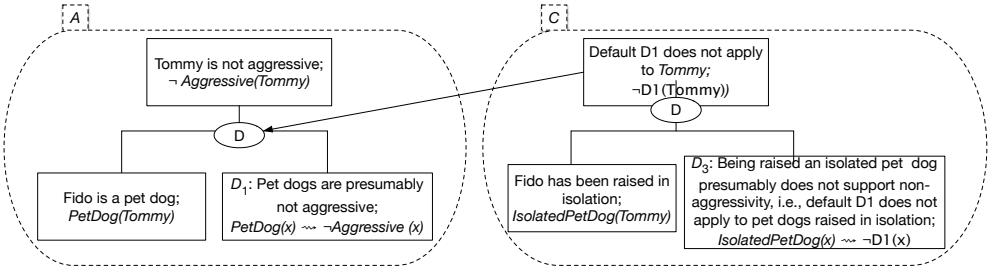

FIGURE 6. Undercutting attack

principle, which affirms that when two defeasible arguments contradict each other, to determine the comparative strength of the two argument, at the point of where they clash, we must compare only the defaults that directly deliver the conflicting conclusions (possibly with the help of deductive inferences). We do not consider the defaults eventually used, in multi-step arguments, to establish the preconditions of the directly conflicting defaults (for a discussion of the last-link principle and a formal definition, see Prakken 2010,Section 6).

In our example, let us assume that we consider that the argument on the left in Figure 5 (Fido presumably is aggressive, given than it is a Doberman) is stronger that the argument on the right (Fido presumably is not aggressive, being a pet dog). According to this priority relation between the two arguments, the first can be said to strictly defeat the second: we should accept the conclusion that Fido is indeed aggressive (and be careful in approaching him).

A second kind of attack against defeasible arguments consists in contesting the support link between the premises and the conclusion of the argument, namely, in denying that in the case at hand, these premises can provide sufficient support for the conclusion (on undercutting, see Pollock 2008). Let us assume that we are dealing with another dog—let us call him Tommy—and let us assume that we know Tommy to be a pet dog, but we also that he has been reared in an isolated mountain hut, having had contact only with his owner, and that we believe that the nonaggressiveness of pet dog toward strangers is mainly due to their experience in previous interactions with a large enough set of humans. We can then reasonably claim that, under these particular circumstances, the fact that Tommy is a pet dog does not adequately support the conclusion that he is friendly toward strangers. This kind of conflict is called *undercutting* (see Figure 6).

An undercutting argument always strictly defeats the argument it attacks, since (contrary to what happens in rebutting) it is not counterattacked by the latter argument. In fact, the undercutter says that the undercut argument does not work in the case at hand, while the undercut argument does not say anything about its undercutter. Note that the undercutter could also be viewed as attacking the particular instance of the default that is applied in the inference. For instance, it may be said that undercutter in Figure 6 denies that the conditional "if Fido is a pet dog, then presumably he is not aggressive" holds, i.e., it denies that the fact that Fido is a pet dog is a reason for him not to be aggressive (given the conditions in which Fido has been raised). However, I prefer to view the undercutter as an

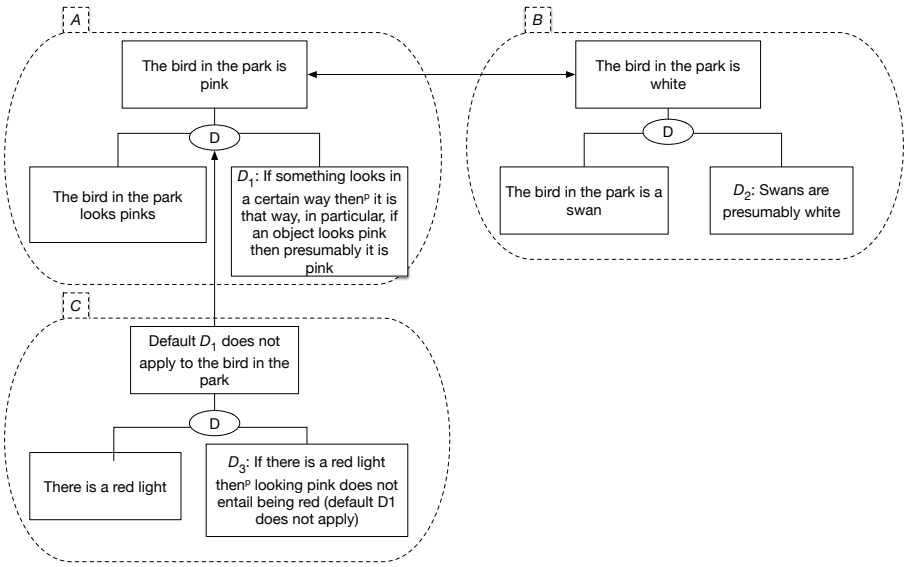

FIGURE 7. Undercutting attack: defeasible perception

attack against a particular inference applying the general default, to stress that the general conditional, stating a presumptive connection, is not affected by the attack.

Let us consider an example pertaining to the epistemology of perceptions (see Figure 7). Assume that in the park I see a bird that to me looks pink (I perceive it in this way), and therefore I conclude that the bird is pink. However, assume that I also see that the bird is a swan, which leads me to conclude that the bird is white, as swans normally are. However, since I know little about swans, I may remain in doubt about the colour of the bird: am I seeing a special swan (are there any pink swans around?) or is my perception of pink misleading me. Assume, however, that I notice that there is a red sunset. Then, as I know that even white things (not only pink ones) look pink under a red light, I will conclude that the fact that the bird looks pink under these conditions does not guarantee that it is indeed pink (it might as well be white): this undercuts the inference from *looking* pink to *being* so.

1.6. Rebutting and Undercutting in the Legal Domain

Let us now us take up rebutting and undercutting in the legal domain. Consider three norms dealing with civil liability (they are somewhat simplified versions of rules in the Italian Civil code, note that I assume that all such norms express the presumptive conditional connection "then presumably", which is abbreviated as then^p in the figures): the first rule (D_1) says that those who cause damage to another through their fault are presumably liable, the second (D_2) that persons lacking capacity are presumably not liable, and the third (D_3) that the incapacity exception presumably does not apply to those who find themselves in a state of incapacity through their own fault (see Figure 8). Assume that we know that John culpably caused damage to Tom (e.g., by deliberately smashing his car). On the basis of this information and of the first norm, we can conclude that John is liable

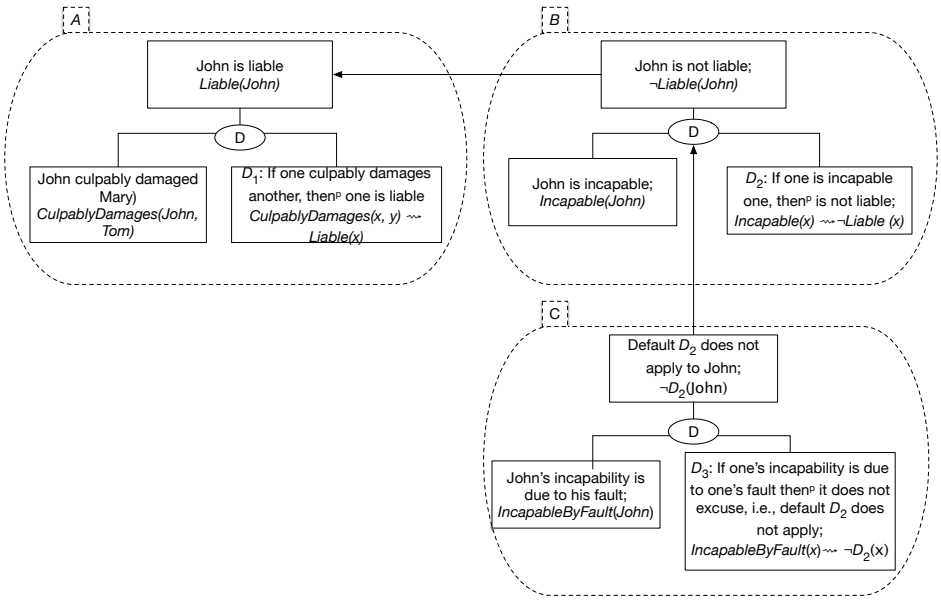

FIGURE 8. Undercutting attack: inapplicability rule

to pay damages (argument A). However, assume that it appears that John lacked capacity at the time of the incident. Then we can have an argument as to why John is not liable (argument B). Indeed, the incapacity exception takes priority over the general liability rule, such that argument B defeats argument A without being defeated by it. Assume, however, that John's incapacity was due to his fault, e.g., to his taking illegal drugs. This provides us with a third argument (C) that undercuts (makes irrelevant) the incapacity exception.

In legal contexts, a different way of undercutting can also be found. This involves those cases in which a legal norm explicitly includes among its preconditions the absence of an "impeditive fact," namely, a fact such that if were established, it would prevent the norm's conclusion being derived (on impeditive facts, see Sartor 1993). This is conveyed by stating that the norm's consequent follows from certain conditions, unless the impeditive fact holds, or by stating that it follows from such conditions if the impeditive fact is not established. The norm's consequent can be derived without needing to establish the absence of the impeditive fact, while establishing that fact would prevent that derivation.

Consider, for instance, the rule in Italian law under which a producer is liable when a product it manufactures harms a consumer, unless it is shown that the producer is not at fault (took all reasonable precautions). Here the impeditive fact is the absence of fault on the producer's side. Let us consider the issue of whether John may be liable as the producer of the motorbike which caused Tom's accident by failing to come to a stop before an obstacle (see Figure 9). It is not necessary to establish John's fault to determine his liability as a producer. In other words, John's liability can be presumed by applying this norm (this is denoted by the

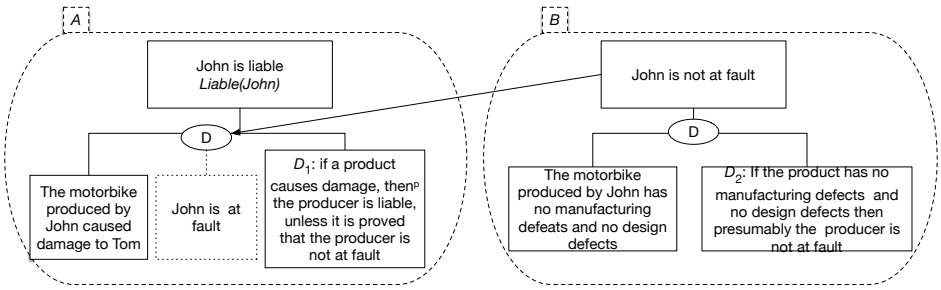

FIGURE 9. Undercutting attack: impeditive fact

dotted lines around this premise). However, if it is established that the motorbike was not defective, John may avoid liability.

1.7. Levels of Abstraction of Arguments

Defaults can have different levels of abstraction, some representing general patterns of inference or inference schemes (Walton et al. 2008), others representing more specific connections between preconditions and conclusions. Indeed, the same conclusion can often be argued by using either a general inference scheme or a more specific rule. Consider, for instance, the issue of the morality of lying, which was the object of a famous controversy between Emmanuel Kant and Benjamin Constant (see Kant 1949). John shows up at Mary's door at asks her whether Bob is at her place. Assume that Bob is in the house, that Mary is aware of this, and that Mary knows that John is armed and intends to kill Bob. The issue is whether Mary should lie, saying that Bob is away so as to save his life.

Let us first consider the argument according to which Mary should not lie. One way to frame this argument is as an argument pertaining to the implementation of moral rules in general. In that case, the premises of the argument could presented as follows:

(1) If rule "if P then Q" is a moral principle, and P is the case, then presumably Q .
(2) The rule "if a statement is a lie, then one should not make the statement" is a moral principle.
(3) The statement that Bob is away is a lie.

By defeasible modus ponens, these premises lead to the conclusion that

(1) Mary should not make the statement that Bob is away.

However, the argument can also be framed in a more specific way, taking the prohibition on lying for granted and using it directly as a premise:

(1) The statement that Bob is away is a lie.
(2) If a statement is a lie, then presumably one should not make the statement.

It seems to me that this second approach fits better our commonsense reasoning, in which we directly use the warrants we endorse, to derive specific conclusions. Considerations pertaining to the foundation or the nature of such warrants are brought in through further arguments. For instance, the adoption of a warrant may be supported by arguments pointing to the consequences of its practice (e.g, in a

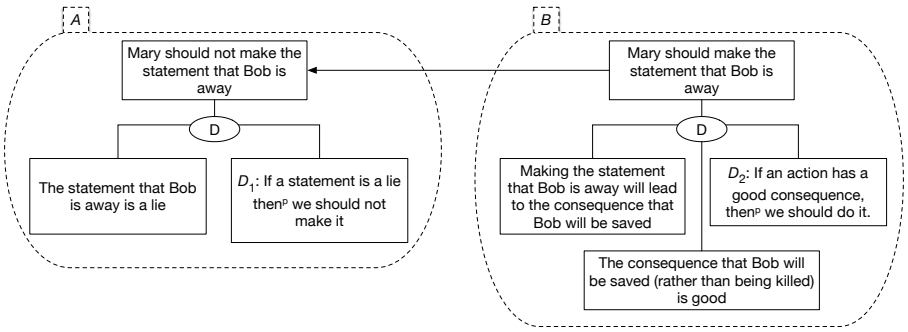

FIGURE 10. Conflicting arguments: strict defeat

rule-utilitarian perspective, the prohibition to lie may be supported by considering the benefit deriving from his generalised practice). Similarly, the strength and function of a warrant can be supported by arguments pointing to its nature (e.g. the fact that a principle pertains to morality may support its superiority over self-interested reasons, or or the fact that it belongs to the law may support its coercive enforceability or its exclusionary nature).

Again, by defeasible modus ponens, these premises lead to the presumable conclusion that

(1) Mary should not make the statement that Bob is away.

Let us now consider an argument why Mary should, on the contrary, lie. To build this argument we can appeal to a different pattern of defeasible inference: call it "inference from good consequences" or teleological argument. According to this pattern, the premises

(1) Making the statement that Bob is away will lead to the consequence that Bob will be saved (rather than being killed by John).
(2) This consequence is good.
(3) If an action has a good consequence then presumably we should do it.

lead to the conclusion that

(1) Mary should make the statement that Bob is away.

The two arguments and their conflict are represented in Figure 10.

If we agree that argument B is stronger than argument A, we should maintain that it strictly defeats argument A, and consequently we should endorse the consequence of B , i.e., that Mary should tell John that Bob is away (even if it is a lie).

1.8. Reinstatement

So far, we have only considered relations between *pairs* of arguments. However, this is insufficient to determine the status of an argument, namely, whether we should accept it or not. More precisely, this is insufficient to establish whether an argument is justified, such that we should accept its conclusion; overruled, such that we should not pay attention to it; or merely defensible, such that we should remain uncertain as to whether to accept it or not (on justified, overruled, and defensible arguments, see Prakken and Sartor 1997). This is because an argument

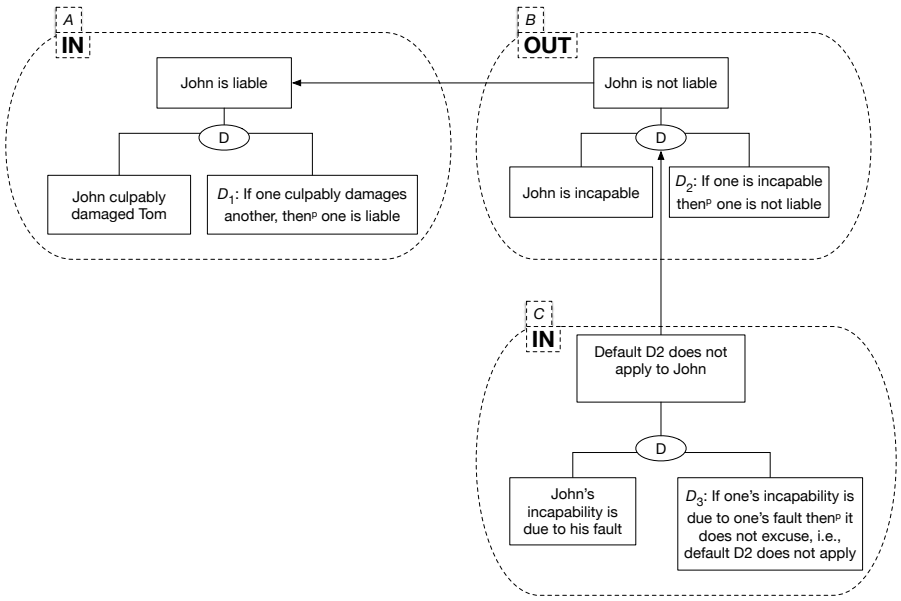

FIGURE 11. Reinstatement

A that is defeated by a counterargument B can still be acceptable when B is in turn defeated by a further argument C: we would have rejected A if we had accepted B , but since we do not accept B (given that it is defeated by C), then A remains acceptable.

To clarify this point it is useful to specify the conditions that an argument should meet to be IN (acceptable) or OUT (inacceptable). The basic idea is that only a defeater which is IN can turn OUT the argument it attacks; a defeater which is OUT is not relevant to the status of the argument it attacks. Thus, we can state the following rules:

(1) An argument $\mathcal{A}$ is IN iff no argument which defeats $\mathcal{A}$ is IN.
(2) An argument $\mathcal{A}$ is OUT iff an argument which defeats $\mathcal{A}$ is IN.

To clarify our analysis let us consider the legal example in Figure 11, which extends Figure 11 with labels denoting the statuses of the corresponding arguments:

Relative to the set the arguments in Figure 11 (A, B , and C), argument C is necessarily IN, since no defeater questions its status. Therefore, argument B is OUT (having a defeater, namely A, which is IN). Consequently, argument A is IN, since it has no defeater which is IN. This is the only assignment of IN and OUT labels that is consistent with rules (1) and (2). Consequently, A is justified, and so is its conclusion (John is liable), B is overruled, and C is justified.

This example shows the connection between dialectics and nonmonotonicity. By introducing new arguments into an argument framework (typically, the set of the arguments proposed in a debate or constructible from a given set of premises), the status of the pre-existing arguments may change relative to that framework: arguments that were justified may now be overruled and arguments that were overruled may now be justified.

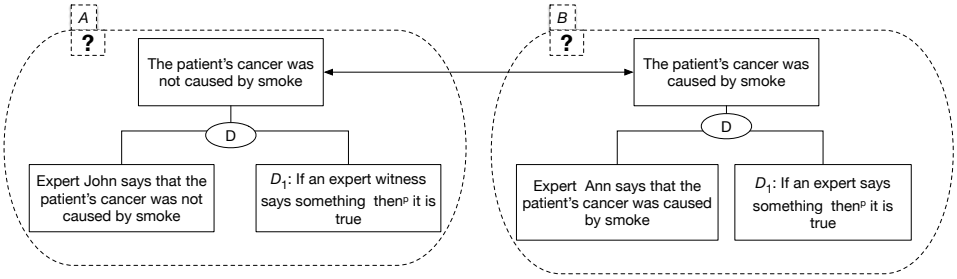

FIGURE 12. Undecided conflict

Rules (1) and (2) above fail to univocally determine the status of those arguments in cases where we have an unresolved conflict (see Figure 12, which depicts the divergent opinions of two experts).

In such a case, which arguments are justified depends on where we start from: if arguments $\mathcal{A}$ and $\mathcal{B}$ attack each other (and neither of them is OUT on other grounds), then if we assume that $\mathcal{A}$ is IN then $\mathcal{B}$ will be OUT, and if we assume that $\mathcal{B}$ is IN, then $\mathcal{A}$ will be OUT. We can deal with this situation by considering all possible assignments of IN and OUT labels to the arguments at stake, consistently with rules (1) and (2) above: an argument is justified if it is IN according to every assignment; it is overruled if it is OUT according every assignment; it is defensible if it is IN according to some assignment and OUT according to some other assignment (Pollock 2008). An equivalent approach by which to assess the status of arguments consists in constructing alternative extensions, namely, maximal sets of consistent arguments: justified, defensible, and overruled arguments are contained in all, some, or no extensions (Dung (1995)).

Unresolved conflicts concerning legal and factual issues are addressed in different ways in the adversarial context of legal disputes, where the judge is assumed to know the applicable law, while the parties should bring evidence on the facts of the case. If an unresolved conflict between competing arguments concerns a legal issue (e.g., there are arguments supporting alternative interpretations of the same source of law), the decision-maker (the judge) should resolve the conflict by assigning priority to one of the conflicting arguments (on defeasible reasoning in legal interpretation, see Walton et al. 2016). If the unresolved conflict concerns a factual premise that is needed to construct an argument, it will be assumed that the factual premises have not been legally substantiated; therefore the factual argument will be OUT.

The dialectical interaction between arguments and counterarguments is reflected in the allocation of burdens of proof and, more generally, of burdens of argumentation. The idea of the burden of proof applies to many dialectical interaction, in context-dependent ways (see (Walton 2008,59)), but it acquires a specific significance in the law (see Sartor 1993, Prakken and Sartor 2009). In a legal case, the party that is interested in establishing a legal outcome bears the burden of presenting and substantiating an argument supporting that outcome. For instance, in the example in Figure 13, plaintiff Mary must provide argument A, establishing John's liability for negligence. She must substantiate the argument's normative premises (the general rule of civil liability) by referring to sources of law,

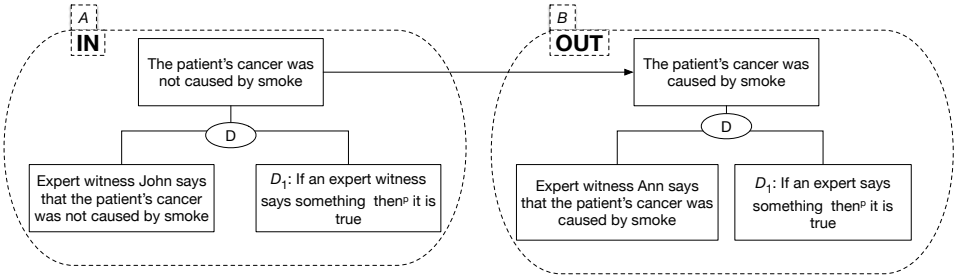

FIGURE 13. Burden of proof

and its factual premise (John culpably damaged Mary) by bringing in appropriate evidence. This argument will be sufficient for Mary to win the case if its premises are accepted and no counterarguments defeating it are provided by John (at least with regard to factual premises, since the judge may independently bring in legal information).

Thus, Mary can be said to bear the burden of proving that John did damage to her, since without establishing this fact she will not be able to construct argument A , which supports the outcome she favours. She does not bear the burden of proving that John was not incapable, since to build argument A , she does not need to establish that fact. On the contrary, John bears the burden of proving that he was incapable, since without establishing this fact, he will not be able to substantiate argument B, which could defeat Mary's argument A (switching A 's status from IN to OUT, in the absence of further interfering arguments).

In general, when a party π_1 fails to construct a certain legally acceptable argument $\mathcal{A}$ supporting her side unless evidence is provided for premise P, we say that π_1 has the burden of proof regarding P. This does not mean that the counterparty π_2 has no interest in P. It is true that π_1 will fail to build the argument based on P if π_1 fails to provide evidence for P, even if π_2 remains inactive. However, if π_1 provides sufficient evidence for P (and the other premises of $\mathcal{A}$ have been established), then π_2 must provide evidence against P, or other counterarguments against $\mathcal{A}$, if he does not want to lose on the basis of $\mathcal{A}$ (on the logic of the burden of proof, and for further refinements, including the distinction among the burden of production, the burden of persuasion, tactical burden, and standards of proof, see Prakken and Sartor 2009, on the connection between defeasibility and proof, see also Sartor 1994, Brewer 2011, Duarte d'Almeida 2013). Figure 13 below exemplifies the context of the burden of proof. Let us assume that the plaintiff (the alleged victim) has the burden of showing that his cancer was caused by smoke and that the standard of preponderance of the evidence applies. Then, even if the two arguments have equal weight, the plaintiff's argument would be strictly defeated by the defendant's argument (to defeat the defendant's argument, the plaintiff's argument must meet the required standard of proof). Thus, in an adversarial legal context governed by the burden of proof, the status assignment of Figure 12 (no justified arguments, two defensible one) would be transformed into the assignment of Figure 13 (one justified and one overruled argument).

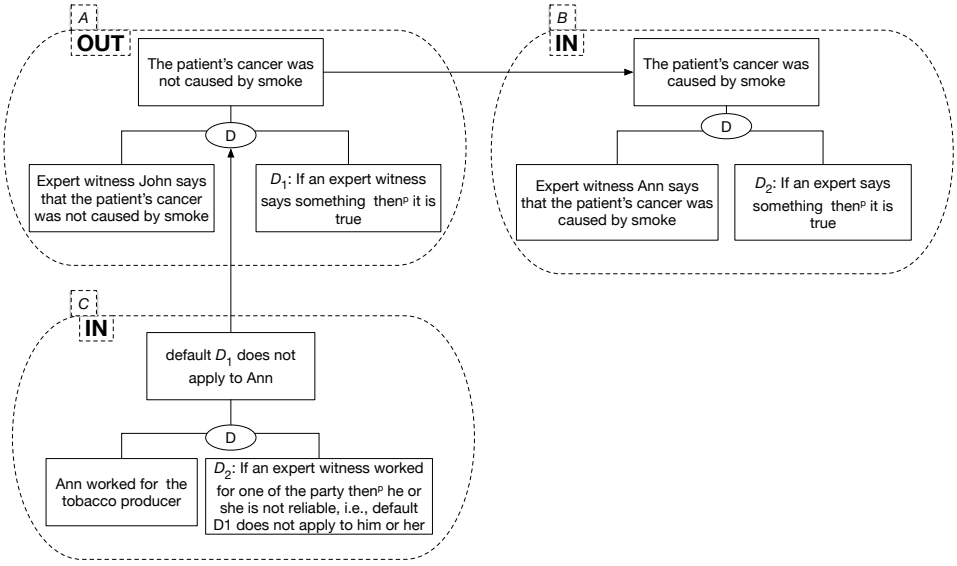

FIGURE 14. Burden of proof and reinstatement

However, the patient may address this situation not only by providing additional evidence, so that his argument outweighs the doctor's argument, but also by undercutting the doctor's argument, e.g., by successfully contesting the reliability of the expert testimony in defence, as shown in Figure 14

1.9. Dynamic Priorities

In the previous examples involving priorities over arguments, we assumed that priorities were given. However, even priorities may be determined by (defeasible) arguments. Usually, a conflict between competing arguments is adjudicated according to the comparative strength of the defaults included in the such arguments. Therefore, priority arguments aim to establish the comparative strength of conflicitng defaults. In the legal domain, where legal norms provide the relevant defaults, priority arguments may appeal to formal legal principles — i.e., criteria which do not refer to the content of the norms at issue— such as the preference accorded to the more recent laws (*lex posterior derogat legi priori*), to the more specific ones (*lex specialis derogat legi generali*), or to those issued by a higher authority (*lex superior derogat legi inferiori*). Priority arguments may also be supported by textual clues, e.g., norms having negative conclusions are usually meant to override previous norms having the corresponding positive conclusions. Finally, priority arguments may refer to the substance of the norms at issue, e.g., assigning priority to the norm that promotes the most important values (legally valuable interests) to a greater extent.

One way to deal with the argumentative role of priority arguments consists in extending the IN and OUT labelling to defeat links between arguments. The previous rules can then be rewritten as follows:

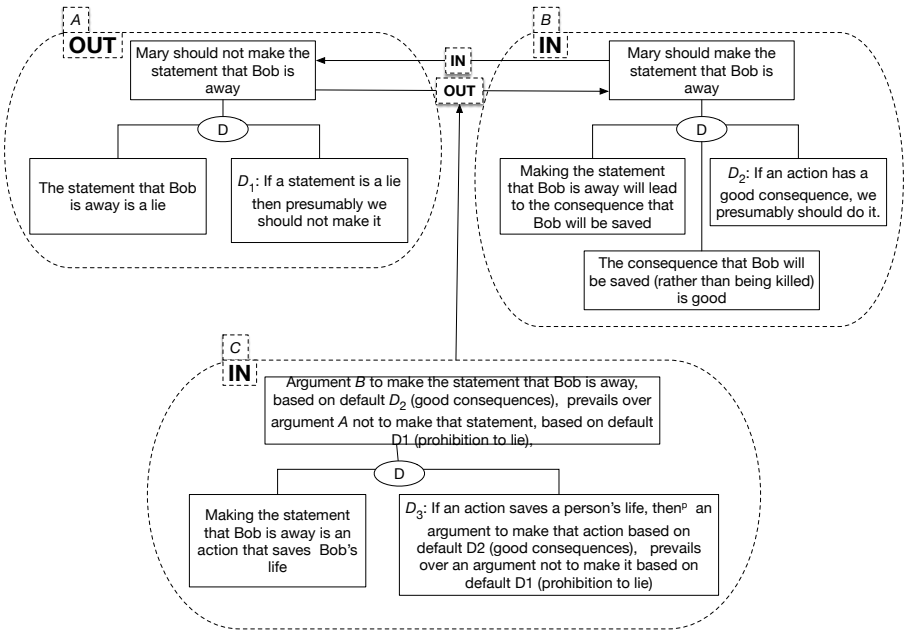

FIGURE 15. Dynamic priorities

(1) An argument $\mathcal{A}$ or a defeat link d is IN iff no argument which is IN defeats $\mathcal{A}$ respectively or $\mathcal{L}$ through a defeat link which is IN.
(2) An argument $\mathcal{A}$ or a defeat link d is OUT iff an argument which is IN defeats respectively $\mathcal{A}$ or d through a defeat link which is IN.

We need to specify when a defeat link is defeated: An argument $\mathcal{C}$ defeats the defeat-link d denoting a rebutting attack from $\mathcal{A}$ to $\mathcal{B}$ when $\mathcal{C}$ states that $\mathcal{B}$ prevails over $\mathcal{A}$.

To clarify this idea let us return to the issue of the admissibility of lying to save a person's life. Let us now add a priority argument (C) stating that, since the statement that Bob is away will save Bob's life, the duty to make the statement, as supported by the argument from good consequences, outweighs the duty not to make it, as supported by the prohibition on lying (Figure 15).

Argument C affirms that argument B (for the duty to say that Bob is away) is stronger than argument A (for the duty not to make that statement). Therefore, we can conclude that the defeat link from A to B is OUT (as a weaker argument cannot rebut a stronger one), while the defeat link from B to A remains IN. Therefore, B strictly defeats A (it defeats it without being defeated by it). Consequently, B is IN and A is OUT: we should tell the lie.

Obviously, the opposite conclusion would follow if we took a different view of priorities, such as the view that deontological arguments, warranted by generalizable rules, always have priority over consequentialist arguments.

1.10. Patterns of Defeasible Reasoning

Various warrants (general defaults) for defeasible reasoning can be identified. The following ones are discussed by (Pollock 1998, 2008):

- *Perceptual inference.* If I have a percept with content P , then I can presumably conclude that P is true. For instance, if I have an image of a red book at the centre of my field of vision, I can conclude that there is a red book in front of me. This conclusion is defeated if I become aware of circumstances that do not ensure the reliability of my perceptions (I am watching a hologram).
- *Memory inference* . If I remember P , then I can presumably conclude that P is true. For instance, my recollection that yesterday I had a faculty meeting lends presumptive support to the conclusion that there was such a meeting. This inference is defeated if I come to believe that my supposed recollection was an outcome of my imagination.
- *Enumerative induction* . If I observe a large enough sample of F s, all of which are G s, then I can presumably conclude that all F s are G s. For instance, if all crows I have ever seen are black, then I can presumably conclude that all crows are black. This inference is defeated if I should see a white crow.
- *Statistical syllogism* . If most F s are G s and an individual a is an F , then I can presumably conclude that a is a G . For instance, assume that (1) the pages of most printed books are even-numbered on their verso side and that (2) the bound pages on my table are a printed book. I can then conclude that these bound pages are even-numbered on their verso side. This inference is defeated if I discover that these bound pages were incorrectly printed with even numbers pages on their recto side.
- *Temporal persistence* . If it is the case that P at time t_1, then presumably P is still the case at a later time t_2. For instance, if my computer was on my table yesterday evening (when I last saw it), then presumably it will still be there. This inference is defeated if I come to know that the computer was moved from the table after I last saw it, and more generally if I have any reason to believe that its location may have changed.

General processes of human cognition, such as abduction and analogy (see Walton et al. 2008) can support further schemes for defeasible arguments, such as the following:

- *Abduction of a cause.* If Q is the case, and P causes Q, then presumably P was the case. For instance, if the grass is wet, and rain causes the grass to be wet, then presumably it has rained. Arguments based on this warrant can be defeated in different ways: by indicating alternative, no less probable, causes of the effect (e.g., somebody has watered the grass), or by showing an inconsistency between the cause (rain) and other states of affairs (e.g., the street is not wet, as it should be if it had rained), etc.
- *Basic analogy.* If P is relevantly similar to Q, and P has property R, then presumably also Q has property R. For instance, if detecting something by just seeing does not count as search, and detecting something by just seeing is relevantly similar to detecting drug with a sniffing dog, then also detecting drug with a sniffing dog does not count as search (for refinements

of the analogy pattern, and for a discussion of the sniffing dog case, see Walton et al. 2008,Ch. 2). Arguments based on analogy can be attacked by questioning or denying that there is a relevant similarity, by pointing to relevant differences, by bringing counterexamples, etc. (I cannot enter here the discussion on what may count as a relevant similarity or difference). In many cases, an analogical conclusion can (or should) also be supported by a more elaborate piece of reasoning, where the aspects that make the similarity relevant are presented as the antecedents of a general warrant (e.g. detecting something without actively interfering does not count as search) which is abducted to explain the common conclusion, (see Brewer 1996 Walton et al. 2008,Ch. 2).

These defeasible warrants are not meant to substitute the logical, philosophical or psychological theories of the phenomena they address, such as perception, induction, abduction or analogy (see for instance, for analogy, Holyoak and Thagard 1996). They should be rather viewed as rules of thumb that may be supported, explained and constrained by such theories.

In the previous examples, I have considered further general defaults, such as those enabling the argument from good or bad consequences or the argument from expert testimony. I have also observed that more specific defaults may be used to construct defeasible arguments: empirical generalisations, as well as legal and moral norms, can be viewed as defaults. In fact, the set of the defaults that may be used in individual and social cognition cannot be reduced to an exhaustive list, since default warrants are justified pragmatically, i.e., because of how well they serve the needs of different practical or epistemic activity types (Walton and Sartor 2013). The successful use of a default warrant in a social activity (such as legal reasoning) critically depends on the extent to which the scheme enjoys shared acceptance, as providing valid support to its conclusions (since the default's acceptance is a crucial precondition of its successful use in arguments meant to convince other people, or to converge with them into shared conclusions). Thus, even abstract legal principles, such as interpretive canons, only justify their conclusions in those legal systems in which they are in fact endorsed and deployed, so as to enjoy the status of social and institutional normative principles.

It is important to stress that defeasible arguments can include multiple steps. For instance, in an argument culminating in the conclusion of a rule, the rule may be supported by an interpretive argument, while rule's factual antecedent may result from arguments assessing the available evidence. Consider the liability case illustrated in Figure 16. The argument for the liability of Doctor Mary includes the following:

- A norm-based argument that Mary is liable, since she harmed her patient and doctors are liable for harming their patients unless they are shown not to be at fault.
- A teleological interpretive argument (a subspecies of the argument from good consequences): the law on doctors' liability must be interpreted in this way, since this interpretation contributes to increasing diligence in the medical profession, which is a good thing.
- An empirical argument based on an expert testimony supporting the conclusion that there was a causal link between the Mary's behaviour and the patient's harm.

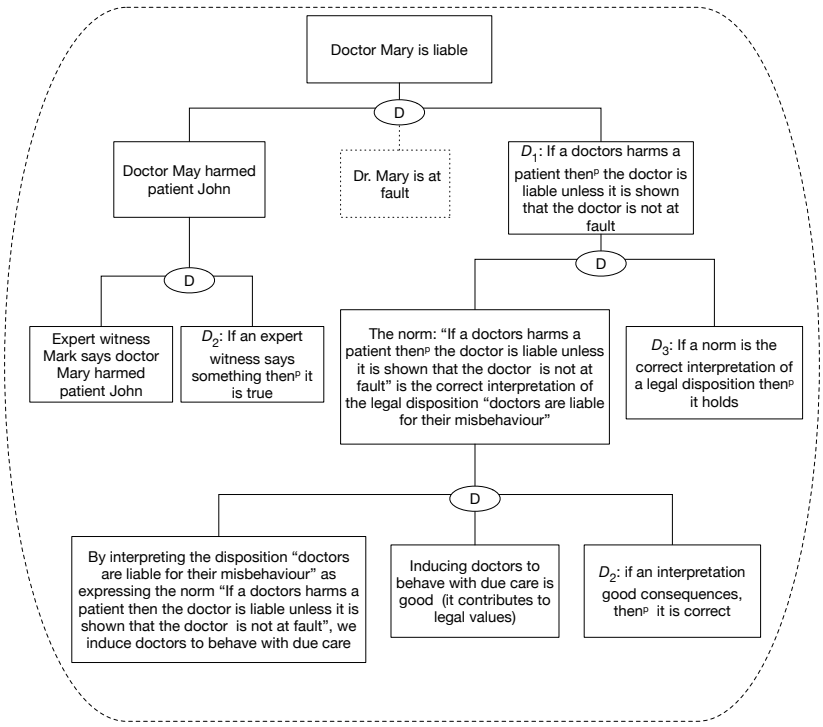

FIGURE 16. Multistep argument

This argument is subject to a series of possible attacks, against each of its subarguments (a subargument being an argument which is included in a larger argument): its top subargument may be undercut by establishing that Mary was not at fault (she used the available medical knowledge correctly); the interpretive subargument can be attacked by contesting the very idea that the proposed interpretation promotes careful behaviour among doctors (on the contrary, it may undermine patient care, since doctors may become too risk-averse, knowing that they may face the difficult task of proving a negative, namely, that they did not act negligently); the empirical subargument can be rebutted by providing a contrary expert opinion, or it can be undercut by challenging the expert's reliability, among other options.

1.11. Legal Systems as Argumentation Bases

We have so far considered arguments and their interactions, i.e., conflicts giving rise to defeat relations. Let us now look at the set of premises that provide the ingredients for constructing a set of interacting arguments.

A set of such premises is not a consistent set of deductive axioms but is rather a repository of materials to be used to build competing arguments and counter-arguments. It is an *argumentation basis,* in the sense of a knowledge base (a set of premises) that can be used for constructing an *argumentation framework* (a set

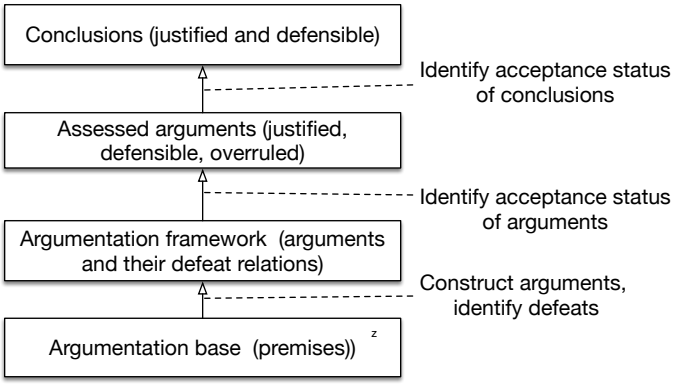

FIGURE 17. Inferential semantics of an argumentation basis

of interacting arguments). In Sartor (1994) I used the term *argumentation framework* (see also Stone Sweet 2004,34) to denote what I here call *argumentation basis*. Here I reserve the term *argumentation framework* to the set of arguments that are constructible from the argumentation basis (Baroni et al. 2011).

Figure 17 (adapted from Baroni et al. 2011) shows a process to determine the inferential semantics of an argumentation basis, namely, the set of all conclusions that are supported by that basis. First, we construct the maximal argumentation framework resulting from the argumentation basis, i.e., we build all arguments that can be obtained by using only the premises in the basis and we identify all defeat relations between such arguments. Then we determine what arguments and defeat links are IN or OUT (for all or some labellings), and consequently establish the status of each argument, i.e., whether the argument is justified, defensible or overruled relatively to the given argumentation basis. Finally, we identify the status of the conclusions of these arguments: the conclusion of justified or defeasible arguments being respectively justified or defensible relatively to the argumentation basis. A different (but equivalent) approach is described in Prakken and Sartor (1997), where the proof of a defeasible conclusion takes place in a game where the proponent of that conclusion has to build an argument (from the argumentation base) and defend it against all possible direct and indirect counterarguments an opponent my construct (from the same argument base)

I shall argue that a legal system itself —considered from an argumentation standpoint, and complemented with the relevant factual evidence— indeed appears to be an argumentation basis rather than a deductive system. In fact, if we accept that the legal system contains general rules and exceptions, conflicting norms, principles expressing incompatible legal interests, argument schemes (abstract warrants) warranting alternative inferences, then we must reject the traditional postulate of the consistency of the law, and consequently we must reject the law's image as an axiomatic base that, when combined with the relevant facts, yields conclusive deductive implications.

On the contrary, a legal system is a heterogeneous, stratified, and conflicting set of legal defaults (legal rules and principles, cases, metarules, accepted argument schemes, etc.) which, when combined with the relevant facts, make it possible to

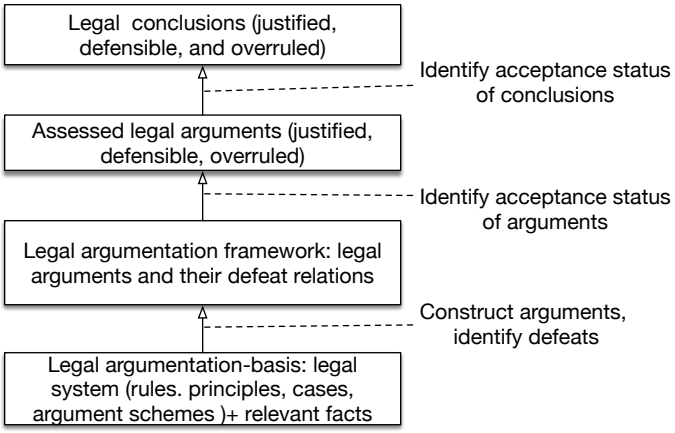

FIGURE 18. Inferential semantics for the law

derive presumptive conclusions. By complementing a legal system (the relevant portion of it) with the evidence establishing the operative facts of a case (facts that match the antecedents of some of the system's norms), we obtain an argumentation basis from which competing presumptive arguments may be constructed. This is shown in Figure 18, that applies to the law the inferential model of Figure 17.

To clarify this idea, let us assume, for simplicity's sake, that the legal system $\mathbb{L}$ in question only contains the three defeasible rules on civil liability included in the arguments in Figure 11 above:

D_1: If one culpably damages another, one is liable: $CulpablyDamages(x, y) \Rightarrow Liable(x)$.
D_2: If one is incapable, one is not liable: $Incapable(x) \Rightarrow \neg Liable\ (x)$.
D_3: If one's incapability is due to one's fault, then it does not excuse, i.e., default D_2 does not apply: $IncapableByFault(x) \Rightarrow negD_2(x)$.

The three factual propositions (possible operative facts) that match the antecedents of these three rules are the following:

P_1: John culpably damages Tom: $CulpablyDamages(John, Tom.$
P_2: John was incapable: $Incapable(John)$.
P_3: John's incapability is due to his fault: $IncapableByFault(John)$.

By complementing $\mathbb{L}$ with appropriate facts (any combination of P_1, P_2 and P_3) we obtain argumentation bases that make it possible to construct different combinations of arguments A , B, and C (different facts being required for each of these arguments).

All these arguments are in principle defeasible, being susceptible to rebuttal or undercutting by appropriate counterarguments, should the latter become available. However, only A and B and can be defeated by counterarguments constructed with

the norms in $\mathbb{L}$, plus corresponding operative facts, since $\mathbb{L}$ does not contain any default that may be used to build a defeater to C.

Let us consider, for instance, argument A in Figure 11. This argument can be constructed from $\mathbb{L}$, complemented by the factual proposition F_1, since the premises for A are constituted by default D_1, which belongs to $\mathbb{L}$, and fact F_1. We can say that argument A can be defeated in $\mathbb{L}$, to mean that $\mathbb{L}$, complemented with appropriate facts, provides the resources for constructing a defeater to A. In fact, A is strictly defeated by B, which can be constructed from $\mathbb{L}$, complemented with factual proposition P_2. Also B can be defeated in $\mathbb{L}$, since B is defeated by C, which can be constructed from $\mathbb{L}$, complemented with the factual proposition P_3. On the other hand, C, while also being a defeasible argument, cannot be defeated in $\mathbb{L}$, since there is no operative fact that would make it possible to rebut or undercut C using only the rules in $\mathbb{L}$.

Note that the fact that an argument can be defeated in $\mathbb{L}$ does not mean that the argument fails to be justified in every argumentation basis obtainable by adding an appropriate set of operative facts to $\mathbb{L}$. For instance, if only the fact that John culpably damaged Tom is added to $\mathbb{L}$, we obtain the argumentation basis $\mathbb{L} \cup \{P_1\}$, from which we can only build argument A . Since no counterargument to A can be constructed from $\mathbb{L} \cup \{P_1\}$, $A$is justified relative to argumentation basis $\mathbb{L} \cup \{P_1\}$ and so is his conclusion: John is liable. If we also add the fact that John was incapable, we obtain the argumentation basis $\mathbb{L} \cup (P_1,\ P_2\}$, relatively to which A is no longer justified, since A's strict defeater B can be constructed. Relatively to $\mathbb{L} \cup \{P_1,\ P_2\}$, B is justified and so is his conclusion: John is not liable. Similarly, A would again be justified, and B would be overruled, relatively to the argumentation basis $\mathbb{L} \cup \{P_1, P_2, P_3\}$, which makes it possible to construct argument C. Thus, relatively to $\mathbb{L} \cup \{P_1, P_2\}$, which originates the argumentation framework $\{A, B, C\}$,in which A's conclusion is justified: John is liable.

An argument that cannot be defeated in a normative system $\mathbb{L}$ may be defeated in larger normative system. Assume, for instance, that through a legislative act or through judicial interpretation, a new norm D_4 is introduced, which is stronger than D_3:

> D_4: If one's incapacity is due to a chronic condition (alcoholism or drug addiction), then the incapacity excuse, i.e., default D_2, does apply: IncapableByChronicalCondition $(x) \Rightarrow\ D_2(x)$.

Then argument C, which could not be defeated in $\mathbb{L}$, can be strictly defeated in $\mathbb{L}' = \mathbb{L} \cup \{D_4\}$. In fact, $\mathbb{L}'$, in combination with the operative fact:

> P_4: John is incapable by a chronical condition (e.g.,alcoholism): IncapableByChronicalCondition $(John)$)

enables us to construct a further argument, let us call it G, that strictly defeats C . Thus, relatively to the argumentation basis $\mathbb{L}' \cup \{P_1,\ P_2,\ P_3,\ P_4\}$, that originates the argumentation framework $\{A,\ B,\ C,\ G\}$, argument A is overruled, while argument B is justified, and so is its conclusion that John is not liable. as shown in Figure 19.

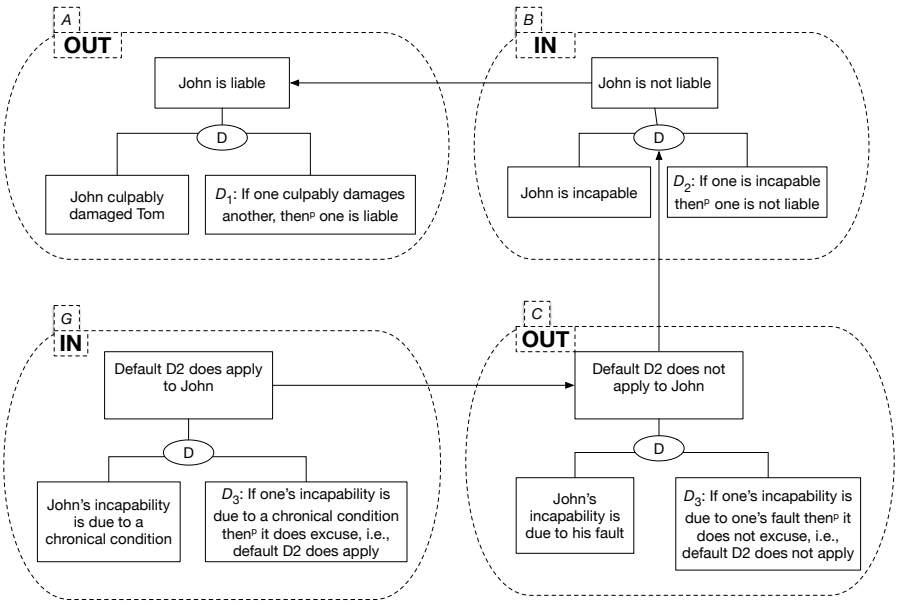

FIGURE 19. Defeat relative to an argumentation basis

CHAPTER 2

Defeasible cognition in the law

This chapter addresses the rationale for defeasibility in law, along with the possibility of using different approaches, such as revision or probability, to deal with uncertainty in legal reasoning. Finally, an account is provided of the emergence of theories of defeasibility in philosophy, logic, and legal theory.

2.1. The Rationale for Defeasibility in Human Cognition

Pollock (1998) argues that defeasibility is a key aspect of human cognition (and more generally, of the cognition of any boundedly rational agent). We start with perceptual inputs and proceed by inferring beliefs from our current cognitive states (our percepts plus the beliefs we have previously inferred). A process so described must satisfy two apparently incompatible desiderata:

- We must form our beliefs on the basis of partial perceptual input (we cannot wait until we have a complete representation of our environment).
- We must be able to take an unlimited set of perceptual inputs into account.

According to Pollock, the only way to reconcile these requirements is by defeasible reasoning. We must adopt beliefs on the basis of a small set of perceptual inputs, but then must be ready to retract these beliefs in the face of additional perceptual inputs, whenever these additional inputs conflict with the initial basis for our beliefs.

Thus, defeasible reasoning appears to have different, but related, functions (see Sartor 2005,Section 2.2, 2.3). The first function consists in providing us with provisional beliefs, on which basis we can reason and act, until we gain information to the contrary.

The second function consists in activating a structured process of inquiry that consists in drawing *pro tanto* conclusions, looking for their defeaters, for defeaters of defeaters, and so on, until stable outcomes are obtained. This process has two main advantages: (1) it focuses the inquiry on relevant knowledge, and (2) it continues to deliver provisional results while the inquiry moves on.

A third function of defeasibility consists in enabling our collective knowledge structures to persist in time, i.e., to continue to work as a shared communal asset, even though each of us is exposed to new information, often challenging the information we already have.

We indeed have two basic strategies for coping with the provisional nature of human knowledge: revision and defeasibility.

Revision assumes that our general knowledge is a set of universal laws. When we discover a case where such universal laws lead us to a false (unacceptable or absurd) conclusion, we must conclude that our theory (or the subsets of it entailing the false conclusion) has been falsified, becoming thus unacceptable (Popper 1959).

Thus, we must abandon some propositions in that theory and replace them with new universal propositions, from which the false conclusion is no longer derivable. Rational strategies for revising a theory have been the object of several studies (see, for instance, Alchourrón et al. 1985, Gärdenfors 1987). In the legal domain, this idea was originally proposed Alchourrón and Makinson (1981) and was subsequently developed by Maranhão (2013).

The other strategy, *defeasibility*, assumes that general propositions are defaults, that are meant govern most cases or the normal cases. Thus, we can consistently endorse such propositions and deny that they apply to certain cases: the exception serves the rule, or at least it does not compromise the rule. To deal with an anomalous case on a defeasibility strategy, we do not abandon the default or change its formulation, but instead we assume that the default's operation is limited on grounds that are different from those that support the use of the default itself. As we saw in the previous example, these grounds may provide an argument that undercuts or rebuts the argument warranted by the default. The idea that legal norms are defaults (rather than strict rules) makes possible a certain degree of stability in legal knowledge: we do not need to change our norms whenever their application is limited through subsequent exceptions or distinctions. However, this perspective does not exclude the need to abandon a norm, when it no longer reflects a "normal" connection, being superseded by subsequent norms (as in implicit derogation), or when it is explicitly removed from the knowledge base (as in explicit derogation: see Governatori and Rotolo (2010)).

2.2. Defeasible Reasoning and Probability

Probability calculus—especially its versions based on the idea of subjective probability—provides an attractive alternative to defeasible reasoning as a method for dealing with limited and provisional information. It has a rich history of successful applications in many domains of science and practice, including legal practice (though its legal applications are still controversial: see Fenton et al. 2016) and has recently found many applications in artificial intelligence.

Consider, for instance, a case where Tom was run over by a car carrying Mary and John, and in which it is not clear who was driving at the time of the accident.

On the probabilistic approach, conflicting evidence does not lead us to incompatible belief —like the belief that John was driving the car when the car ran over Tom, and the belief Mary was driving the car on the same occasion— between which a choice is needed. We rather come to the consistent view that incompatible hypotheses have different probabilities. For instance, on the basis of the available evidence, we may consistently conclude that there is a 40 percent probability that John was driving, and a 60 percent probability that Mary was doing it. Probabilistic inference uses probability calculus to determine the probability of an event on the basis of the probability of other events. For instance, if there is an 80 percent probability that Tom will have problems walking because he has been run over, there is a 32 percent probability (40 percent * 80 percent) that Tom will have such problems having been run over by John, and a 48 percent chance (60 percent * 80 percent) that he will have such problems having been run over by Mary. Here I cannot enter probability calculus or discuss the many difficult issues related to it, especially when ideas of probability and causation are combined, or when Bayesian

reasoning is used to determine the probability of a hypothesis in light of the evidence. I will merely highlight three issues that make probability calculus inadequate as a general approach for dealing with uncertainty in legal reasoning.

The first issue is that of practicability: we often do not have enough information to assign numerical probabilities in a sensible way. For instance, how do I know that there is a 40 percent probability that John was driving and a 60 percent probability that Mary was driving? In such circumstances, it seems that we must attribute probabilities arbitrarily or, no less arbitrarily, we must assume that all alternative ways in which things may have turned out have the same probability.

The second issue is conceptual: although it makes sense to ascribe probabilities to factual propositions, it makes little sense to assign probabilities to legal rules and principles, unless we are making predictions. A legal decision-maker does not usually decide to use a normative premise by assessing the probability that the premise holds.

The third issue relates to psychology: humans tend to face situations of uncertainty by choosing to endorse hypothetically one of the available epistemic or practical alternatives (while keeping open the chance that other options may turn out to be preferable), and by applying their reasoning to this hypothesis (while possibly, at the same time, exploring what would be the case if things turn out to be different). We do not usually assign probabilities and then compute what further probabilities follow from such an assignment. When we have definite beliefs or hypotheses, we are usually good at developing inference chains, storing them in our minds (keeping them dormant until needed), and then retracting any such chains when one of its links is defeated. Conversely, we are bad at assigning numerical probabilities, and even worse at deriving further probabilities and revising probability assignments in light of further information.

Our inability to work with numerical probabilities certainly figures among the many failures of human cognition (like our inability to quickly execute large arithmetical calculations). In fact, computer systems exist which can handle efficiently complex probability networks (otherwise termed *belief networks* or *Bayesian networks*). They perform very well in certain domains by manipulating numerical probabilities much faster and more accurately than a normal person (see (Russell and Norvig 2010,Ch. 13)). However, our bias toward exploring alternative scenarios, and defeasibly endorsing one of them, does have some advantages: it focuses cognition on the implications of the most likely situations, it supports making long reasoning chains, it facilitates building scenarios (or stories) which may then be evaluated according to their coherence, it enables us to link epistemic cognition with binary decision-making (it may be established that we have to adopt decision Q if P is the case, and NON-Q if P is not the case). There is indeed psychological evidence that humans develop theories even under situations of extreme uncertainty, when no reasonable probability assignment can be made.

The limited applicability of probability calculus in many domains does not exclude that there may be various practical and legal issues where statistics and probability provide decisive clues, as when scientific evidence is at issue.

Recently, approaches have been developed that try to combine defeasible reasoning and probability by working out the likelihood that different premises and combinations of them will be used in making arguments and that these will interact with other arguments. Such approaches would lead to probabilistic refinements of

the IN and OUT labelling previously considered: rather than just saying that an argument is IN or OUT, we could establish that it has a certain probability of being IN or OUT relative to an argumentation basis whose premises or combinations of them are assigned certain probabilities (Riveret et al. 2012, Hunter 2013).

2.3. Defeasibility in the Law

Defeasible reasoning characterises the law at different levels.

First, clues to the defeasibility of legal reasoning are embedded in the very language of legal sources. As we saw in the previous example, the legislator itself often suggests how to construct defeaters to certain arguments. For example, to indicate that liability in tort can be excluded by appealing to self-defence or a state of necessity, the legislator may use any of the following formulations:

- *Unless clause* . One is liable if one voluntarily causes damage, unless one acts in self-defence or in a state of necessity.
- *Explicit exception* . One is liable if one voluntarily causes damage. One is not liable for damages if one acts in self-defence or in a state of necessity.
- *Presumption* . One is liable if one voluntarily causes damage and one does not act out of self-defence or a state of necessity. The absence of both is presumed.

According to all these formulations, to build an argument to the effect that one must make good some damage, it is normally sufficient to ascertain that one voluntarily caused that damage, but this argument is defeated by counterarguments appealing to the fact the person turns out to have acted either out of necessity or in self-defence.

Defeasibility is also an essential feature of *conceptual constructions* in the law. Legal concepts must be applied to such a diverse range of instances that they can at best offer a tentative and generic characterization of the objects to which they apply, a characterization that must be supplemented with exceptions. General legal concepts presuppose defeasibility: the requirement of absolute rigour in defining and applying concepts—the demand that all features which are included in, or entailed by, a concept apply to each of its instances—would paradoxically run counter to the very possibility of being "logical" in the sense of using general concepts. In fact, even the definitions of the legal concepts that can be found in statutes and codes reflect the stepwise defeasible process of establishing legal qualifications: first, a general discipline is established for a certain legal genus (e.g., the genus "contract"); special exceptions are then introduced for species within this genus (e.g., the species contract of sale); finally, further exceptions may be introduced for specific subspecies (e.g., the sale of real estate). Consequently, when using conceptual hierarchies, we must apply to a certain object the rules governing the category in which it is included, but only insofar as no exceptions emerge concerning a subcategory in which that object is also included.

Defeasibility can be deliberately established by the legislator, but it may also result from the evolution of legal knowledge: after a general rule has been established, exceptions are often provided for those cases where the rule appears to be inadequate.

This is typically the evolution of judge-made law, where general *rationes decidendi* are often limited by way of *distinctions* , that is, by way of exceptions introduced for specific contexts (on defeasibility and precedents, see Prakken and

Sartor 1998, Horty 2011). In such cases, judges often leave the original default rule unchanged and add a new, prevailing rule that addresses the specific situations requiring a distinction. For instance, in the *Monge* case (US Supreme Court, 28 Feb. 1974, No. 6637), the judges introduced an exception to the idea that contracts of employment at will (lacking any set term) could be terminated by both parties regardless of the reason ("for any reason or no reason at all"). They stated that "a termination by the employer of a contract of employment at will which is motivated by bad faith or malice or based on retaliation [...] constitutes a breach of the employment contract." Correspondingly, on the basis of this rule the dismissed employee Olga Monge could build an argument (her dismissal was a breach of contract, being based on malice and retaliation) that could defeat the employer's argument that she could be legitimately dismissed on the ground that her contract was at will. Note that the judges could also have revised the original rule into a new rule: "a contract can be terminated by both parties for any reason unless the employer is terminating the contract motivated by bad faith or malice or based on retaliation." The new rule would have triggered the same dialectical exchange, as long as the unless clause was interpreted as attributing to the employee the burden of proving bad faith, malice, or retaliation.

Finally, we need to also consider the procedural aspect of defeasibility. As noted, this aspect concerns the fact that defeasible reasoning activates a structured process of inquiry in which we draw prima facie conclusions, look for their (prima facie) defeaters, look for defeaters of defeaters, and so on, until stable results can be obtained. A process like this one reflects the natural way in which legal reasoning proceeds. This is especially the case in the law's application to particular situations, when we have to consider the different, and possibly conflicting, legal rules that apply to such situations and must work out conflicts between these rules.

The defeasibility of legal reasoning also reflects the dialectics of judicial proceedings, where each party provides arguments supporting his or her position, and these arguments conflict with the arguments made by the other party. The debate of the parties is usually transferred to the judicial opinion that takes in the results of the dispute and determines its output. To convincingly justify a judicial decision in a case involving genuine issues, it is not sufficient to state a single argument; it is necessary to establish that the winning argument prevails over all arguments to the contrary, especially those that have been presented by the losing party, or that the latter arguments have to be rejected on other grounds.

Finally, doctrinal work cannot avoid being contaminated by the dialectics of legal proceedings, since its main function consists in providing general arguments and points of view to be used in judicial debates. From this perspective, doctrinal reasoning may be viewed as consisting in an exercise in *unilateral dialectics* , understood as a disputational model of inquiry in which "one develops a thesis against its rivals, with the aim of refining its formulation, uncovering its basis of rational support, and assessing its relative weight" (Rescher 1977,47).

The significance of defeasibility in legal reasoning has been recently confirmed by the psychological experiments by Gazzo Castañeda and Knauff (2016), which show how both lawyers and laypersons reason defeasibly when applying legal norms. When presented with a legal conditional, in its usual formulation (If somebody kills a person, then he or she should be punished for manslaughter), and with an instance of the antecedent condition (Bert killed a person), most participants in the

experiment conclude for the conditional's conclusion (Bert should be punished for manslaughter), but withdraw this conclusion when told that an exculpatory circumstance (because of a psychological disorder, Bert was unable to control his actions) also obtains. The experiments also show that lawyer are better than laypersons in withdrawing legal conclusion when faced with legally recognised exceptions, having a more precise knowledge of such exceptions and of their role in legal reasoning.

2.4. Overcoming Legal Defeasibility?

Some authors have suggested that the law ought to be recast into a set of deductive axioms that would lead to consistent outcomes in any possible factual situation. This reformulation of the law would eliminate normative conflicts, and therefore would leave no room for legal defeasibility. This idea has been affirmed by Alchourrón and Bulygin (1971): the legislator and the doctrinal jurist should combine their efforts towards providing axiomatic reformulation of the law, or at least of particular sections of it. Just as Euclid developed an axiomatic model of geometry, and as modern natural science and social science (especially economics) have developed axiomatic models for their theories, so the legislator and the jurist should axiomatise the law. By adding to such an axiomatisation a description of a specific case, we should obtain a set of premises from which the obligations and entitlements of the parties in the case can be deduced.

Alchourrón (1996a,b) claimed that the ideal of the axiomatization of the law should inspire legislation and doctrine. It could contribute to bringing legal studies and scientific method together: just as in science the phenomena to be explained, the *explanandum* , should be the logical consequences of a set of premises, the *explanans* , containing scientific laws and the description of particular facts, so in law the content of a legal conclusion (the decision) should be the deductive consequence of a set of premises including both general norms and the description of specific facts. Systemic interpretation should have the task of making exceptions explicit, by embedding their negation into the antecedent of the concerned legal norm (a prima facie norm "if φ then ψ" which is subject to exception χ, should be rewritten as "if φ and not χ then ψ").

It seems to me that even if such a reformulation of the law were feasible (with regard to all exceptions that could be identified by legal scholars), it is doubtful that it would be useful, i.e., that it would make the law easier to understand and apply. Legal prescriptions would need to become much more complex, since every rule would have to incorporate all its exceptions. In addition, such a representation of the law would not be able to model the dynamic adjustment that takes place—without modifying the wording of existing rules—whenever new information concerning the conflicting rules and the criteria for working out their conflicts is taken into consideration. Finally, by rejecting defeasible reasoning, we would forfeit the law's ability to provide provisional outcomes while legal inquiry moves on.

The need to represent the law in ways that facilitate defeasible reasoning does not imply that the current way of expressing legal regulations in statutes and regulatory instruments cannot be improved. On the contrary, considerable improvements in legislative technique are required to cope with the many tasks entrusted to modern legal systems. However, such improvements should not be aimed at producing a conflict-free set of legal rules, just for the sake of logical consistency. They should rather be aimed at producing legal texts that can more easily be understood and

applied. This objective requires skilful use of the very knowledge structures (such as conceptual hierarchies, speciality, or the combination of rules and exceptions) that enable defeasible reasoning.

Accepting defeasibility in the law has significant implications both for the way we use legal knowledge and for the structure of such knowledge. On the one hand, deductive inference can be complemented with defeasible arguments. On the other hand, the acceptance of defeasibility leads us to viewed the law as an argumentation basis containing conflicting pieces of information as well as the criteria for resolving some of these conflicts. It is important to stress the difference between an argumentation basis and a deductive axiomatic base. While a deductive axiomatic base is consistent and flat, an argumentation basis is conflictive and possibly hierarchical: it includes reasons clashing against one another, reasons for preferring one reason to other reasons, and reasons for applying or not applying certain reasons given particular conditions.

Both strategies just mentioned, namely, representing the law as an axiomatic base and representing it as an argumentation basis, may be justified in different contexts. The first strategy may be appropriate when we want to deepen our analysis of a small set of norms and anticipate as much as possible all instances of their application, finding a precise solution for each of them. The second strategy, however, more directly corresponds to the logical structure of non-formalized legal language (which expresses the law as setting out rules and exceptions, principles, preference criteria, etc.), and it reflects the ways in which legal reasoning proceeds when dealing conflicting pieces of information: rules and exceptions, different values needing to be balanced, different norms implementing different values, competing standards indicating what norms and values ought to prevail in case of conflict, and so on.

An argumentation basis may be transformed into an axiomatic knowledge base whose deductive conclusions include all outcomes that would be defeasibly justified relatively to the given argumentation basis (assuming that all the facts of the case are known). The dialectical interaction between reasons for and against certain conclusions, and between grounds for preferring one argument to another, would be transformed into a set of conclusive connections between legal preconditions and legal consequences. Flattening legal information in this way, however, would entail a loss of information: the deductive knowledge base would not include a memory of the choices from which it derives, and therefore it would not contain the information needed to reconsider such choices—it would not, for example, contain the information on which it was decided that a certain principle would outweigh a competing principle or that a certain interpretation was preferable. To understand the articulation of the relevant legal reasons, we would need to go back to the original argumentation basis.

Consider, for instance, the domain of privacy. Under EU regulation law the processing of personal data is admissible only for a specific purpose that is communicated to the person concerned. Moreover, such processing is in general admissible only with that person's consent. These constraints are justified by the need to protect values such as individual self-determination and dignity. However, there is a large set of exceptions to the consent principle, namely, different scenarios in which data can be processed without consent. These exceptions are justified by the need to protect the competing rights of others, as well as certain social values.

Moreover, we have cases where consent alone is insufficient to make data processing permissible, further requirements being necessary (like the authorization of a data protection authority for genetic data), and for each such exception specific rationales can be found that guide interpreters in determining the contents and limits of the exception. Finally, there may be cases where personal data may be processed even beyond the explicitly stated legislative scenarios, on the basis of an authorization which a data protection authority issues to protect the rights of others, but which overrides the right to privacy. To determine whether a data protection authority has made legitimate use of its powers, we need to consider the importance of the values at stake (privacy, freedom of expression, economic freedom, health, etc.) and evaluate whether they have been balanced in a way that respects legal (in particular, constitutional) constraints. We could try to reduce this multilevel argumentation basis to a set of flat rules, but what we would obtain is a representation removed from the original legal texts (laws, regulations, authorizations), and whose contents and rationales are much more difficult to grasp.

2.5. The Emergence of the Idea of Defeasibility in Law and Ethics

Though that formal logics for defeasible reasoning have been developed only recently, we can may find references to defeasibility in the history of philosophical and legal reasoning.

A famous fragment by Aristotle apparently characterises legal reasoning as defeasible (Aristotle, *Nicomachean Ethics* , 1137b) in the sense that legal conclusions derived from general norms may have to be rejected in the face of particular cases having exceptional features that make those conclusions inadequate:

> All law is universal, and there are some things about which it is not possible to pronounce rightly in general terms; therefore, in cases where it is necessary to make a general pronouncement, but impossible to do so rightly, the law takes account of the majority of cases, though not unaware that in this way errors are made. And the law is none the less right; because the error lies not in the law nor in the legislator, but in the nature of the case, for the raw material of human behaviour is essentially of this kind. So, when the law states a general rule, and a case arises under this that is exceptional, then it is right, where the legislator, owing to the generality of his language, has erred in not covering that case, to correct the omission by a ruling such as the legislator himself would have given if he had been present there, and as he would have enacted if he had been aware of the circumstances. (Aristotle, *Nicomachean Ethics* , 1137b)

Cicero distinguishes presumptive (probabilis) and necessary argumentation (Cicero, *De inventione* , Book 1, Section 44). He provides various patterns (warrants) for presumptive inferences: the (natural) meaning of a sign (e.g., blood traces indicate participation in a violent action), what happen usually (e.g., mothers love their children), common opinion (e.g., philosophers are atheists), or similarity (if it is not discreditable to the Rodians to lease their port-dues, then it is not discreditable even to Hermacreon to rent them). Moreover he considers how (defeasible) arguments may be refuted:

> All argumentation is refuted when one or more of its assumptions is non granted, or when, the assumptions having been granted, it is denied that the conclusion follows from them, or when it is shown that the kind itself of the argumentation is faulty, or when against a strong argumentation another argumentation equally strong or stronger is put forward (Cicero, *De inventione* , Book 1, Section 79).

The second and the fourth items in Cicero's list seem to correspond to what we called undercutting and rebutting, respectively, namely, those attacks that are peculiar to defeasible arguments.

The Aristotelian approach to the dialectics of rule and exception is developed by Aquinas:

> [I]t is right and true for all to act according to reason: And from this principle it follows as a proper conclusion, that goods entrusted to another should be restored to their owner. Now this is true for the majority of cases: But it may happen in a particular case that it would be injurious, and therefore unreasonable, to restore goods held in trust; for instance, if they are claimed for the purpose of fighting against one's country. And this principle will be found to fail the more, according as we descend further into detail, e.g., if one were to say that goods held in trust should be restored with such and such a guarantee, or in such and such a way; because the greater the number of conditions added, the greater the number of ways in which the principle may fail, so that it be not right to restore or not to restore. (Aquinas 1947,I–II, q. 94, a. 4)

The idea of defeasibility in the legal domain is precisely outlined by G. W. Leibniz, who characterises legal presumption as defeasible inference, arguing that in presumptions

> the proposed statement necessarily follows from what is established as true, without any other requirements than negative ones, namely, that there should exist no impediment. Therefore, it is always to be decided in favor of the party who has the presumption unless the other party proves the contrary. (Leibniz 1923, De Legum Interpretatione, A VI iv C 2789)

Leibniz argues that all laws are defeasible: legal norms support presumptive conclusions, which are subject to exceptions established by other norms. He also points at the connection between defeasibility and burden of proof:

> every law has a presumption, and applies in any given case, unless it is proved that some impediment or contradiction has emerged, which would generate an exception extracted from another law. But in that case the charge of proof is transferred to the person who adduces the exception. (Leibniz, De Legum Interpretatione, A VI iv C 2791)

Turning from law to morality, we can find a notion of defeasibility in the work of David Ross, an outstanding Aristotelian scholar and moral philosopher who developed a famous theory of prima facie moral obligations (Ross 2002, 1939). Espousing a pluralist form of moral intuitionism, Ross relates defeasibility to the possibility that, in concrete cases, moral principles may be overridden by other moral principles:

> Moral intuitions are not principles by the immediate application of which our duty in particular circumstances can be deduced. They state [...] prima facie obligations. [...] [We] are not obliged to do that which is only prima facie obligatory. We are only bound to do that act whose prima facie obligatoriness in those respects in which it is prima facie obligatory most outweighs its prima facie disobligatoriness in those aspects in which it is prima facie disobligatory. (Ross 1939, 84–5)

Ross links the notion of defeasibility to the idea of outweighing, a key notion in reason-based approaches to practical reasoning. The ideas that moral reasoning consists in balancing reasons and the idea of defeasibility are indeed connected, under the assumption that we can legitimately make moral assessments also on the basis of partial knowledge of the situations we face., i.e., even when we are not guaranteed to have taken into account all relevant reasons. The fact that certain reasons support a certain action only provide a defeasible support to that action: these reasons justify that action in the absence of outweighing reasons to the contrary, but would fail to support the outcome in presence of the latter reasons. Consequently, if we believe that that the reasons justifying the actions are present and we are not aware of reasons to the contrary, we should conclude that the action is presumably justified (on the basis of the information we have). If we come to believe that outweighing reasons are present, we should withdraw this conclusion.

Indeed, defeasibility may make the appeal to general ethical principles compatible with the particularistic view that any moral principle or reason may be overridden or be inapplicable depending on the circumstances (Dancy 2004). As Horty (2007, 2012) has argued, moral principles should be viewed as defaults, that link reasons to actions (or obligations to act), and support such actions as long as they are not rebutted by reasons to the contrary or undercut by reasons against their application. Logics for defeasible reasoning provide formal accounts of the view that practical reasoning consists in the assessment of competing reasons for action by a bounded cogniser.

Although the notion of defeasibility is quite familiar in legal practice and in doctrinal work, it was not extensively discussed and analysed until recently. This notion was brought to the attention of legal theorists by H. L. A. Hart (1951,152):

> When the student has learnt that in English law there are positive conditions required for the existence of a valid contract, [. . .] he has still to learn what can defeat a claim that there is a valid contract, even though all these conditions are satisfied. The student has still to learn what can follow on the word "unless," which should accompany the statement of these conditions. This characteristic of legal concepts is one for which no word exists in ordinary English. [...] [T]he law has a word

> which with some hesitation I borrow and extend: This is the word "defeasible," used of a legal interest in property which is subject to termination of "defeat" in a number of different contingencies but remains intact if no such contingencies mature.

References to the defeasibility of legal arguments can be found in important approaches to legal reasoning. For instance, Viehweg (1965) argued that lawyers approach specific problem situations, not by reasoning from a complete and consistent system of universal axioms, but by referring to an open, unordered, inconsistent, undetermined list of *topoi* (points of view, usually expressed as maxims) addressing the relevant features of the different situations that come up. Such *topoi* are usually defeasible, since they may fail to apply under particular situations. Consider, for instance, the legal *topos* that nobody can transfer to another person more rights than those he or she possesses (*nemo plus juris in alium transferre potest quam ipse habet*). This rule does not apply to some exceptional cases in which a buyer in good faith can acquire property from an apparent seller that is not the actual owner.

Similarly, Perelman and Olbrechts-Tyteca (1969)Obrechts-Tyteca (1969 focus on the distinction between deductive demonstration and argumentation, affirmed that, contrary to demonstration, argumentation is always in principle open to challenge or reconsideration (see (Blair 2012,127)).

2.6. The Idea of Defeasibility in Logic and AI

Logic and artificial intelligence have played a key role in providing a precise analysis of defeasibility (see Ginzberg 1987 for a collection of seminal contributions on nonmonotonic reasoning, Horty 2001 and Koons 2009 for a discussion of nonmonotonic logics, and (Blair 2012,Ch. 9) on defeasibility in the context of argumentation theories).

Pollock (2010)observes that Chisholm (1957) was the first epistemologist to use the term *defeasible*, taking it from Hart (1951). Among the philosophers who have addressed aspects of defeasibility is Stephen Toulmin whose approach to reasoning is based on the idea that inference rules or warrants connect data and conclusions of arguments. In the following passage he claims that some of these warrants are defeasible:

> Warrants are of different kinds, and may confer different degrees of force on the conclusions they justify. Some warrants authorise us to accept a claim unequivocally, given the appropriate data [. . .]; others authorise us to make the step from data to conclusion either tentatively, or else subject to conditions, exceptions, or qualifications (Toulmin 2003,100).

According to Toulmin, defeasibility has a special place in the law:

> Again, it is often necessary in the law-courts, not just to appeal to a given statute or common-law doctrine, but to discuss explicitly the extent to which this particular law fits the case under consideration, whether it must inevitably be applied in this particular case, or whether special facts may make the case

> an exception to the rule or one in which the law can be applied only subject to certain qualifications (Toulmin 2003,101).

Defeasibility is also addressed by Nicholas Rescher, who deals with it in connection with dialectics (Rescher 1977) and presumptive reasoning (Rescher 2006). Rescher (1977,6) describes defaults as "provisoed assertions", having the logical form P/Q and meaning that:

> "P generally (or usually or ordinarily) obtains provided that Q" or "P obtains, other things being equal, when Q does" or "when Q , so ceteris paribus does P " or "P obtains in all (or most) ordinary circumstances (or possible worlds) when Q does" or "Q constitutes prima facie evidence for P ."

The assertion of P under proviso Q , combined with the assertion of Q , constitutes and argument for P , though Q does not "entail, imply or ensure P", but makes Q only "normal, natural, and only to be expected" (Rescher 1977,7).

The most influential and comprehensive model of defeasibility is the one provided by John Pollock, who as noted introduced the ideas of undercutting and rebutting, as well as the technique of labelling defeasible inference graphs to determine their justification status (see Pollock 1995, 2010).

Particularly influential in contemporary research on informal logic has been the account of defeasible reasoning provided by Doug Walton. According to (Walton 1996,42-43)

> presumptive reasoning is neither deductive nor inductive in nature, but represents a third distinct type of reasoning of the kind classified by Rescher (1976) as plausible reasoning, an inherently tentative kind of reasoning subject to defeat by special circumstances (not defined inductively or statistically) or a particular case.

Walton et al. (2008) identify a number of distinct argumentation patterns, called argument schemes, each of which can be challenged by appropriate critical questions acting as pointers to possible defeaters.

In artificial intelligence and logic, some formal approaches have been developed to capture the normality assumption embedded in defeasible reasoning: things are assumed be normal unless we have evidence to the contrary. This assumption can be modelled by minimising the extension of predicates that express abnormality conditions (McCarthy 1980). A similar idea underlies negation by failure, used in logic programming: atomic propositions are assumed to be false unless they can be shown to be true (Clark 1978). Preferential defeasible logics (see Kraus et al. 1990) are based on the idea that the defeasible implications of a set of premises are those propositions that are true in the most normal models (situations) that satisfy those formulas.

The idea of defeasible reasoning as the application of default inference rules supporting non-deductive presumptive inferences has been developed by Reiter (1980).

An elegant and broadly scoped model of reasoning with defaults, meant to capture the link between reasons and the conclusions they favour, has recently been proposed by Horty (2007, 2012).

A large amount of AI research has been recently developed which merges defeasible reasoning and argumentation (Rahwan and Simari 2009). In particular, the abstract account of argumentation proposed by Dung (1995) has been very influential. Its abstractness lies in the fact that it focuses on attack (defeat) relations between arguments without considering these arguments' internal structure.

2.7. Defeasibility in Research on AI & Law

In AI & law, defeasible reasoning has been the subject of much research starting from the end of the 1980s. Much of this work focuses on defeasible argumentation (for a survey, see Prakken and Sartor 2015). The possibility of using negation by failure to model defeasible reasoning and burdens of proof in the law was suggested by Sergot et al. (1986). The issue of defeasibility in legal reasoning was first identified by Thomas Gordon (1988, 1995)), who later developed the Carneades system into a computable framework for defeasible reasoning (Gordon et al. 2007).

Hage (1997) proposed the idea of rule application as a general pattern for defeasible reason, where rules deliver their consequences only when they are shown to be both valid and applicable (applicability meaning, in Hage's terminology, the rule's antecedent conditions are satisfied). In his framework, a legal rule works as an exclusionary reason, such that arguments applying the rule defeat arguments based on excluded reasons (but they may be defeated by arguments based on other reasons).

Prakken and Sartor (1996) developed the first model of defeasible reasoning in law which includes reasoning with (defeasible) rules and with priorities among such rules. The model has been extended to cover the burden of proof (Prakken and Sartor 2009), and has been applied to various aspects of legal reasoning, such as reasoning with precedents (Prakken and Sartor 1998). Prakken has developed the idea of prioritised argumentation in several technical contributions (Prakken 2010, Modgil and Prakken 2013).

The idea of legal reasoning as defeasible argumentation has also been developed by Loui and Norman (1995), who have analysed the way a single defeasible legal inference may result from the compression of various inference steps, and may be attacked by unpacking it and addressing these steps.

Bench-Capon (2003) has developed the idea of value-based argumentation, namely, the idea that preferences between arguments are determined by the values endorsed by the audience to which the arguments are directed. Bench-Capon and Sartor (2003) have studied how alternative defeasible theories (sets of premises) can be constructed to explain cases, and how they may be prioritised.

Governatori et al. (2004) have shown how defeasible argumentation can be captured by using defeasible logic, in the manner originally proposed by Nute (1994). Extensions of defeasible logic have been used to capture different aspects of legal reasoning, such as the timing of legal effects (Governatori et al. 2005) and changes in the law (Governatori and Rotolo 2010).

Finally, I should mention the rich research line on the use of defeasible legal argumentation in the evidence domain and its connections with other approaches to evidence (see Verheij et al. 2016).

2.8. Defeasibility in Legal Theory

The idea of defeasibility remains highly controversial, as evidenced by the contributions contained in a recent collection (Ferrer Beltran and Ratti 2012b).

Carlos Alchourron, a leading legal logician, has opposed the ideal of defeasible reasoning, arguing for a combination of systematic interpretation and deduction: systematic interpretation should merge rules and exceptions into a coherent whole to which deduction could be applied (see Alchourrón 1996b,a). Other legal theorists, such as Alexander Peczenik (2005); Hage and Peczenik (2000,115ff.)) and NeilkMacCormick (1995), have on the contrary argued that defeasibility plays a significant role in legal reasoning (see also Brozek 2004).

It is no easy task to review the legal theorists' approaches to defeasibility, since such theorists have advanced different understandings of defeasibility, which often do not comport with the idea of defeasibility as nonmonotonic reasoning. Brozek (2014) has pointed out different ways in which defeasibility is understood in Ferrer Beltran and Ratti (2012b):

> Ferrer Beltrán and Ratti consider, inter alia, the following formulation: "a norm is defeasible when it has the disposition not to be applied even though it is indeed applicable" (Ferrer Beltran and Ratti 2012a,31). Frederick Schauer, in turn, claims that "the key idea of defeasibility [...] is the potential for some applier, interpreter, or enforcer of a rule to make an ad hoc or spur-of-the-moment adaptation in order to avoid a suboptimal, inefficient, unfair, unjust, or otherwise unacceptable, rule-generated outcome," and concludes that "defeasibility is not a property of rules at all, but rather a characteristic of how some decision-making system will choose to treat its rules" (Schauer 2012,81 and 87). Jorge L. Rodríguez says that "when we express a conditional assertion, we assume the circumstances are normal, but admit that under abnormal circumstances the assertion may become false", and—transferring this characteristic of defeasibility into the domain of law—claims that "legal rules [are defeasible since they] specify only contributory, yet not sufficient, conditions to derive the normative consequences fixed by legal system" ((Rodriguez 2012,88)). Finally, Riccardo Guastini claims that legal rules are defeasible since "there are fact situations which defeat the rule although they are in no way expressly stated by normative authorities in such a way that the legal obligation settled by the rule does not hold anymore." (Guastini 2012,183)

All the foregoing formulations point to interesting aspects of legal reasoning and to the practice of defeasible reasoning in the law. I would argue, however, that they fail to provide convincing redefinitions or clarifications of the notion of defeasibility. I have argued that defeasibility applies to three objects:

- *Arguments* . A defeasible argument is an internally valid argument that may be defeated by counterarguments that do not challenge the argument's premises but rebut its conclusions or undercut the link between its premises and its conclusion.

- *Inference* . A defeasible inference is nonmonotonic, in the sense that it makes it possible to derive conclusions that may no longer be derivable if additional premises are added.
- *Conditionals* . A conditional is defeasible when it has the logical structure of a default, i.e., when it links a merely presumptive (non-conclusive) consequent to its antecedent.

These three aspects are different faces of the same issue. A defeasible argument $\mathcal{A}$ consists in a nonmonotonic inference: if we expand the argumentation basis from which $\mathcal{A}$ is constructed with premises that enable the construction of a defeater $\mathcal{B}$ to $\mathcal{A}$, the conclusion of A will no longer justified relatively to the expanded argumentation basis, and in this sense, no longer derivable from it. Correspondingly, default conditionals make it possible to construct defeasible arguments, i.e., nonmonotonic inferences: the results obtained through defeasible modus ponens can be defeated by rebutters or undercutters.

According to this idea of defeasibility, a legal norm can be said to be *defeasible* whenever all the following conditions are jointly possible:

- The norm is accepted (being valid and being generally applicable in the special-temporal domain under consideration)
- The norm's antecedent is also accepted
- The norm's consequent is rejected.

As we have seen in the examples above (for instance, in Figure 8), a defeasible norm N can be modelled as a default, i.e., in the logical form "if $P(\mathbf{x})$ then presumably $Q(\mathbf{x})$", i.e., as $N(\mathbf{x}) : P(\mathbf{x}) \Rightarrow Q(\mathbf{x})$, where $\mathbf{x}$ is the list of the variables in the norm (and the default would stand for the set of its ground instances). In fact., the inferences (arguments) warranted by that norm—i.e., arguments having the form $(P(\mathbf{a}),\ N(\mathbf{x}) :\ P(\mathbf{x}) \Rightarrow Q(\mathbf{x}),\ therefore\ Q(\mathbf{a})$ where a is an individual case, namely, a list of values for variables $\mathbf{x}$ —can be rejected, given appropriate conditions, without rejecting the norm or is antecedent. This would happen whenever the premise for building a rebutter (a stronger norm having the form $N_1(\mathbf{x}) : R(\mathbf{x}) \Rightarrow \neg Q(\mathbf{x})$ or an undercutter (a norm having the form $N_2(\mathbf{x}) : R(\mathbf{x}) \Rightarrow \neg N_1(\mathbf{x})$) is available. This notion of the defeasibility also applies to the more abstract view of the applications of a norm as involving a meta-level warrant such as "If the norm 'If $N(\mathbf{x}) : P\ (\mathbf{x})$ then presumably $Q\ (\mathbf{x})$' is valid and $P\ (\mathbf{a})$ is the case, then presumably $Q\ (\mathbf{a})$" (as in the model proposed by Hage 1997). In the following I will speak of argument warranted by a norm, to cover both models of norm-based reasoning.

If the coexistence of the three conditions above is impossible, then the norm N at issue can be said to be *strict* or *indefeasible,* and can be modelled as a material conditional, having the form: "for all $\mathbf{x}$, if $P(x)$ then $Q(x)$", i.e. $\forall x(P(x) \rightarrow Q(x))$. More plausibly an indefeasible norm N could be represented through a universal strict conditional "for all x, if $P(x)$, then necessarily $Q(x)$", i.e, $\forall x(P(x) \twoheadrightarrow Q(x))$ (or possibly as a strict rule, which may not allowing for contraposition, see Prakken 2010). The strict conditional, which is here denoted with the arrow, $\twoheadrightarrow$ expresses the idea that the correlation between $P(x)$ and $Q(x)$ does not depend on the present factual situation (on the actual world), but would rather hold in every possible factual situation (the norm being unchanged). If we accept the indefeasible norm N and also accept that its antecedent holds in any possible context, we must accept that also the norm's consequent holds in that context. N could not be the object of exceptions in a strict sense, namely, of provisions stating that the unmodified

norm does not apply when an impeding circumstance $E(\boldsymbol{a})$ is established. To avoid the effect of N to be triggered in circumstance $E(\boldsymbol{a})$, we would have to substitute it with the new norm $\forall \mathbf{x}\ (P(\mathbf{x}) \bigwedge \neg E(\mathbf{x}) \ \twoheadrightarrow Q(\mathbf{x})$. Because of this change the norm's effect could be established in a case only when both predicates P and $\neg E$ are established in that case. Rather than E being an impeditive fact capable of blocking the application of the norm, $\neg E$ would become a negative constitutive fact that must be established, for that effect to be triggered.

Let us consider for instance a norm linking the causation of harm to the obligation to compensate the victim. If the norm were defeasible, it would mean that if any individual x culpably harms another individual y, then presumably x must compensate y, and it could be modelled in the logical form: $N(x,\ y) : CulpablyHarms(x, y) \Rightarrow MustCompensate(x, y)$. If the norm were indefeasible, it would rather mean that for all individuals x and y, if x harms y then necessarily x has to compensate y, and it could be modelled in the logical form: $\forall x,\ y(CulpablyHarms(x, y) \twoheadrightarrow MustCompensate(x, y))$.

Thus, from the perspective here developed, the defeasibility of a norm pertains to its content, as expressible in its logical form, and therefore is not affected by the fact that the norm may be declared invalid: this may happen, under appropriate conditions, for both defeasible and indefeasible norms. Similarly, the defeasibility of a norm is not affected by the fact that the norm may be modified or substituted through judicial interpretation or through legislation. Both defeasible and indefeasible norms can be modified by new legislation or case law. The difference rather pertains to the necessity of a modification to introduce an exception:

- If exceptions to a norm can be introduced without changing the norm (without affecting its content or meaning), then the norm is defeasible, regardless of whether exceptions are expressed or implicit and whether they closed or open, and regardless of what authority and procedure that is needed for introducing exceptions (for an analysis of different kinds of exceptions, see Celano 2012). In particular, it is irrelevant to the defeasibility of a norm, whether exceptions to it can be introduced through judicial interpretation, or only through legislation (or through new constitutional norms).
- If the only way to legitimately exclude the application of a norm to cases having feature E consists in changing that norm, extending in its antecedent with the negation of E ($\neg E$), then the norm is indefeasible.

In their analysis of the notion of defeasibility in the law Ferrer Beltran and Ratti (2012a, 36) distinguish three cases: (1) the norm's validity is defeasible, in the sense that it depends on defeasible criteria, (2) the norm is externally defeasible, in the sense that the "conditions of applications contain implicit exceptions whose scope has not been determined", and (3) the norm's normative content is defeasible in the sense that it specifies operative facts that are contributory conditions for the production of the norm's legal effect. In particular, they say that in the third case "the norm' s antecedent contains implicit exceptions which may not be exhaustively identified." They also affirm that in cases (1) and (2), it is not the norm itself which is defeasible, but rather the criteria for its validity or application, while the norm should be represented as a material conditional.

As Ferrer Beltran and Ratti (2012a, 36)) rightly observe, only their third case of defeasibility is really significant: the first two cases depend on meta-norms on

validity or application that are defeasible in the third sense, namely, according to their content. However, the way in which they describe defeasibility by content differs from the approach here adopted in three regards.

Firstly, it makes the implicitness of the exceptions to a norm a necessary condition for the defeasibility of the norm. On the contrary, I have argued that even explicit exceptions to a norm presuppose the defeasibility of that norm, since they give rise to the pattern that characterises defeasible argumentation: the absence of the exception is not needed to construct an argument warranted by the norm, though that argument can be defeated by arguments warranted by the exception.

Secondly, the characterisation of the antecedent of a defeasible norm as providing contributory conditions for the norm's conclusion fails to capture the idea of defeasible connection between the antecedent and the conclusion of that norm. As I observed in Section 3, a contributory condition for a conclusion may fail to provide any presumptive support for that conclusion. This is the case when a norm has a conjunctive antecedent so that its application results in a linked argument (see Figure 2). On the other hand, the antecedent of defeasible norm can be described as a contributory reason for the norm's conclusion. A genuine contributory reason for a legal conclusion should indeed provide on its own sufficient presumptive support to that conclusion, i.e., it should match the antecedent of a legal default, and so enable the construction of a separate defeasible argument. Separate defeasible argument sharing the same conclusion may contribute to a stronger convergent argument supporting the same conclusion (see Figure 3 and Figure 4).

Finally. it is not clear to me why the issue of determining where a norm is defeasible or not should be specifically addressed as a "matter of interpretation". It is a matter that pertains to the determination of the logical structure of the norm at issue, an issue that could pertain to interpretation or not depending on how one understand the notion of interpretation, i.e., as concerning every ascription of meaning to a text, or only the ascription of meaning meant to address some doubts (see Dascal and Wróblewski 1988). Interpretation in the first sense obviously covers the determination of every aspect of the content of a norm having a textual source, and therefore it also covers the determination of the norm's logical structure. In fact, to determine whether a norm is defeasible or not, we have to consider —depending on whether we are approaching the norm from a socio-legal or from doctrinal perspective— (a) the way in which the norm is comprehended and used by the community of its users (those who endorse/follow/apply it) or (b) the way in which the norm should correctly be comprehended and used by the same community. This determination would involve an empirical assessment according to (a) or a normative assessment according to (b). This assessment would be no less (and no more) dependent on interpretation than the determination of any other aspects of the norm's content, such as the structure and the components of its antecedent and its consequent. Finally, the issue of determining whether a legal norm is defeasible would be utterly trivial if we were to adopt —both empirically and normatively— the view that all legal norms are defeasible in this abstract sense —i.e., the assumption that no strict legal norm exists, as a matter of fact— following Leibniz's suggestion.

It seems to me that we need to distinguish clearly two aspect concerning a norm's defeasibility. The first aspect, to which we have referred in this contribution by using the term "defeasible", pertains the intrinsic logical structure of a

norm: is the norm meant to establish a presumptive or a conclusive link between its antecedent and its conclusion? Defeasibility so understood is a counterfactual property: to say that a norm *N,* having antecedent P and conclusion Q (I leave the variables implicit, for simplicity's sake), is defeasible just means that it is in principle possible to reject an argument for Q warranted by N , while accepting both N and P : we can imagine a system $\mathbb{L}$ and a factual constellation $\mathbb{F}$, such that with regard to the argumentation basis $\mathbb{L} \cup \mathbb{F}$ both N (unchanged) and P are justified (e.g. being unchallenged), but the argument that delivers Q on the basis of N and P is rebutted or undercut, by arguments constructible from $\mathbb{L} \cup \mathbb{F}$.

Once we have determined that N is intrinsically defeasible, we can address further issues. One issue concerns determining whether N- warranted arguments can be defeated in the legal system $\mathbb{L}$ containing N , i.e., whether a rebutting or undercutting counterargument to an N-warranted argument can be mounted by using only norms in $\mathbb{L}$, plus appropriate operative facts (see Section 11). A different issue concerns determining whether N could be defeated given the possible (permitted or empowered) judicial modifications of the current legal system L, namely, whether judicial construction/interpretation could introduce in $\mathbb{L}$ new norms that enable the construction of defeaters against the application of N . Obviously answering either of these issues may or will require the interpretation of the legal system $\mathbb{L}$ under consideration (on the connection between the possibility that a norm is defeated and interpretation, see Duarte 2011,135).

Finally, the idea of a norm's defeasibility as pertaining to the logical structure of that norm, leads me to address one further claim by Ferrer and Ratti, namely, the view that whenever a norm's validity or application is determined by defeasible metarules, the norm itself must be indefeasible. Since a norm's defeasibility only concerns the logical structure of that norm, the fact that a norm is defeasible does not exclude (nor require) that its validity as well as the domain of its intended application are governed by defeasible criteria. Inapplicability rules, however, may presuppose the defeasibility of the norm that they address: a rule stating that norm N is inapplicable under exceptional circumstances E , is usually meant enable the construction of undercutters to N- warranted arguments, namely, arguments having the following form: E is the case, if E than N does not apply (does not warrant its conclusion), therefore N does not apply: $E, E \Rightarrow \neg N$, therefore $\neg N$ (see argument C in Figure 8).

In conclusion, I think that, notwithstanding the multifarious creative ways in which legal theorists have framed the idea of defeasibility, it would better to stick to the more limited and precise concept on which other disciplines —such as logic, philosophy, and computing— converge, namely, the view that defeasible reasoning is nonmonotonic and that the antecedent of a defeasible norm provides only presumptive support to the norm's conclusion. The considerations that have been presented as alternative analyses of the concept of defeasibility should rather be rephrased are pertaining to the ways in which (a) arguments applying defeasible legal norms can be rebutted or undercut, or (b) existing legal norms, both defeasible or indefeasible ones, can be abrogated or modified.

2.9. Conclusion

Defeasible reasoning is a key aspect of legal reasoning and problem-solving. Therefore, theories and logics of defeasible can greatly contribute to the study of legal argumentation and legal justification.

Recognising the strength of the connection between defeasibility and the law does not require abandoning logical rigour. On the contrary, it favours adopting logical models that precisely match certain important structures of legal knowledge, certain frequent patterns of legal reasoning, and of the dialectics of legal interaction. Argument-based theories of defeasible reasoning provide the most advantageous approach to address defeasibility in legal contexts.

I have argued that legal theory should addressed defeasibility using a shared conceptual framework and focus with the other disciplines — in particular, logic and computing— which have so far addressing defeasible reasoning. This does not exclude that the legal theory can provide useful contribution to the study of defeasibility. In fact, the law provides a rich set of structures and patterns for defeasible reasoning. Therefore, the analysis of patterns of defeasibility in the law can contribute not only to legal theory and (computable) legal logic, but also to the development of general theories and logical models of defeasibility.

CHAPTER 3

Presumptions

I have claimed that defeasible reasoning consists in presumptive inference. More exactly, in a defeasibly valid argument , the premises only provide presumptive support for the conclusion: if we accept the premises we should also accept the conclusion, but only so long as we do not have prevailing arguments to the contrary. As example of such presumptive/defeasible inferences, we have considered reasoning patterns consisting in the application legal norms or of common sense warrants.

In this general sense, defeasibility and presumptivity can be seen as german concepts, or rather the two terms can be viewed as synonymous.

However, there is a distinct and more precise notion of presumptiveness in the law. Under this distinct notion of presumption, only some legal rules quality of presumption rule and only some inferences can be properly be viewed as presumption-based. In the following I will provide a model of presumption in the law

3.1. Presumptiveness as defeasibility

3.2. Presumptions in the law

CHAPTER 4

Burdens of Proof

CHAPTER 5

Defeasibility in Legal Interpretation

CHAPTER 6

Balancing and proportionality

CHAPTER 7

Defeasibility in Evidential reasoning

APPENDIX A

Classical logic

APPENDIX B

Logics for defeasible reasoning

Bibliography

Alchourrón, C. E. (1996a). Detachment and defeasibility in deontic logic. *Studia Logica 57*, 5–18.

Alchourrón, C. E. (1996b). On law and logic. *Ratio Juris 9*, 331–48.

Alchourrón, C. E. and E. Bulygin (1971). *Normative Systems*. Springer.

Alchourrón, C. E., P. Gärdenfors, and D. Makinson (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic 50*, 510–30.

Alchourrón, C. E. and D. Makinson (1981). Hierarchies of regulations and their logic. In R. Hilpinen (Ed.), *New Studies on Deontic Logic*, pp. 123–48. Reidel.

Alexy, R. (2002). *A Theory of Constitutional Rights*. Oxford University Press.

Aquinas, T. (1947). *Summa Theologica*. Benzinger Brothers.

Baroni, P., M. Caminada, and M. Giacomin (2011). An introduction to argumentation semantics. *The Knowledge Engineering Review 26*, 365–410.

Bench-Capon, T. J. M. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation 13*, 429–448.

Bench-Capon, T. J. M. and H. Prakken (2006). Justifying actions by accruing arguments. In P. E. Dunne and T. J. M. Bench-Capon (Eds.), *Computational Models of Argument. Proceedings of COMMA 2006*, pp. 247–258. IOS Press.

Bench-Capon, T. J. M. and G. Sartor (2003). A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence 150*, 97–142.

Blair, J. A. (2012). *Groundwork in the Theory of Argumentation*. Springer.

Brewer, S. (1996). Exemplary reasoning: Semantics, pragmatics and the rational force of legal argument by analogy. *Harvard Law Review 109*, 923–1028.

Brewer, S. (2011). Logocratic method and the analysis of arguments in evidence. *Law, Probability and Risk 10*, 175–202.

Brozek, B. (2004). *Defeasibility of Legal Reasoning*. Zakamycze.

Brozek, B. (2014). Law and defeasibility: a few comments on the logic of legal requirements. *Revus 23*, 165–170.

Celano, B. (2012). True exceptions: Defeasibility and particularism. In J. Ferrer Beltran and G. B. Ratti (Eds.), *The Logic of Legal Requirements*, pp. 268–87. Oxford University Press.

Chisholm, R. M. (1957). *Perceiving: A Philosophical Study*. Cornell University.

Clark, K. L. (1978). Negation as failure. In H. Gallaire and J. Minker (Eds.), *Logic and Data Bases*, pp. 293–332. Plenum. (1st ed. 1978.).

Dancy, J. (2004). *Ethics Without Principles*. Oxford University Press.

Dascal, M. and J. Wróblewski (1988). Understanding and interpretation in pragmatics and in law. *Law and Philosophy 7*, 203–24.

Duarte, D. (2011). Linguistic objectivity in norm sentences: Alternatives in literal meaning. *Ratio Juris 24*, 112–39.

Duarte d'Almeida, L. (2013). A proof-based account of legal exceptions. *Oxford Journal of Legal Studies 33*, 133–68.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*–person games. *Artificial Intelligence 77*, 321–57.

Fenton, N., M. Neil, and D. Berger (2016). Bayes and the law. *Annual Review of Statistics and Its Application 3*, 51–77.

Ferrer Beltran, J. and G. B. Ratti (2012a). Defeasibility and legality: A survey. In J. Ferrer Beltran and G. B. Ratti (Eds.), *The Logic of Legal Requirements: Essays on Defeasibility*, pp. 11–38. Oxford University Press.

Ferrer Beltran, J. and G. B. Ratti (Eds.) (2012b). *The Logic of Legal Requirements: Essays on Defeasibility*. Oxford University Press.

Gärdenfors, P. (1987). *Knowledge in Flux*. MIT.

Gazzo Castañeda, L. E. and M. Knauff (2016). Defeasible reasoning with legal conditionals. *Memory and Cognition 44*, 499–517.

Ginzberg, M. L. (Ed.) (1987). *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann.

Gordon, T. F. (1988). The importance of nonmonotonicity for legal reasoning. In H. Fiedler, F. Haft, and R. Traunmüller (Eds.), *Expert Systems in Law: Impacts on Legal Theory and Computer Law*, pp. 111–26. Attempto.

Gordon, T. F. (1995). *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Kluwer.

Gordon, T. F., H. Prakken, and D. N. Walton (2007). The Carneades model of argument and burden of proof. *Artificial Intelligence 171*, 875–96.

Governatori, G., M. J. Maher, D. Billington, and G. Antoniou (2004). Argumentation semantics for defeasible logics. *Journal of Logic and Computation 14*, 675–702.

Governatori, G. and A. Rotolo (2010). Changing legal systems: legal abrogations and annulments in defeasible logic. *Logic Journal of IGPL 18*, 157–94.

Governatori, G., A. Rotolo, and G. Sartor (2005). Temporalised normative positions in defeasible logic. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law (ICAIL 2005)*, pp. 25–34. ACM.

Guastini, R. (2012). Defeasibility, axiological gaps, and interpretation. In J. Ferrer Beltran and G. B. Ratti (Eds.), *The Logic of Legal Requirements*, pp. 182–92. Oxford University Press.

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

Hage, J. C. and A. Peczenik (2000). Law, morals and defeasibility. *Ratio Juris 13*, 305–25.

Hart, H. L. A. (1951). The ascription of responsibility and rights. In A. Flew (Ed.), *Logic and Language*, pp. 145–66. Blackwell. (1st ed. 1948–1949.).

Hitchcock, D. (2017). *On Reasoning and Argument: Essays in Informal Logic and on Critical Thinking*. Springer.

Holland, J. (2012). *Signals and Boundaries Building Blocks for Complex Adaptive Systems*. MIT.

Holland, J., K. J. Holyoak, R. E. Nisbett, and P. R. Thagard (1989). *Induction. Processes of Inference, Learning and Discovery*. MIT.

Holyoak, K. and P. Thagard (1996). *Mental Leaps: Analogy in Creative Thought.* MIT.

Horty, J. (2001). Nonmonotonic logic. In L. Goble (Ed.), *The Blackwell Guide to Philosophical Logic*, pp. 336–61. Blackwell.

Horty, J. F. (2007). Reasons as defaults. *Philosopher's Impring 8*(3), 1–28.

Horty, J. F. (2011). Rules and reasons in the theory of precedent. *Legal theory 10*, 1–33.

Horty, J. F. (2012). *Reasons as Defaults.* Oxford University Press.

Hunter, A. (2013). A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning 54*, 47–81.

Kant, I. ([1797] 1949). On a supposed rights to lie from altruistic motives. In L. White Beck (Ed.), *Critique of Practical Reason and Other Writings in Moral Philosophy,*, pp. 346–50. Chicago University Press.

Koons, R. (2009). Defeasible reasoning. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.*

Kraus, S., D. Lehmann, and M. Magidor (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence 44*, 167–207.

Leibniz, G. W. (1923). *Sämtliche Briefe und Werke.* Akademie-Verlag.

Loui, R. P. and J. Norman (1995). Rationales and argument moves. *Artificial Intelligence and Law 3*, 159–89.

MacCormick, D. N. (1995). Defeasibility in law and logic. In Z. Bankowski, I. White, and U. Hahn (Eds.), *Informatics and the Foundations of Legal Reasoning*, pp. 99–117. Kluwer Academic.

Maranhão, J. S. A. (2013). Defeasibility, contributory conditionals, and refinement of legal systems. In J. Ferrer Beltran and G. B. Ratti (Eds.), *The Logic of Legal Requirements*, pp. 53–76. Oxford University Press.

McCarthy, J. (1980). Circumscription – a form of non-monotonic reasoning. *Artificial Intelligence 13*, 27–39.

Modgil, S. and H. Prakken (2013). A general account of argumentation with preferences. *Artificial Intelligence 195*, 361–97.

Nute, D. (1994). Defeasible logic. In *Handbook of logic in artificial intelligence and logic programming. Volume 3: Nonmonotonic reasoning and uncertain reasoning*, pp. 353–395. Oxford University Press.

Peczenik, A. (2005). *Scientia Juris: Treatise of Legal Philosophy and General Jurisprudence - Volume 4.* Springer.

Perelman, C. and L. Olbrechts-Tyteca (1969). *The New Rhetoric: A Treatise on Argumentation.* University of Notre Dame Press. (1st ed. in French 1958.).

Pollock, J. L. (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person.* MIT.

Pollock, J. L. (1998). Perceiving and reasoning about a changing world. *Computational Intelligence 14*, 498–562.

Pollock, J. L. (2008). Defeasible reasoning. In J. E. Adler and L. J. Rips (Eds.), *Reasoning: Studies of Human Inference and its Foundations*, pp. 451–70. Cambridge University Press.

Pollock, J. L. (2010). Defeasible reasoning and degrees of justification. *Argument and Computation 1*, 7–22.

Prakken, H. (2005). A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial*

Intelligence and Law (ICAIL 2005), pp. 85–94. ACM.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation 1*, 93–124.

Prakken, H. and G. Sartor (1996). Rules about rules: Assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law 4*, 331–68.

Prakken, H. and G. Sartor (1997). Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics 7*, 25–75.

Prakken, H. and G. Sartor (1998). Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law 6*, 231–87.

Prakken, H. and G. Sartor (2009). A logical analysis of burdens of proof. In H. Kaptein, H. Prakken, and B. Verheij (Eds.), *Legal Evidence and Proof: Statistics, Stories, Logic*, pp. 223–53. Ashgate.

Prakken, H. and G. Sartor (2015). Law and logic: A review from an argumentation perspective. *Artificial Intelligence 227*, 214–45.

Prakken, H. and G. A. W. Vreeswijk (2001). Logics for defeasible argumentation. In *Handbook of philosophical logic*, pp. 219–318. Springer.

Rahwan, I. and G. R. Simari (2009). *Argumentation in Artificial Intelligence*. Springer.

Raz, J. (1985). Authority, law, and morality. *The Monist 68*, 295–323.

Reiter, R. (1980). Logic for default reasoning. *Artificial Intelligence 13*, 81–132.

Rescher, N. (1977). *Dialectics: A Controversy-oriented Approach to the Theory of Knowledge*. State University of New York Press.

Rescher, N. (2006). *Presumption and the Practices of Tentative Cognition*. Cambridge University Press.

Riveret, R., A. Rotolo, and G. Sartor (2012). *Norms and Learning in Probabilistic Logic-Based Agents*, Volume Deontic Logic in Computer Science. 11th International Conference, DEON 2012, Bergen, Norway, July 16-18, 2012. Proceedings. Springer.

Rodriguez, J. (2012). Against defeasibility of legal rules. In J. Ferrer Beltran and G. B. Ratti (Eds.), *The Logic of Legal Requirements*, pp. 89–107. Oxford University Press.

Ross, W. D. ([1930]2002). *The Right and the Good*. Clarendon.

Ross, W. D. (1939). *Foundations of Ethics*. Clarendon.

Russell, S. J. and P. Norvig (2010). *Artificial Intelligence. A Modern Approach* (3 ed.). Prentice Hall.

Sartor, G. (1993). Defeasibility in legal reasoning. *Rechtstheorie 24*, 281–316.

Sartor, G. (1994). A formal model of legal argumentation. *Ratio Juris 7*, 212–26.

Sartor, G. (2005). *Legal Reasoning: A Cognitive Approach to the Law*. Springer.

Sartor, G. (2013). The logic of proportionality: Reasoning with non-numerical magnitudes. *German Law Journal 14*, 1419–57.

Schauer, F. F. (2012). Is defeasibility an essential property of law? In J. Ferrer Beltran and G. B. Ratti (Eds.), *The Logic of Legal Requirements*, pp. 77–88. Oxford University Press.

Sergot, M. J., F. Sadri, R. A. Kowalski, F. Kriwaczek, P. Hammond, and H. Cory (1986). The British Nationality Act as a logic program. *Communications of the ACM 29*, 370–86.

Stone Sweet, A. (2004). *The Judicial Construction of Europe*. Oxford University Press.

Toulmin, S. ([1958] 2003). *The Uses of Argument*. Cambridge University Press. (1st ed. 1958.).

Verheij, B., F. Bex, S. Timmer, C. Vlek, J.-J. Meyer, S. Renooij, and H. Prakken (2016). Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk 15*, 35–70.

Viehweg, T. (1965). *Topik und Jurisprudenz. Ein Beitrag zur rechtswissenschaflichen Grundlagenforschung* (3 ed.). Beck.

Walton, D. and G. Sartor (2013). Teleological justification of argumentation schemes. *Argumentation 27*, 111–142.

Walton, D., G. Sartor, and F. Macagno (2016). An argumentation framework for contested cases of statutory interpretation. *Artificial Intelligence and Law 24*, 51–91.

Walton, D. N. (1996). *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates.

Walton, D. N. (2006). *Fundamentals of Critical Argumentation*. Cambridge University Press.

Walton, D. N. (2008). A dialogical theory of presumption. *Artificial Intelligence and Law 16*, 209–43.

Walton, D. N. (2013). *Methods of Argumentation*. Cambridge University Press.

Walton, D. N., C. Reed, and F. Macagno (2008). *Argumentation Schemes*. Cambridge University Press.