



EUI Working Papers

LAW 2009/02

DEPARTMENT OF LAW

DEFEASIBILITY IN LEGAL REASONING

Giovanni Sartor

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF LAW

Defeasibility in Legal Reasoning

GIOVANNI SARTOR

EUI Working Paper **LAW** 2009/02

This text may be downloaded for personal research purposes only. Any additional reproduction for other purposes, whether in hard copy or electronically, requires the consent of the author(s), editor(s). If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper or other series, the year, and the publisher.

The author(s)/editor(s) should inform the Law Department of the EUI if the paper is to be published elsewhere, and should also assume responsibility for any consequent obligation(s).

ISSN 1725-6739

© 2009 Giovanni Sartor

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract

I shall first introduce the idea of reasoning, and of defeasible reasoning in particular. I shall then argue that cognitive agents need to engage in defeasible reasoning for coping with a complex and changing environment. Consequently, defeasibility is needed in practical reasoning, and in particular in legal reasoning.

Keywords

Legal reasoning – defeasibility

Contents

1 Reasoning Schemata and Reasoning Instances	1
2 The Adoption of a Reasoning Schema	3
3 Conclusive and Defeasible Reasoning	4
4 Validity and Truth-Preservation	5
5 Monotonic and Nonmonotonic Reasoning	5
6 The Rationale of Defeasibility	7
7 The Logical Function of Defeasible Reasoning Schemata	9
8 Collision and Defeat	10
9 Collisions and Incompatibility	11
10 Undercutting Collisions	12
11 Preference-Based Reasoning	13
12 Reinstatement	14
13 Undercutting in Practical Reasoning	16
14 Defeasible Reasoning and Probability	17
15 The Idea of Defeasibility in the Practical Domain	19
16 Defeasibility in Legal Language	20
17 Defeasibility in Legal Concepts and Procedures	21
18 Overcoming Legal Defeasibility?	22
19 Conclusion	24
Bibliography	26

Defeasibility in Legal Reasoning*

Giovanni Sartor
European University Institute, Florence
CIRSFID, University of Bologna

I shall first introduce the idea of reasoning, and of defeasible reasoning in particular. I shall then argue that cognitive agents need to engage in defeasible reasoning for coping with a complex and changing environment. Consequently, defeasibility is needed in practical reasoning, and in particular in legal reasoning.

1 Reasoning Schemata and Reasoning Instances

A reasoning agent proceeds in ratiocination through discrete reasoning steps. Each step is characterised by its pre-conditions (some mental states already possessed by the reasoner) and its post-conditions (some new mental states, to be acquired through performing the reasoning step).¹ Correct ratiocination consists of rational transitions from pre-conditions into post-conditions, namely transitions take place according to certain patterns, or *reasoning schemata*, which constitute, and may be validated by, rationality.² In general, I shall represent such schemata in the following form:

Reasoning schema: *Name*

$$\frac{A_1; \dots; \text{AND } A_n}{B_1; \dots; \text{AND } B_m} \text{ IS A REASON FOR}$$

where *Name* is the name of the schema, A_1, \dots, A_n are the pre-conditions of the schema (cognitive states the reasoner possesses before instantiating the schema), and B_1, \dots, B_m are its post-conditions (cognitive states the agent possesses after instantiating the schema). I will say that the set of all the pre-conditions in a schema constitutes its *reason*, and the set of all its post-conditions forms its *conclusion*. I will also speak respectively of a *subreason* or of a *subconclusion* to refer to a single pre-condition or to a single postcondition, that is, to refer to a single mental state contained in the reason or in the conclusion.³

*Supported by the EU projects ONE-LEX (Marie Curie Chair), ESTRELLA (IST-2004-027665), and ALIS (IST-2004-027968).

¹Here I am considering the sequential component of reasoning, namely, ratiocination, rather than the so-called heuristic component. The heuristic component works in different ways (in making analogies, suggesting hypotheses, etc.) and provides its outcomes to the sequential component. The latter, while being unable to achieve such outcomes on its own, may subject them to its scrutiny. For some observations on this distinction, and on its connections with the usual dycotomy context of discovery vs context of justification, see G. Sartor, *Legal Reasoning: A Cognitive Approach to the Law*, vol. 5, Treatise on Legal Philosophy and General Jurisprudence (Berlin, Springer, 2005), Ch. 2.

²This idea is developed by the epistemologist John Pollock, whose theory of reasoning provides the basic inspiration for the model here presented. See J. L. Pollock, *Cognitive Carpentry: A Blueprint for How to Build a Person* (New York, N. Y., MIT Press, 1995), J. L. Pollock and J. Cruz, *Contemporary Theories of Knowledge* (Totowa, N. Y., Rowman and Littlefield, 1999).

³This understanding of the term “reason” corresponds to a common way of using it (my reason for believing *A* is that I believe that *B*, my reason for intending to do *A* is that I desire that *B* and I believe that by doing this I will achieve *C*). However, it does not correspond to the way in which this term is often used in legal theory, namely as a fact justifying a certain action (J. Raz, *Practical Reason and Norms* [London, Hutchinson, 1975]). In the sense in which I am using the term “reason”, a reason is not an external fact, but a cognitive state of the agent, and a reason does not justify an action, but it justifies the adoption of a further mental state, whose adoption is justified according to the procedure of rationality (as long as the agent instantiates the cognitive state constituting the reason).

Reasoning schemata are *formal*, in the sense that they apply to all contents having a certain structure or *logical form*, though natural language can express such contents in different ways.⁴ Any specific instance of a reasoning schema—any combination of mental states matching the reasoning schema—constitutes a *reasoning instance*. For example, consider the following reasoning schema, named *sylogism*:⁵

Reasoning schema: *Sylogism*

- (1) believing that all *P*'s are *Q*'s; AND
 (2) believing that *a* is *P*
 ————— IS A REASON FOR
 (3) believing that *a* is *Q*

Note that the reason of this schema includes two components or *subreasons*:

- the belief in a general proposition, which is usually called the *major premise* of the *sylogism*, and
- the belief in a particular (individual or concrete) proposition, which is usually called the *minor premise* of the *sylogism*.

The major and the minor premise are connected by the fact that a predicate (*P*) occurs in both of them. The reasoning schema *sylogism* is instantiated by following reasoning instance, which embodies its pattern:

Reasoning instance: *Sylogism*

- (1) believing that all Mondays are days on which there is a flight to Barcelona; AND
 (2) believing that today is Monday
 ————— IS A REASON FOR
 (3) believing that today is a day on which there is a flight to Barcelona

In the following I am only considering cognitive states represented by beliefs, rather than considering also non-doxastic cognitive states, such as perceptions in the epistemic domain, and preference, desires, intentions or wants in the practical domain. However, among the beliefs I also include practical beliefs, such the belief that something is preferable to something else, that something is a value, or that it is obligatory or permitted to perform a certain actions.⁶ Therefore, I can omit, in the representation of a reasoning schemata, the indication that such schemata concern beliefs as cognitive states of a reasoning agent—rather than the propositions providing the contents of such beliefs—and just indicate the believed propositions. Consequently, the above instance of the *sylogism* schema can be represented as follows (omitting, for semplicity, also the words “is a reason for” and the name of the instantiated schema).

- (1) all Mondays are days on which there is a flight to Barcelona;
 (2) today is Monday
 —————
 (3) today is a day on which there is a flight to Barcelona

Here are two legal examples of normative syllogisms. The first corresponds to the so-called *judicial syllogism*:

⁴One may even wonder whether there is a universal *language of thought* in which each cognitive content may be uniquely expressed (J. Fodor, *The Modularity of Mind* [Cambridge, Mass., MIT Press, 1983]).

⁵I use the expression *sylogism*, since this pattern of reasoning corresponds, in the legal domain, to what is usually called *judicial syllogism* (J. Wróblewski, “Legal Syllogism and Rationality of Judicial Decision” [1974] *5 Rechtsstheorie*, 33–45). However, I need to remind the reader that the Aristotelian theory of syllogism was not concerned with propositions referring to specific individuals, like proposition (2) and (3) in our schema. The syllogistic figure which comes closest to our *sylogism* would be the Barbara mood, according to which the couple of (universal) premises (1) [all *P*'s are *Q*'s] and (2) [all *Q*'s are *R*'s] leads to the (universal) conclusion that (3) [all *P*'s are *R*'s].

⁶I will not consider here how practical beliefs connect to non-doxastic conative states (see Sartor, *Legal Reasoning: A Cognitive Approach to the Law*).

Reasoning instance: *Syllogism*

- (1) all thieves ought to be punished; AND
- (2) John is a thief

- (3) John ought to be punished

The second concerns the type of reasoning which is involved in referring to authoritative sources of law, which can be called *meta-syllogism*.

Reasoning instance: *Meta-syllogism*

- (1) all rules issued by the head of the law school are binding; AND
- (2) rule [it is forbidden to smoke in the premises of the law school] was issued by the head of the law school

- (3) rule [it is forbidden to smoke in the premises of the law school] is binding

While in ordinary syllogisms a rule is used for deriving normative qualifications of people or objects, in meta-syllogism meta-rule (a rule about rules) is used for inferring properties of rules (more generally, of normative propositions) or relations between rules.

2 The Adoption of a Reasoning Schema

Let us specify what it means for a reasoner to *adopt a reasoning schema*. When saying that a reasoner j adopts a reasoning schema, I mean that j has a particular inclination, corresponding to, and validated by, rationality: Whenever j instantiates all preconditions of the schema, then j will also tend to instantiate the schema's post-conditions. For instance, when saying that j adopts schema *detachment*, I mean that j has the following inclination: Whenever j believes both A and [if A then B], j will also tend to believe B .

I need however to clarify what I mean by j having such an inclination. One cannot perform all inferences required by every rational reasoning schema: This would lead one to acquire an infinite number of useless cognitive states, and therefore to fill one's head with useless contents. The simplest inference rules of propositional logic are sufficient to lead an overzealous reasoner into such a hopeless condition. Consider, for example, the following schema:

Reasoning schema: *Disjunction introduction*

- (1) A

- (2) A OR B

The inferences that are enabled by schema *disjunction introduction* look obvious: Any proposition entails its disjunction. For instance, a reasoner believing that [today is Tuesday] can safely come to believe that [today is Tuesday OR today is Thursday]. This process unfortunately may continue: As from proposition A one infers proposition [A OR B], from the latter proposition one can infer [A OR B OR C], and so on infinitely.

This issue is linked to the so-called problem of *logical omniscience*: One cannot derive (and endorse) all implications of one's beliefs. The word "cannot" in the previous sentence, however, can be read in two different senses. In a first sense, it points to a serious limitation of our cognitive powers: We cannot (are unable to) infer many important truths that follow from what we already know (progressing in the discovery of these truths is the difficult task of mathematicians and logicians). In a second sense, which is the one I am now considering, the assertion that we cannot be logically omniscient rather refers to the futility of a misguided cognitive effort: We cannot (we should not, since it would be silly or unreasonable) derive all useless or trivial implications of our current beliefs.

The way out of the latter problem consists in limiting oneself to performing only those inferences that may be relevant for one's epistemic interests, according to the priorities determined by these interests (and

according to the available time and energy). Therefore we cannot view reason-schemata neither as absolute necessities, forcing one to draw whatever irrelevant conclusions they indicate, nor as pure possibilities, which one can randomly implement or disregard. Reasoning schemata rather express a *necessity* that is *conditioned* to the utility or relevance of their use. Thus, one believing both P and $[\text{if } P \text{ then } Q]$ should acquire belief in Q only if one has some interest in establishing whether Q holds (and one has nothing more important to do). Otherwise a rational reasoner should refrain from making the inference.⁷

Finally, I am presenting only reasoning schemata which are “rational”, in the sense that they pertain to rationality (they indicate ways in which rational cognition proceeds), and are validated by rationality itself, when reflexively applied to the evaluation of its own procedures. Therefore the fact that an agent endorses a reason (the precondition of a rational reasoning schemata) not only leads that agent, as a matter of fact, to endorse the conclusion of the schema, but it also justifies, and guides it to the endorsement of that conclusion: the fact that I currently believe both “ P ” and “ $\text{IF } P \text{ THEN } Q$ ”, not only leads me, as a matter of fact, to believe “ Q ”, but also justifies me in accepting “ Q ”, and this justification appears convincing to my rationality itself (as long as I continue to believe both premises, and, in the case of defeasible reasoning, I have no prevailing reason countering such an inference). This justification does not exclude that the belief in “ Q ” is false (incorrect). This may happen when the premises (the reason) are wrong: the conclusions I derive from false beliefs, according to correct inferences, are likely to be also false. However this fact does not make the derivation of such conclusions unjustified, as long as one maintains belief their premises (to avoid deriving false conclusion from false premises, one should retract such premises, rather than refraining from deriving the conclusions they entail). So, the belief on a conclusion may be unjustified, since a mistake was performed in one step of the chain leading the reasoner to that conclusion, but still the reasoner would be justified (on the basis of its previous beliefs, and until he withdraws such beliefs) in performing those inference steps in that chain which are correct taken in isolation.

3 Conclusive and Defeasible Reasoning

Two classes of reasoning schemata must be distinguished: conclusive and defeasible ones. The basic difference between the two kinds can be summarised as follows.

A *conclusive* reasoning schema indicates a cognitive transition that can operate regardless of any further information the agent possesses, as long as the reasoner instantiates the pre-conditions in the schema: Whenever one endorses the reason of the schema, one may safely endorse its conclusion.

Definition 3.1 *Conclusive reasoning schema.* A reasoning schema R is conclusive if one can always adopt R 's conclusions while endorsing R 's premises (and one should never reject R 's conclusions while endorsing R 's premises).

The distinctive feature of a *defeasible* reasoning schema is that it may be defeated by further information to the contrary: The schema indicates a transition that only operates when one has no prevailing beliefs (or, more generally, mental states) against applying the schema or against adopting its conclusions. When one endorses the premises of a defeasible schema but has such prevailing beliefs to the contrary, one should refrain from adopting the conclusion of the schema (and withdraw that conclusion if one has already adopted in by instantiating that schema).

Definition 3.2 *Defeasible reasoning schema.* A reasoning schema is defeasible if one should, under certain conditions, refrain from adopting its conclusions though endorsing its premises.

In the following sections I shall consider some important aspects of conclusive and defeasible reasoning, analysing their commonalities and differences.

⁷On the connection between reasoning and interest, for an architecture for interest-driven reasoning, see Pollock, *Cognitive Carpentry: A Blueprint for How to Build a Person*.

4 Validity and Truth-Preservation

The strong connection between reason and conclusion that characterises conclusive reasoning schemata can be linked to the idea of *truth-preservation*. Conclusive inference schemata are truth preserving: Whenever (in any possible situation where) the premises of a conclusive schema are true, then also the conclusions of the schema are true. Said otherwise, it is impossible that the premises of a conclusive schema are true and its conclusions are false.

Definition 4.1 *Truth preservation.* A reasoning instance R is truth preserving if necessarily, whenever R 's premises are true, then also R 's conclusions are true. Similarly, a reasoning schema R is truth-preserving if all R 's instances are truth preserving.

Truth-preservation is a very useful and important property, but it does not characterise all rational reasoning patterns. Therefore, it does not delimit the scope of logical reasoning, if by *logic* I mean rational reasoning, or rational ratiocination.

It is true that many authors tend to limit logic to truth-preserving reasoning: They view logic as having the specific function of providing truth-preserving ways of reasoning. Moreover, many logicians refer to truth-preservation by using the word *valid*. For instance, it is said “an argument is VALID if it is logically impossible for all the premises to be true, yet the conclusion false”⁸, or, just to take another citation, that “an argument is called *valid* when its premises entail its conclusion, in other words, if the premises can't be true without the conclusion also being true”.⁹ Correspondingly, it is also said that logic is the study of valid reasoning.

Obviously, there is nothing wrong in using the word *valid* in this specific sense (for which there is a long and very respectable tradition), nor in defining the word *logic* so that it only concerns truth-preserving reasoning. However, these definitions lead people (especially when they are not familiar with formal reasoning and with the technical meaning of logical notions) to the idea that any form of reasoning that is not truth-preserving is “invalid,” in the generic sense of being wrong, arbitrary, or unreasonable.

To avoid this connotation sneaking into our discourse (and to avoid confusion with the sense in which the word *valid* is used in the law, for example when discussing legal sources), I shall refrain from using *valid* in the sense of “truth preserving” (and *invalid* in the sense of “non truth-preserving”). This allows us to avoid qualifying all defeasible inferences as being “invalid”: Though they are not truth-preserving (it may happen that their preconditions hold, but their conclusions fail to be true), defeasible reasoning schemata are indeed “valid” or “sound” forms of reasoning, in the sense of being appropriate ways of approaching certain cognitive tasks.

5 Monotonic and Nonmonotonic Reasoning

Conclusive reasoning schemata provide for *monotonic reasoning*: The conclusions that can be derived by a reasoner that only uses conclusive reasoning always grow, as long as the reasoner is provided with further input information. More exactly, if a conclusion A can be conclusively derived from a set of premises S_1 , then A can also be derived from whatever set S_2 resulting from the addition of further premises to S_1 (from whatever set S_2 such that $S_1 \subset S_2$).

On the contrary, when one draws defeasible inferences from a set of premises S_1 , it may happen that, by adding further premises to S_1 , one obtains a set S_2 which does not entail some conclusions one could derive from S_1 alone. Defeasible reasoning schemata license *non-monotonic* reasoning: Their conclusions may need to be abandoned when new information is available.¹⁰

⁸M. Sainsbury, *Logical Forms: An Introduction to Philosophical Logic* (Oxford, Blackwell, 2001).

⁹W. Hodges, “Elementary Predicate Logic”, in *Handbook of Philosophical Logic. Volume I: Elements of classical logic*, ed. by D. Gabbay and F. Günthner (Dordrecht, Kluwer, 1983), 1–131, 1.

¹⁰On non-monotonic reasoning, see M. L. Ginzberg, ed., *Readings in Nonmonotonic Reasoning* (Los Altos, Cal., Morgan Kaufmann, 1987), which collects the contributions which have originated research in this domain. For an

Reasoning schema: *Conclusive syllogism*

- (1) all Y 's are Z 's; AND
 (2) x is Y
 ————— IS A CONCLUSIVE REASON FOR
 (3) x is Z

Reasoning schema: *Defeasible syllogism*

- (1) all Y 's are normally Z 's; AND
 (2) x is Y
 ————— IS A DEFEASIBLE REASON FOR
 (3) x is Z

Table 1: *Conclusive and defeasible syllogism: reasoning schemata*

Reasoning instance: *Conclusive syllogism*

- (1) all bachelors are unmarried; AND
 (2) John is a bachelor
 ————— IS A CONCLUSIVE REASON FOR
 (3) John is unmarried

Reasoning instance: *Defeasible syllogism*

- (1) pet dogs are normally unaggressive; AND
 (2) Fido is a pet dog
 ————— IS A DEFEASIBLE REASON FOR
 (3) Fido is unaggressive

Table 2: *Conclusive and defeasible syllogism: reasoning instances*

In Table 1 you can see the reasoning schemata for defeasible and conclusive syllogism, which are applied in the reasoning instances of Table 2. According to the first reasoning instance, when one believes that [all bachelors are unmarried] and that [John is a bachelor], one can conclusively conclude that [John is unmarried]. According to the second instance, when one believes that [Pet dogs are normally unaggressive] and that [Fido is a pet dog] one can defeasibly conclude that [Fido is unaggressive].

The difference between conclusive and defeasible reasoning emerges most clearly when one acquires beliefs that contradict the conclusions of one's previous inferences.

Let us assume, for example, that Mary—who knows that [bachelors are unmarried] and [husbands are married]—after meeting John at a dinner party comes to believe, according to John's statements, that [John is a bachelor]. This leads Mary to believe, according to schema *conclusive syllogism*, that [John is unmarried]. However, the day after a friend tells Mary that [John is Lisa's husband]. This should lead her to conclude, still according to *conclusive syllogism*, that [John is married], which contradicts her belief that [John is unmarried]. At this stage, Mary has no choice but to abandon the premises of one of these inferences. If she sticks to the idea that John is a bachelor, she needs to withdraw the belief that John is Lisa's husband, while if she accepts the idea that John is Lisa's husband, she needs to withdraw the belief that John is a bachelor.

In defeasible reasoning, a different approach is required. Let us assume, for instance, that one endorses all of the following propositions: [Fido is a pet dog], [pet dogs are normally unaggressive], [Fido is a Doberman], [Dobermans are normally aggressive]. According to schema *defeasible syllogism*, this set of

introduction, see also G. Brewka, *Nonmonotonic Reasoning: Logical Foundations of Commonsense* (Cambridge, Cambridge University Press, 1991).

(1) Fido is a pet dog;	(2) pet dogs are normally unaggressive
(3) Fido is unaggressive	
(1) Fido is a Doberman;	(2) Dobermans are normally aggressive
(3) Fido is aggressive	

Table 3: *Two defeasible inferences*

premises licences both of the defeasible inferences in Table 3. Let us assume that the second inference in Table 3 is stronger (more reliable) than the first one. In such a situation, we are not required to withdraw any of the premises of the weaker inference (withdraw the belief that pet dogs are normally unaggressive, or that Fido is a pet dog). We can maintain the premises of both inferences, that is, we can keep on believing both of the following reasons (sets of premises): {Fido is a Doberman; Dobermans are normally aggressive}, {Fido is a pet dog; Pet dogs are normally not aggressive}. However, we shall refrain from deriving the conclusion that [Fido is aggressive].

6 The Rationale of Defeasibility

Defeasible reasoning schemata, as I have observed in Section 3, are not truth-preserving: When believing the premises of a defeasible schema, we are lead to endorse the conclusions of the schema, though these conclusions are not truth-preservingly implied by our premises (and they may indeed turn out to be false, even when the premises hold true). However, failure to satisfy truth preservation does not entail logical faultiness. On the contrary, epistemology has come to identify various kinds of sound defeasible inference.¹¹ Here I list a few of them:

- *Perceptual inference.* Having certain perceptual contents is a defeasible reason to believe in the existence of corresponding external objects. More generally, having a percept with content φ is a defeasible reason to believe φ . For instance, having an image of a red book at the centre of my visual field is a defeasible reason to believe that there is a red book in front of me. This conclusion is defeated if I come to know that there are circumstances which do not ensure the reliability of my perceptions (I am watching a hologram).
- *Memory inference.* Recalling φ is a defeasible reason to believe φ . For instance, my memory that yesterday I had a faculty meeting is a defeasible reason for me to believe that indeed there was such meeting. This inference is defeated if I come to know that the supposed memory was an outcome of my imagination.
- *Enumerative induction.* Observing a sample of F 's all of which are G 's is a defeasible reason to believe that all F 's are G 's. For instance, believing that all crows I have ever seen are black is a defeasible reason to believe that all crows are black. This inference is defeated if I perceive one white crow.
- *Statistical syllogism.* Believing that [most F 's are G 's AND a is an F] is a defeasible reason to believe that a is a G . For instance my beliefs that [most printed books have even-numbered pages on their left side] and that [the volume on the top of my table is a printed book] are defeasible reason for me to believe that this volume has even-numbered pages on its left side. This inference is defeated if I discover that this volume was wrongly printed with even-numbered pages on its right side.

¹¹Pollock, *Cognitive Carpentry: A Blueprint for How to Build a Person*52ff.

- *Temporal persistence.* Believing that φ is the case at time t_1 is a defeasible reason to believe that φ is still the case at a later time t_2 . For instance, my belief that my computer was on the top of my table yesterday evening (when I last saw it) is a defeasible reason for me to believe that my computer is still there. This inference is defeated if I come to know that the computer was moved from the table after yesterday evening.

Similar defeasible reasoning schemata may be identified also for practical, and in particular legal reasoning. For instance, we endorse the conclusion of a rule on the basis of our belief the rule is valid and that the rule's antecedent is satisfied. This inference, however, can be defeated either by a stronger rule to the contrary or by showing that the rule at issue is inapplicable. Another typical defeasible inference pattern characterises teleological reasoning, where we conclude for the adoption of plan of action (or the endorsement of a rule), given that it appears that such plan is able to achieve a certain value, and that its impact on all values at issue is better than the impact of any other plan we have been able to identify so far. This inference, however, is defeated if we discover a different plan which would give a better outcome.

In general, there is nothing strange or pathological in defeasible reasoning. On the contrary, defeasibility is the natural way in which an agent can cope with a complex and changing environment.

We do not even need to view defeasibility only in cognitive agents: An agent only endowed with fixed or conditioned reflexes may exhibit what may be viewed as a form, or at least as an evolutionary antecedent, of defeasible reasoning. Consider a reactive agent having two reflexes r_1 and r_2 such that:

- according to r_1 , stimulus s_1 (tasting good) triggers action a_1 (eat it!);
- according to r_2 , stimulus s_2 (feeling hot) triggers action a_2 (get rid!);
- r_2 is stronger than r_1 .

Assuming that the strength of each reflex is proportional to its importance for an agent's survival or reproduction, the most useful thing to do (and thus the solution that should have been chosen by natural evolution) when incompatible reflexes are triggered would be to implement the stronger of the two reflexes. Accordingly, when confronted with stimuli s_1 and s_2 (a tasty, but burning bite of food), a well-adapted reactive agent, rather than staying inactive or choosing randomly, will execute r_2 and do a_2 (get rid of the food). Therefore we may conclude that, in a certain sense, reflex r_2 defeats r_1 : Given only stimulus s_1 , the agent would react with a_1 , but given the combination of s_1 and s_2 , the agent will react with a_2 .

Though one may correctly speak of defeasible reflexes, defeasibility acquires its fullest meaning for cognitive agents: For such agents defeasibility consists in having certain cognitive states and withdrawing them when further cognitive results become available. This is what happens, as we have seen in the above pages, in *defeasible ratiocination*: The agent acquires through reasoning certain provisional cognitive states, and later may retract them, as a result of further reasoning.

To refer to the provisional conclusions of defeasible reasoning, usually the qualification *prima facie* is used. However, qualifying all defeasible conclusions as being *prima facie* suggests that all results of defeasible reasoning are obtained on the basis of the only information that is immediately available to the reasoner. This suggestion is misleading, since a defeasible conclusion may also be adopted after an accurate inquiry. Possibly a better terminological choice¹² consists in qualifying the outcomes of defeasible reasoning as *pro-tanto* conclusions, namely, conclusions which, through being justified on the basis of the information so far considered, may be withdrawn on the basis of further information.

Note that a defeasible belief is not necessarily so strong as to justify acting accordingly. One may be aware that further inquiry may provide reasons against maintaining that belief. In such a case, one may resist acting on the basis of a *pro-tanto* conclusion. It will depend on the circumstances of the case, and mainly on the depth of the inquiry so far performed and on the need of providing a quick answer, whether rationality requires jumping from *pro-tanto* acceptance into action, or rather deferring action until the issue has been better examined.¹³

¹²Suggested by A. Peczenik, "Scientia Juris", in *Treatise of Legal Philosophy and General Jurisprudence - Volume 2* (Berlin, Springer, 2006) Sec. 5.1.3.

¹³The important idea of *rational deferment* goes back to the mediaeval philosopher John Buridan, who was unjustly ridiculed with the famous story of Buridan's ass (J. Buridan, *Quaestiones super decem libros Ethicorum*

Consider for example the case of a person who asks his tax lawyer whether he should pay taxes on money he earned abroad. Assume that the lawyer finds a rule stating that also money earned abroad should be taxed. However, the lawyer is aware that a number of exemptions exist, concerning different countries and different types of revenue (though she is not aware of the content and the preconditions of such exceptions). Therefore she should tell the client that she only *pro-tanto* (namely, on the basis of the information she has so far considered) believes that the money he earned abroad is not taxed. She needs to look further into tax law for providing a sufficiently reliable answer.

7 The Logical Function of Defeasible Reasoning Schemata

According to the analysis I developed in the previous section, defeasible inference schemata seem to have a twofold function.

The first function is that of providing the reasoner with provisional thoughts, on the basis of which one may reason and if necessary act, until one has new information to the contrary. In this spirit, John Pollock¹⁴ relates defeasible reasoning to a general feature of human cognition. He argues that normally one starts with perceptual inputs and goes on inferring beliefs from one's current cognitive states (one's percepts plus the beliefs which one has previously inferred). Such a belief-formation process must satisfy apparently incompatible desiderata:

- One must be able to form beliefs on the basis of a partial perceptual input (one cannot wait until one has a complete representation of one's environment).
- One must be able to take into account an unlimited set of perceptual inputs.

Defeasibility is the way to reconcile such requirements:

The only obvious way to achieve these desiderata simultaneously is to enable the agent to adopt beliefs on the basis of small sets of perceptual inputs but then retract them in the face of additional perceptual inputs if those additional inputs conflict in various ways with the original basis for the beliefs. This is a description of *defeasible reasoning*. Beliefs are adopted on the basis of arguments that appeal to small sets of previously held information, but the beliefs can later be retracted in the face of new information.¹⁵

The second function for defeasible reasoning is that of activating a structured process of inquiry, based upon drawing *pro-tanto* conclusions, looking for their defeaters, for defeaters of defeaters, and so on, until stable results can be obtained. This process has two main advantages: (1) it focuses the inquiry on relevant knowledge, and (2) it continues to deliver provisional results while inquiry goes on.

A third function of defeasibility is that of enabling our knowledge structures to persist to a certain degree over time remaining a shared asset, while each one of us is constantly exposed to new information, often challenging the information we already have. We have indeed two basic strategies for coping with the provisional nature of human knowledge have two strategies for dealing with the provisional nature of human knowledge.

The first strategy consists in viewing our persistent knowledge as a set of universal propositions, which may be falsified by any particular fact contradicting them. When we discover a case where such universal propositions lead us to false (or unacceptable or absurd) conclusions, we must conclude that our theory (the set of the propositions leadin been falsified (proved to be unacceptable). Thus we must abandon one or some such propositions and substitute them with new universal propositions, from which the false conclusion is no longer derivable.¹⁶

Aristotelis ad Nicomachum [Frankfurt am Main, Minerva, 1968]).

¹⁴Pollock, *Cognitive Carpentry: A Blueprint for How to Build a Person*; J. L. Pollock, "Perceiving and Reasoning about a Changing World" (1998) 14 *Computational Intelligence*, 489–562.

¹⁵Pollock, *Cognitive Carpentry: A Blueprint for How to Build a Person* 40.

¹⁶According to the hypothetical deductive model described by Karl Popper (K. R. Popper, *The Logic of Scientific Discovery* [London, Hutchinson, 1959]). On the revision of knowledge in face of change, see the groundbreaking

The second strategy consists on the contrary, in assuming that we must keep our general propositions even when their isolated application would lead us to wrong conclusions. We need to assume that such propositions only concern the majority of cases, or the normal cases, so that the exception serves the rule, or at least, does not damage it rule. To deal appropriately with the anomalous case, then, it not necessary the case to abandon the rule or to change its formulation, but rather to assume that the operation of the rule is limited on grounds different from the grounds supporting the rule itself, grounds which may consist in an exception, in a prevailing principles to the contrary, in conflicting rules coupled with criteria for solving rule-conflicts (for instance, the traditional principles, according to which a higher, more special or later law norm over a lower, more general or earlier one). This enables a certain degree of stability in legal knowledge, though it does not exclude the need to abandon a defeasible general rule, when it does no longer reflect a “normal” connection, when it is superseded by subsequent information (as in implicit derogation to legal norms), when it is explicitly removed from the knowledge-base (as in explicit derogation).

8 Collision and Defeat

In order to have a first look into the working of defeasible reasoning, let us consider an example that is frequently referred to by epistemologists.

Assume that I believe that swans are normally white, and that I am told that there is a swan in the park. This enables the following defeasible inference (an instance of the schema *statistical syllogism*):

- (1) most swans are white;
- (2) the bird in the park is a swan

- (3) the bird in the park is white

However, when I look out of the window, I see that the bird in the park, although being unmistakably a swan, looks kind of pinkish. This prompts the following perceptual inference:

- (1) I am having a pink image of the bird in the park

- (2) the bird in the park is pink

Thus, I am pushed towards conflicting conclusions, supported by competing defeasible inferences (according to the first inference, the bird is white, according to the second one it is pink). Assume that, being a moderate empiricist, I consider the perceptual inference to be stronger than the statistical one. Therefore, I should abandon my *pro-tanto* belief that the swan is white and accept (though provisionally) that it is pink. In such a case, we say that the inference concluding that the swan is white is defeated by the perceptual inference according to which it is pink.

Cognitive processes like the one we have just considered can be explained by introducing two notions, collision and defeat.

Definition 8.1 *Collision.* Let M be a reason for adopting Q and M^* be a reason for adopting Q^* . We say that there is a collision between M and M^* , when the combined cognitive state which consists in endorsing both of M and M^* does not support adopting both of Q and Q^* .

When one finds oneself in a collision, one is prevented from performing both colliding inferences. One may be prevented from performing just one of them, or one may be prevented from performing both. Those inferences which are prevented by the collision, are said to be *defeated*, while the reason that prevents deriving a conclusion is said to be a *defeater*:¹⁷

work of Alchourrón, Gärdenfors and Makinson: C. E. Alchourrón, P. Gärdenfors, and D. Makinson, “On the Logic of Theory Change: Partial Meet Functions for Contractions and Revisions” (1985) 50 *Journal of Symbolic Logic*, 510–30, P. Gärdenfors, *Knowledge in Flux* (Cambridge, Mass., MIT Press, 1987).

¹⁷See Pollock and Cruz, *Contemporary Theories of Knowledge* 195ff.

Inference \mathcal{A}	Inference \mathcal{B}
(1) Most swans are white	(1) I perceive the image of a pink bird in the park
(2) The bird in the park is a swan	
(3) The bird in the park is white	(2) The bird in the park is pink

Table 4: *Rebutting collision: Inference \mathcal{A} collides with inference \mathcal{B}*

Definition 8.2 *Defeat.* Let M be a reason for adopting Q . Premise M^* defeats (is a defeater for) M , iff the combined state consisting in endorsing both of M and M^* does not support adopting Q .

In the example above, mental state m (believing that most swans are white and that the bird in the park is a swan) collides with mental state m^* (having the percept that that the bird in the park is pink). State m alone was a reason for adopting a q (believing that the bird in the park is white), according to *statistical syllogism*. However, the combined cognitive state consisting in believing m and having percept m^* is no reason for adopting q : Because of the defeater m^* , the inference from m to q is blocked or defeated. The conflict between the two inferences \mathcal{A} and \mathcal{B} originated by the two colliding reasons m and m^* is reproduced in Table 4.¹⁸

9 Collisions and Incompatibility

The type of collision exemplified at the end of previous paragraph (conflict of inferences leading to incompatible conclusion) can be called *rebutting collision*. As a first approximation, rebutting collision can be defined as follows:

Definition 9.1 *Rebutting collision.* There is a rebutting collision between reasons M and M^* when

- M is a reason to adopt Q ,
- M^* is a reason to adopt Q^* , and
- Q is incompatible with Q^* .

However, only in some cases can incompatibility be assessed by looking only at the directly concerned proposition (more exactly, cognitive states): These are the cases when such propositions would be incompatible in all logically possible situations. For instance, it cannot be that case that Tom both stole a car and did not steal it; or that Mary both is obliged to repair some damage and is not obliged to do so.

In many other cases, to assess the incompatibility of propositions one needs to consider further information, concerning meaning connections, causal links, or further facts. For example, to assess that being pink and being white are incompatible states of one same object (our swan) one must know that an object cannot have two colours at the same time. Similarly, to know that there is a conflict between the fact that others detain and process one's personal data and one's free self-determination (as affirmed by privacy supporters), much psychological and sociological background-knowledge is to be assumed. Finally, to know that low inflation and full occupation are incompatible (in a certain economic context) one must have a lot of economical information.

¹⁸Here and in the following, I shall refer to sequences of propositions—and in particular to inferences and arguments—by using symbols \mathcal{A} , \mathcal{B} , ...

Inference \mathcal{A}	Inference \mathcal{B}
(1) I perceive a pink bird in the park	(1) There is red light outside
(2) The bird in the park is pink	(2) Perceiving a pink object under red light does not warrant that it is pink
(3) My perceiving a pink bird does not warrant that it is pink	(3) My perceiving a pink bird does not warrant that it is pink

Table 5: *Undercutting collision: Inference \mathcal{A} collides with inference \mathcal{B}*

Therefore, very often the incompatibility of two conclusions cannot be immediately detected by the reasoner, but is rather the result of finding relevant information and of bringing it to bear through reasoning processes.¹⁹

10 Undercutting Collisions

Besides rebutting collisions, there is another way in which reasons can collide. This happens when the reasoner has a reason to believe that, in the present circumstances, a reasoning schema does not apply, since under those circumstances the reason of the schema provides no reliable support to its conclusions.

Let us consider a variation of the ornithological example introduced in Section 8. Assume that I am seeing (or rather having the vision of) a pink bird, and that I come to believe, according to a perceptual inference, that the bird is indeed pink. I notice, however, that there is beautiful sunset now, throwing red light over all things. I know that red light makes white objects look pink. Therefore, I conclude that it would be unreasonable for me to believe that the swan is pink on the basis of the only fact that it looks pink to me (under the present conditions the pink-looking swan might as well be white).

Note that this reasoning does not tell me that the swan is white, since pink objects would still look pink under red light: It only undermines the previous inference, without providing a different conclusion (Table 5).

This type of collision (collision between a reason M , and a reason M^* , indicating that M is unreliable) shall be called *undercutting collision*. More exactly, we can define undercutting collisions as follows:²⁰

Definition 10.1 *Undercutting collision.* There is an undercutting collision between reasons M and M^* when

- M is a reason for adopting Q ,
- M^* is a reason for believing that M does not support Q .

Under these conditions we also say that M^* undercuts M .

For instance, my awareness that there is red light is a reason for believing that the fact that the bird looks pink does not support concluding that it is pink. Such awareness *collides with*, and more exactly *undercuts*, my considering the pink appearance of the bird as a reason for concluding that it is pink.

¹⁹For an illuminating discussion, see C. Perelman and L. Olbrechts-Tyteca, *Traité de l'argumentation: la nouvelle rhétorique*, 5th ed. (Brussels, Éditions de l'université de Bruxelles, 1988), sec. 46

²⁰According to Pollock and Cruz, *Contemporary Theories of Knowledge* 196.

11 Preference-Based Reasoning

When one endorses two colliding reasons, one cannot derive the conclusions of both of them: At least one of these reasons is defeated.

We may thus distinguish the following two cases:

- If one reason, assume R_1 , prevails over the other, R_2 , we may reject R_2 and endorse R_1 (so that only R_2 is defeated).
- If, on the contrary, none of the two reasons prevails, both reasons are defeated.

Thus, it appears that a reason R_1 *defeats* a reason R_2 , whenever R_1 collides R_2 , and R_2 does not prevail over R_1 . When, additionally, R_1 prevails over R_2 (item 1 above) then R_1 *strictly defeats* R_2 (defeats R_2 , but is not defeated by it).²¹

In case of rebutting collision the stronger reason outweighs and strictly defeats its competitor. Consider for example, a recent case that had to be addressed by the Italian privacy authority. It concerned the case of a woman who requested for health reasons to access data concerning the DNA of her father (the data was stored in the database of a hospital), who did not give his consent. Therefore, the authority needed to balance the privacy-based inference (the father's data could not be provided since sensitive data cannot be communicated without the consent of the data subject) and the inference based upon the right to health (one has a right to obtain what is needed for one's health, such as, for that woman, access to the father's DNA). In such a case the health-based inference was considered to be preferable to the privacy-based inference, so that the latter was viewed as being strictly defeated, and the former inference dictated the outcome of the case.

In case of undercutting collisions, on the contrary the undercutter prevails: It strictly defeats the attacked reason. This seems indeed the most reasonable way of approaching undercutting. If I just have a reason R_1 to believe that the bird in the park is pink, and a reason R_2 to conclude that R_1 is unreliable, I should not conclude that the bird is pink (on the basis of the unreliable reasons R_1). Similarly, assume that I find in a law text two rules: a rule r_1 , and a rule r_2 saying that r_1 is inapplicable under certain circumstances. If these circumstances are satisfied in the present case I should conclude that r_1 is indeed inapplicable, and refrain from deriving its conclusion.

In conclusion, when facing a collision, we should reason as follows:

- in rebutting collisions, we should compare the strength of the conflicting reasons, and assume that any reason that is not stronger than its competitor is defeated (only stronger reasons prevail);
- in undercutting collisions, we should assume that the undercutter always prevails.

Let us consider a further example on undercutting. Assume that I have heard two different accounts of the same event, from two friends, John and Mary:

- John tells me A ,
- Mary tells me $\text{NON } A$ (A is not the case), and
- I consider both of them to be sufficiently reliable, under normal circumstances (so that I would have believed each one of them, if the other had kept silent).

It seems that I should view the statements of Mary and John as defeating one another, and refrain from forming any belief on the matter (neither A nor $\text{NON } A$), unless I can assume that one of the two conclusions more reliable than the other. In the latter case, I should provisionally (defeasibly or *pro-tanto*) believe the content of the more reliable statement, and reject the other.

Similarly, assume that I come to know both of the following facts:

²¹Note that the propositions that (a) [R_2 does not prevail over R_1] and that (b) [R_1 prevails over R_2] are not equivalent: The first also holds when the outcome of the conflict of the two reasons is a draw (is undetermined), while the second requires that the conflict is positively decided in favour of R_1 (I assume that the prevailing-over relation is antisymmetric).

- Tom intentionally caused harm to Mary's property, driving into her fence, and
- he acted in a state of necessity, to avoid hitting a child who was crossing the street.

Under such conditions, I should conclude that Tom has no duty to compensate Mary's damage, since the conclusion that he is not liable (having acted for the necessity of saving another's life) prevails over the conclusion that he is liable (having intentionally damaged another's property).

This way of reasoning assumes, however, that one has a way of determining what conclusion (if any) is more strongly supported. In some cases, one may do that by adopting mathematical methods. For example, one may use probability calculus, and assume that the strength of each conclusions corresponds to its probability, which can be computed by combining the probabilities of its pre-conditions. Other numerical calculations of the comparative strength of beliefs have also been proposed, which diverge under some respects from standard probability.²²

I cannot here discuss the merit of numerical methods for assessing credibility, which would require us to address the technicalities of probability and statistics. Let us just observe, in general, that there are certain specific legal issues (in particular in the domain of evidence) where numerical calculi can provide appropriate answers. However, these calculi do not provide a generally applicable solution for dealing with incompatible conclusions in moral and legal reasoning. In the law, it is rather the case that one needs to engage in priority reasoning, that is, one needs to bring to bear further information, and to decide accordingly which inference is to prevail. Though this information rarely is numerical, it enables in most cases a sufficiently precise assessment.

12 Reinstatement

I need to introduce a further idea that explains the characteristic procedure of defeasible reasoning: the notion of *reinstatement*. When a defeater is strictly defeated by a further inference, then the inference originally attacked by the defeat may recover its capacity to establish its conclusion.

Let us develop further the pink-swan example. As before, assume that the bird of which I am having a pink vision is a swan, so that I would conclude for its whiteness, unless this conclusion was defeated by my perception of its pinkness. Assume also that I realise that the sun is setting, and that its light makes all white things look pink. My awareness of this undercuts the conclusion that the bird is pink (see Table 6 on the facing page). As a consequence of the perceptual inference being undercut, the inference for whiteness is reinstated: I again *pro-tanto* believe that the bird is white (being a swan).

Let us consider another example, concerning the difference between rebutting and undercutting. Let us assume that a detective is investigating the violent death of John, of which Mary, John's inconsolable girlfriend, is accused. The detective believes that Mary loved John, but has evidence that her clothes were stained with John's blood. The information he has allows him to build two inferences, which rebut one another: the inference according to which Mary did not kill John, since she loved him (inference \mathcal{A}), and the inference that she killed him, since her clothes were stained with John's blood (inference \mathcal{B}).

Assume that the detective also believes that inference \mathcal{B} is preferable to inference \mathcal{A} (he gives more credit to chemistry than to psychology). This allows the detective to endorse the conclusion of inference \mathcal{B} : At this stage of the inquiry he forms the belief that Mary killed John (inference \mathcal{B} defeats inference \mathcal{A} , but inference \mathcal{A} does not defeat \mathcal{B}).

However, assume that the detective discovers that Lisa (John's previous girlfriend) tried to frame Mary, by staining Mary's dress with John's blood. Again, this alone would not be a reason to believe that Mary did not kill John, but rather a reason for considering that the inference \mathcal{B} (from Mary's having blood-stained clothes to her being the murderer) is unreliable, and therefore to view this inference as being undercut.

The latter inference, let us call it \mathcal{C} , by undermining inference \mathcal{B} results in reinstating inference \mathcal{A} . Thus, we pass from the following situation (I use a smaller character for defeated inferences):

²²See, for example, Pollock, *Cognitive Carpentry: A Blueprint for How to Build a Person*.

Inference B is preferable to inference A

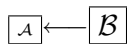
<i>Inference A</i>	<i>Inference B</i>
(1) Mary loved John	(1) Mary's clothes were stained with John's blood
(2) people normally do not kill their loved ones	(2) if one's clothes are stained with the victim's blood, then normally one has killed the victim
(3) Mary has not killed John	(3) Mary has killed John

Table 6: *Defeat by rebutting*

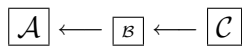
Inference B is preferable to inference A

<i>Inference A</i>	<i>Inference B</i>	<i>Inference C</i>
(1) Mary loved John	(1) Mary's clothes were stained with John's blood	(1) Lisa stained John's clothes with blood
(2) people normally do not kill their loved ones	(2) if one's clothes are stained with the victim's blood, then normally one has killed the victim	
(3) Mary has not killed John	(3) Mary has killed John	(2) Mary's clothes being stained with John's blood does not warrant that Mary killed John

Table 7: *Reinstatement through undercutting*



where inference A is strictly defeated by inference B , into the situation:



where inference B is strictly defeated by inference C , and consequently A is reinstated. Thus given all of A , B , and C , the detective would conclude, according to A , that Mary did not kill John (since she loved him).

A further type of undercutting collision is provided by inferences presupposing that one does not have a certain mental state: These inferences collide exactly with one's adoption of that mental state. Such inferences are made when one assumes that, if a certain proposition A were true, one would have come to believe A , or at least one would possess the information that enables one to infer A .²³ Therefore, when

²³This seems to be the rationale of the so-called autoepistemic logic (R. C. Moore, "Possible-World Semantics for Autoepistemic Logic", in *Readings in Nonmonotonic Reasoning*, ed. by M. L. Ginsberg [Los Altos, Cal., Morgan Kaufmann, 1987], 137–42), and of the use of negation by failure in logic programs (K. L. Clark, "Negation as

<i>Inference A</i>	<i>Inference B</i>
(1) If I do not believe that one is guilty, then I presume that one is innocent	(1) I have evidence proving that Mary is guilty
(2) I do not believe that Mary is guilty	
(3) Mary is innocent	(2) Mary is guilty

Table 8: *Undercutting presupposition*

failing to form the belief that A , one may conclude that A does not hold and reason accordingly. However, when one positively finds that A holds, one should abandon inferences based upon the assumption that A does not hold. For example, a judge who does not believe that a person is guilty should base his verdict upon the thesis that the assumption that the person is innocent (*in dubio pro reo*), as shown in Table 8.

It may be argued that this kind of situation takes place more generally in the law, whenever a rule makes a certain legal effect dependent upon the fact that certain facts are not shown to hold.²⁴ This is indeed the typical way of reasoning with legal *presumptions*. Consider for example the rule in the Italian civil code, which says that one is presumed to have received a message when the message arrived at one's address, unless one proves that one had no possibility of accessing the message. The judge, according to this rule, will be able to conclude that one has received a message on the basis of the sole fact that the message arrived at one's address, under the assumption that the addressee could read the message. However, if there is evidence showing that the addressee was under the impossibility of accessing the message, this conclusion will be defeated.

13 Undercutting in Practical Reasoning

Undercutting also applies to practical reasoning. The need to act on the basis of defeasible conclusions is particularly apparent when one—pressed by many goals and commitments—needs to quickly find plans for achieving each of these goals, and adopt those plans without further ado (even when one is aware that there may be better plans which one could find, having additional time at one's disposal). However, as I have observed above, if a better plan becomes available before action takes place, one should abandon one's previous intention, and adopt the new plan.

Consider, for example, the inference leading me to adopt the intention of leaving on Monday for Barcelona, since this plan is sufficiently good: It is better than staying at home (the inactivity option), and is better than any plan I have so far considered. This inference gets defeated when I build a plan for leaving on the Sunday before, which allows me to get a cheaper ticket and to make a short visit to the main monuments of the city.

In general, in teleological reasoning one moves from (1) having a goal G and (2) believing that plan (instruction) P_1 is a satisfactory way of achieving G , into (3) intending to realize P_1 . However, this inference can be undercut when the reasoner comes to believe that a different plan, let us call it P_2 , is better than P_1 , as a way of achieving G . The defeat schema seems therefore to be the following:

Failure", in *Readings in Nonmonotonic Reasoning*, ed. by M. L. Ginsberg [Los Altos, Cal., Morgan Kaufmann, 1987], 311–25).

²⁴See G. Sartor, "Defeasibility in Legal Reasoning", in *Informatics and the Foundations of Legal Reasoning*, ed. by Z. Bankowski, I. White, and U. Hahn (Dordrecht, Kluwer Academic, 1995), 119–57.

<i>Inference A</i>	<i>Inference B</i>
(1) I have the goal of getting the new book of my friend Henry Prakken	(1) Plan p_2 (getting a free author's copy) is a more convenient way of getting Henry's book
(1) Plan p_1 (buying it on line) is sufficiently good way of achieving this goal	
(1) I intend to implement plan p_1	(1) the convenience of plan p_1 does not support its adoption

Table 9: *Teleological defeat*

Defeating schema: *Teleological defeat*

- (1) believing that plan P_2 is a better way of achieving goal G than plan P_1
 _____ IS AN UNDERCUTTING DEFEATER AGAINST
 (2) using *teleological inference*, for adopting P_1 as a way of achieving G

Assume for example that a judge has so far decided cases on product liability by requiring that customers prove the producers' fault. Assume that the judge adopted this policy since she believed that this was a satisfactory way to induce producers to take care in order to avoid releasing faulty products. Assume, however, that the same judge comes now across some law-and-economics literature that shows that strict liability provides a stronger incentive to preventing damage to the customer. The judge should then be induced to abandon the fault liability approach in favour of strict liability (for simplicity's sake, I discount consideration concerning the need that the judge respects legislation or precedents, and co-ordinates her action with the action of her colleagues, of legislators and of the citizens).

Here is the corresponding defeat instance:

Defeating instance: *Teleological defeat*

- (1) believing that strict liability is a better way of preventing damage to the customers than fault liability
 _____ IS AN UNDERCUTTING DEFEATER AGAINST
 (2) using *teleology*, for adopting fault liability as a way of preventing damage to the customers

Another example of this type of defeat is represented in Table 9.

14 Defeasible Reasoning and Probability

We may wonder whether defeasible reasoning is the only, or the best, way to deal with incomplete information. In particular, we need to consider the main alternative to it, that is, *probability calculus*, and in particular the versions of probability calculus that are based upon the idea of *subjective probability*. In fact, probability calculus has more solid scientific credentials than any logic for defeasible reasoning, and a rich history of successful applications in many domains of science and practice.

Following the probability approach, the reasoner—rather than facing incompatible beliefs (like the belief that John was driving the car, when the car run over Tom, and the belief Mary was driving the car, in the same occasion), and then having to make a choice—would come to the consistent conclusion that incompatible hypothesis have different probabilities (for instance, one would reach the conclusion that there is a 40% chance that John run over Tom, and a 60% chance that Mary did it). Probabilistic inference, as it is well known, determines the probability of an event on the basis of the probability of other events, according to the mechanism of probability calculus: If there is a 80% chance that Tom will have walking

problems because of having been run over, there is a 32% chance ($40\% * 80\%$) that Tom will have such problems having been run over by John, and 48% chance ($60\% * 80\%$) that he will have such problems having been run over by Mary.

Here I cannot introduce probability calculus, nor discuss the many difficult issues which are related to it (especially when ideas of probability and causation are combined). I shall just point at a few issues concerning the probability calculus, which make it inadequate as a general solution for dealing with uncertainty in legal reasoning.

The first issue concerns practicability: Often we do not have enough information for assigning numerical probabilities in a sensible way. For instance, how do I know that there is a 40% probability that John was driving and 60% probability the Mary was driving? In such a case, it seems that either we arbitrarily attribute probabilities, or, with equal arbitrariness, we assume that all alternative ways in which things may have gone have the same probability.

The second issue is a conceptual one: Though it makes sense to ascribe probabilities to epistemic propositions, it makes little sense to assign it to practical information. What does it mean that a certain desire (goal) or intention (instruction), has a certain probability? What does it mean that a value, a normative connection, an obligation have a certain probability?

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 the other options may turn out to be preferable), and applying their reasoning to this hypothesis (while possibly, at the same time, exploring would be the case, if things turned out to be different). We do not usually assign probabilities, and then compute what further probabilities follow from such an assignment.

This cognitive behaviour corresponds to the reasoning skills of which we are naturally endowed. Humans, once they have definite beliefs or hypotheses, are able to develop inference chains, store them in their minds (keeping them unconscious until needed), and then retract any of such chains when one of its steps is defeated. On the contrary, humans are bad at assigning numerical probabilities, and even worse in deriving further probabilities and in revising a probability assignment in the light of further information.

Our incapacity of working with numerical probabilities certainly is one of the many failures of human cognition (like our incapability of quickly executing big arithmetical calculations). In fact, computer systems exist which, using complex nets of probabilities (these are called *belief networks* or *probability networks*), perform very well in certain domains by manipulating numerical probabilities much quicker and more accurately than a normal person.²⁵

However, our bias toward adopting a binary, rather than a probabilistic approach (endorsing one alternative, rather than assigning probabilities to all of them), in face of uncertainty has some advantages: It focuses cognition on the implication of most likely situation, it makes it easier to make long reasoning chains, it facilitates building scenarios (or stories) which then may be evaluated according to their coherence, it enables linking epistemic cognition with binary decision-making (it may be established that one has to adopt decision *B* if *A* is the case, and *NON B* if *A* 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.

In a social context a binary approach makes it easier to replicate the reasoning and thinking of other people: One can forecast more easily a binary choice by another than the ascription of probabilities, and such a binary choice can become the focus of social expectations.²⁶

The limited applicability of probability calculus in the practical domains does not exclude that there are various practical and legal issues where statistics and probability provide decisive clues, as when

²⁵See, for a short introduction, S. J. Russell and P. Norvig, *Artificial Intelligence. A Modern Approach*, 2nd ed. (Englewood Cliffs, N. J., Prentice Hall, 2003), chap. 14. For a collection of important contributions, see G. Shafer and J. Pearl, eds., *Readings in Uncertain Reasoning* (Los Altos, Cal., Morgan Kaufmann, 1990).

²⁶This aspect is particularly emphasised by N. Luhmann, *Rechtssystem und Rechtsdogmatik* (Stuttgart, Kohlhammer, 1974), chap. 6, sec. 1.

scientific evidence is at issue.²⁷

15 The Idea of Defeasibility in the Practical Domain

The notion of defeasibility, before becoming the focus of much AI research²⁸ was deployed by some epistemologists²⁹. Even before that, however, the idea of defeasibility was frequently applied and sometimes studied in the domain of practical (moral and legal) reasoning.

The word *defeasible* is a traditional legal term, indicating the possibility that a legal instrument is voided in special circumstances. In fact, it is a typical feature of law and morality that they can only offer standards appropriate for normal situations. When exceptional circumstances occur, these standards may need to be put aside.³⁰ This is well expressed in a famous Aristotelian citation, where defeasibility is considered not a fault, but a natural feature of legal rules:

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.³¹

Similarly, Aquinas makes the following precise observation:

[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.³²

The very notion of defeasibility is at the centre of the work of David Ross, an outstanding Aristotelian scholar and moral philosopher, who developed a famous theory of *prima-facie* moral obligations.³³ Ross, who endorsed a pluralist form of moral intuitionism, relates defeasibility to the possibility that moral principles are overridden by other moral principles in concrete cases:

²⁷On scientific evidence, see, among others S. Haack, "Truth and Justice, Inquiry and Advocacy, Science and Law" (2003) 7 *Associations*, 103–14, and S. Haack, *Defending Science within Reason* (Amherst, N. Y., Prometheus, 2003), 233ff.

²⁸For a collection of seminal contributions, see Ginzberg, *Readings in Nonmonotonic Reasoning*.

²⁹In particular, see J. L. Pollock, *Knowledge and Justification* (Princeton, N. J., Princeton University Press, 1974), and R. M. Chisholm, *Theory of Knowledge*, 2nd ed. (Englewood Cliffs, N. J., Prentice Hall, 1977)

³⁰On defeasibility in the law, cf. among others: T. F. Gordon, "The Importance of Nonmonotonicity for Legal Reasoning", in *Expert Systems in Law: Impacts on Legal Theory and Computer Law*, ed. by H. Fiedler, F. Haft, and R. Traunmüller (Tübingen, Attempto, 1988), 111–26; Sartor, "Defeasibility in Legal Reasoning"; D. N. MacCormick, "Defeasibility in Law and Logic", in *Informatics and the Foundations of Legal Reasoning*, ed. by Z. Bankowski, I. White, and U. Hahn (Dordrecht, Kluwer Academic, 1995), 99–117.

³¹Aristotle, *Nicomachean Ethics* (Oxford, Oxford University Press, 1954), 1137b.

³²T. Aquinas, *Summa Theologiae* (Allen, Texas, Benzinger Bros, 1947), I-II, q. 94, a. 4.

³³W. D. Ross, *The Right and the Good* (Oxford, Clarendon, 1930); , W. D. Ross, *Foundations of Ethics* (Oxford, Clarendon, 1939).

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.³⁴

The notion of defeasibility, quite usual in legal practice and in doctrinal work, was brought to the attention of legal theorists by H.L.A Hart³⁵, whose observations anticipate the current debate on the topic:

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.

In legal defeasibility two aspects must be distinguished:

1. the inventive, heuristic reasoning required for understanding the inadequacy of a general rule in a particular context and for stating a corresponding exception or rebuttal;
2. the structure of the reasoning that takes into account exceptions, rebuttals, and presumptions in order to reach conclusions justified by all available knowledge.

The second aspect does not fall outside logic. Legal defeasibility implies that for justifying a legal conclusion it is normally, but not always, sufficient to select from the available pool of premises a subset supporting that conclusion. In fact, the inference of that conclusion may be defeated, if an exception or an “unless” clause is satisfied, or if a relevant presumption is found to be contradicted.

16 Defeasibility in Legal Language

The legislator often explicitly foresees the defeasibility of legal rules, by using different linguistic structures. For example, to establish that tort liability is excluded by self-defence or 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 state of necessity. The absence of self-defence and of a state of necessity is presumed.

According to all these formulations, for concluding that one must make good a certain damage it is normally sufficient to ascertain that one voluntarily caused that damage, but this inference is defeated if the person turns out to have acted either in a state of necessity or in self-defence.

By distinguishing circumstances that have to be positively established to derive a certain conclusion from circumstances susceptible to blocking this derivation, the law divides the *burden of proof* between the parties. The plaintiff, who is interested in establishing a certain legal conclusion, has the burden of establishing the supporting circumstances. The defendant, once the plaintiff has succeeded, has the burden

³⁴Ross, *Foundations of Ethics* 84-5.

³⁵H. L. A. Hart, “The Ascription of Responsibility and Rights”, in *Logic and Language*, ed. by A. Flew (Oxford, Blackwell, 1951), 145–66, 152.

of establishing the impeding circumstances. Though defeasibility may not fully explain the dialectics of legal procedures, it provides their logical background.³⁶

17 Defeasibility in Legal Concepts and Procedures

Defeasibility is also an essential feature of *conceptual constructions* in the law. Legal concepts have to be applied to such a diverse domain of instances that they can at best offer a tentative and generic characterisation of the objects to which they apply, a characterisation which has to 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 one of its instances—would paradoxically run against 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 (for example, contract), then special exceptions are introduced for species of this genus (like the sale contract), and finally further exceptions may be introduced for specific subspecies (like the sale of real estates). Consequently, when using conceptual hierarchies we must apply to a certain object the rules concerning the category in which it is included only in so far as no exceptions emerge concerning a subcategory in which that object is also included.

Defeasibility can be consciously 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 means of *distinctions*, namely, by means of exceptions introduced for specific contexts.

General attitudes of legal reasoning can also be explained and justified as *defeasible presumptions*. For example, interpretation standards and even judicial precedents (*rationes decidendi*) bind judges only defeasibly, in the sense that they should be followed only in so far as they are not overridden by (strong) reasons to the contrary.³⁷ Even the obligation to apply legislation is considered to be defeasible by many authors, who believe that statutory texts are not legally binding when completely unacceptable from a moral standpoint.³⁸

Similarly, the due consideration for widespread social attitudes, legal traditions, socially accepted common-sense rules (so much stressed by theorists of argumentation) can be distinguished from blind conservatism only when those factors are given the statute of defeasible presumption. This was, by the way, also the spirit of the principle of inertia advocated by Chaim Perelman³⁹, according to which any opinion adopted in the past should only be abandoned when having sufficient reasons to do that. More generally, defeasible endorsement is the appropriate cognitive attitude with regard to what Aristotle calls *endoxa*, meaning, any “opinion held by everyone or by the majority or by the wise—either all of the wise or the majority or the most famous of them—and which is not paradoxical”.^{40,41}

³⁶For a logical model of the burden of proof, obtained by extending a logic for defeasible reasoning, see H. Prakken and G. Sartor, “Formalising Arguments about the Burden of Persuasion”, in *Proceedings of the 11th international conference on Artificial intelligence and law (ICAIL-2007)* (New York, N. Y., ACM, 2007). For a logical analysis of presumptions, see H. Prakken and G. Sartor, “More on Presumptions and Burdens of Proof”, in *Proceedings of JURIX 2008*, ed. by E. Francesconi, G. Sartor, and D. Tiscornia (Amsterdam, IOS, 2008).

³⁷R. Alexy, *A Theory of Legal Argumentation: The Theory of Rational Discourse as Theory of Legal Justification* (Oxford, Clarendon, 1989), 274-279.

³⁸A. Peczenik, *On Law and Reason* (Dordrecht, Kluwer, 1989); , R. Alexy, *Begriff und Geltung des Rechts* (Freiburg, Steiner, 1992).

³⁹C. Perelman and L. Olbrechts-Tyteca, *The New Rhetoric: A Treaty on Argumentation*. (Notre Dame, Ind., University of Notre Dame Press, 1969), 142.

⁴⁰Aristotle, *Topics* (New York, N. Y., Random House, 1941), 104a8-13.

⁴¹Aristotle, *Topics*).

Finally, also the procedural aspect of defeasibility needs to be considered. This aspect concerns the fact that, as observed above, defeasible reasoning activates a structured process of inquiry—based upon drawing *prima facie* conclusions, looking for their (*prima facie*) defeaters, looking for defeaters of defeaters, and so on—until stable results can be obtained. Such a process reflects the natural way in which legal reasoning proceeds. This is particularly the case in the application of the law to particular situations, when one has to bring to bear on particular situations the different, and possibly conflicting legal rules that apply to them, and adjudicate the conflicts between these rules.

Klaus Guenther⁴² has affirmed that the application of the law is characterised by a *sense of appropriateness*, intended as the ability of impartially taking into consideration all different features of the considered situation, according to all valid rules which may apply to it. Defeasible reasoning can indeed be viewed as a formalisation of Günther's sense of appropriateness. The model of defeasible reasoning is indeed based on the distinction between on the one hand the adoption of general defeasible rules, each of which only takes into account a few aspects of possible situations, and on the other hand the application of this rules to concrete situations, where one needs to take into account all aspects that are relevant according to all applicable rules.

In fact in the law, as in “any subject on which difference of opinion is possible, the truth depends on a balance to be struck between two sets of conflicting reasons”⁴³, the point by which “truths have reached the point of being uncontested” (ibid., 51) is far from being attained. Therefore, in legal reasoning the use of *positive logic*, which relates a thesis to its supporting grounds, must be supplemented with critical discussion of the opinion to the contrary, that is, by that *negative logic* which “points out weaknesses in theory or errors in practice, without establishing positive truths” (ibid, 109). Logics for defeasible reasoning have the merit of unravelling exactly the basic structure of Mill's negative logic.

Defeasibility of legal reasoning also reflects the dialectics of judicial proceedings where each party provides arguments supporting its position, which conflict with the arguments of the other party. This debate may also be transferred into the judicial opinion that resumes the results of the dispute and determines its output. To convincingly justify a judicial decision in a case involving serious issues, it is not sufficient to produce a single argument, but it is necessary to establish that the winning argument prevails over arguments to the contrary, especially those that have been presented by the losing party.

Also 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. In this perspective, doctrinal reasoning may be viewed as consisting in an exercise of *unilateral dialectics*, intended 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”.⁴⁴

18 Overcoming Legal Defeasibility?

Some authors have suggested that the law ought to be recast in a set of consistent axioms, which would lead to compatible outcomes, according to deductive inference, in any possible factual situation. This reformulation of the law would eliminate normative conflicts, and therefore would allow us to avoid legal defeasibility (at least of the rebutting type). This idea has been affirmed in particular by Carlos Alchourrón and Eugenio Bulygin:⁴⁵ the legislator and the doctrinal jurist should join their effort toward providing axiomatic reformulation of the law, or at least of particular sections of it. As Euclid developed an axiomatic model of geometry and as modern natural and social science (in particular economics) have developed

⁴²K. Günther, *The Sense of Appropriateness* (Albany, N. Y., State University of New York Press, 1993); , K. Günther, “A Normative Conception of Coherence for a Discursive Theory of Legal Justification” (1989) 2 *Ratio Juris*, 155–66.

⁴³J. S. Mill, “On Liberty”, in *On Liberty and Other Essays*, ed. by J. Gray (Oxford, Oxford University Press, 1991), 5–130, 41.

⁴⁴N. Rescher, *Dialectics: A Controversy-oriented Approach to the Theory of Knowledge* (Albany, N. Y., State University of New York Press, 1977), 47.

⁴⁵C. E. Alchourrón and E. Bulygin, *Normative Systems* (Vienna, Springer, 1971).

axiomatic models of their knowledge, so the legislator and the jurist should axiomatise certain domains of the law. By adding to such an axiomatisation the description of specific cases (a description to be accomplished in the same language used for the axiomatisation), we should obtain a set of premises from which the legal discipline of such cases can be deduced. Accordingly, the model of the deductive application of the law should be extended beyond the so called judicial syllogism so as to include the whole of predicate logic and deontic logics.

In a more recent work⁴⁶ the same Carlos Alchourrón has reaffirmed that the ideal of the axiomatisation of the law should inspire legislation and doctrine, and can contribute bring legal studies and scientific method together: as in science the phenomena to be explained, the *explanandum*, should be logical consequences of a set of premises, the *explanans*, containing scientific laws and the description of particular facts,⁴⁷ so in the law the content of a particular legal conclusion (decision) should be the deductive consequence of a set of premises including both general norms and the description of specific facts.

We need to consider, however, that it is very doubtful whether such a reformulation of the law is really feasible, and, even when it were feasible, it is doubtful whether it would be useful.

It is indeed doubtful whether such an axiomatisation would really make the law easier to understand and apply. Legal prescriptions would need to become 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 adjudicating their conflicts is taken into consideration. Finally, by rejecting defeasible reasoning, one would lose its capacity of providing provisional outcomes while the inquiry goes on.

The need of representing the law in ways which facilitate defeasible reasoning do not imply that the current way of expressing legal regulations in statutes and regulatory instruments cannot be improved. Large improvements in legislation techniques on the contrary are required to cope with the many tasks that need to be carried out by modern legal systems. However such improvements should not aim at producing a conflict-free set of legal rules, just for the sake of logical consistency. They should rather aim at producing legal texts that can be more easily understood and applied.⁴⁸

This objective requires a skilful use of the very knowledge structures (such as conceptual hierarchies, speciality, or the combination of rules and exceptions) that enable defeasible reasoning. Such structures are largely used also outside the law, in situations when one has to express precisely ways of dealing with complex and changing situations. For example, conceptual hierarchies (enabling defeasible inheritance) have become a standard programming technique in object-oriented programming (the mainstream programming methodology nowadays), while defeasible reasoning (in the form of negation by failure) provides a core function of logic programming.⁴⁹

Accepting defeasibility in the law has significant implications both for the ways of using legal knowledge and for the structure of such knowledge. On the one hand, deductive inferences are complemented with defeasible arguments, supporting their conclusion only to the extent that they are not validly attacked (rebutted or undercut) by further arguments. On the other hand, a conflict-free set of legal axioms is substituted with an argumentation framework containing both conflicting pieces information and the criteria for solving such conflicts. It is important to stress the difference between an argumentation framework and a deductive axiomatic base. While a deductive axiomatic base is consistent and flat, an argumentative framework is conflictual and hierarchical: it includes reasons clash one against the other, reasons for preferring certain reason to the certain other, reasons for applying or not not applying certain reasons given certain conditions.

Both strategies just mentioned, namely, representing law as an axiomatic base or representing it as an argumentation framework may be justified, in different context. The first strategy may be appropriate

⁴⁶C. E. Alchourrón, "On Law and Logic" (1996) 9 *Ratio Juris*, 331–48.

⁴⁷C. G. Hempel, *Philosophy of Natural Sciences* (Englewood Cliffs, N. J., Prentice-Hall, 1966).

⁴⁸Gordon, "The Importance of Nonmonotonicity for Legal Reasoning".

⁴⁹For a classical introduction to logic programming, see R. A. Kowalski, *Logic for Problem Solving* (New York, N. Y., North Holland, 1979).

when we want to deepen our analysis of small set of norms and anticipate as much as possible all possible instances of their application, finding a precise solution for each of them. The second strategy, however, corresponds more directly to the logical structure of non-formalised legal language (which express the law indicating rules and exceptions, principles, preference criteria, etc.), and reflects the ways of legal reasoning, when it applies to its peculiar contents: rules and exceptions, different values to balance, different norms implementing different values, standards indicating what norms and values ought to prevail in the case of conflict, etc.

We may transform an argumentation framework into an axiomatic knowledge base whose deductive conclusions include all outcomes which would be defeasibly justified given the argumentation framework (assuming that all the facts of the cases were completely known). The dialectical interaction between reasons for and against certain conclusions, and of the grounds for preferring one reason to the other would be transformed into the description of conclusive connections between legal preconditions and legal consequences (the connection which hold on the basis of the given argumentation framework and of the arguments is enables). Flattening legal information in this way, however, entails a loss of information: the deductive knowledge base does not include memory of the choices from which it derives, and therefore it does not contain the information needed to reconsider such choices, for instance, for the purpose of establishing new interpretative choices. We need to go back to the original argumentative framework when we want to understand the articulation of legal reasoning, and expand or revise its premises on the basis of new information.

Consider for instance the domain of privacy. Here we have (at least under the EU regulation), the idea that processing personal data is only admissible for a specific purpose, communicated to the concerned person. Moreover, in general it is only admissible when there is consent by that person. This constraint are justified by the need of protecting values such as individual self-determination and dignity. However, there is a large set of exceptions to the consent principle, namely different hypotheses when data can be processed also without such consent. These exceptions are justified by the need of protecting competing rights of others, as well as certain social goals. Moreover, we have cases where consent alone is insufficient to make data processing permissible, further requirements been necessary (such as, for genetic data, the authorisation of the privacy authority), and for each such exceptions specific rationales can be found, which guide in determining the contents and limits of such exceptions. Finally there may be cases where personal data may be processable, even beyond the explicitly stated legislative hypotheses, since there is an authorisation by that data protection authority, conceded since access to data was required for protecting rights of others, prevailing upon the right to privacy. To determine whether the privacy 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 which respects legal (and constitutional constraints). We could try to reduce this multi-levelled argumentative framework to a set of flat rules, but we would obtain a representation that is removed from the original text, and whose contents and rationales are much more difficult to grasp.

19 Conclusion

I hope to have convinced the reader that defeasible reasoning is an essential aspect of legal problem-solving, and that the logical structures enabling such reasoning are an essential feature of legal knowledge. There is one last issue to be addressed, to which I will devote some concluding remarks. Can we give to defeasible reasoning a rigorous logical form, so that we can check precisely whether a certain conclusion is defeasibly derivable from a defeasible knowledge base?

The answer, I believe, is positive. Various logical approaches have been devised for modelling defeasible reasoning—default logic, autoepistemic logic, circumscription, preferential logics, metalogics—and they can be shown to converge towards the intuitive understanding of defeasible reasoning.⁵⁰ In the legal

⁵⁰For an analysis of the merits different approaches and of their application to the legal domain H. Prakken, *Logical Tools for Modelling Legal Argument: A Study of Defeasible Reasoning in Law* (Dordrecht, Kluwer, 1997).

domain argument-based models of defeasible reasoning—according to which defeasibility results from the interaction of conflicting arguments—have attracted most attention. Among such models I take the liberty of mentioning the one I have developed together with Henry Prakken,⁵¹ whose basic ideas are common to other argument-based approaches to legal defeasibility.⁵² In this model legal rules and principles are viewed as reasoning warrants that, when certain antecedent reasons are available (being provided as part of the input, or as the result of previous inference steps), support the derivation of legal conclusions. In this way basic legal arguments are obtained, which consist in chains of warrants leading to legal conclusions. Such arguments may be attacked (rebutted or undercut) by further arguments. When the attack succeeds, we say that the attacked argument is defeated: this happens when the attacked argument is rebutted by a stronger argument, or when it is undercut. For determining the relative strength of the arguments at issue, we have to appeal to further arguments, telling us which one of the contradictory arguments is to be preferred, on what grounds.

As a first approximation, we can then say that the conclusion of the defeated argument needs to be retracted. However, the picture is complicated by the idea of reinstatement: defeated arguments can be recovered if their defeaters are, in their turn, contradicted by stronger arguments (or undercut). So, we obtain a model of legal reasoning as the dialectical interaction of competing inferences: the outcome of this competition determines what conclusions will be legally justified in the framework of the available legal knowledge. We are consequently able to distinguish what conclusions are warranted, namely, justified with regard to a defeasible knowledge base (having no justified or defeasible attacker), what conclusions are overruled (being strictly defeated by a justified argument) and what conclusions are merely defensible (being attacked by an argument which is neither justified or overruled).⁵³

This account can be precisely stated in a very intuitive way in the dialectical form of a dispute, where an proponent (the party who advances the initial thesis, namely, the conclusion of the initial argument) interacts with an opponent (a critic), who contests that argument (who advances the even-level arguments). Both parties use in the best way the available knowledge, using argument that contrast the argument of the counterparty. The proponent wins if he has always the last word, proposing bottom argument against which the opponent cannot replicate. The opponent wins if she finds a criticism to which the proponent is unable to reply. Thus, the idea is that a statement is warranted (relative to a certain body of information) if its proponent can successfully defend it in a dispute, or ‘argument game’ with an opponent. The reasoner who checks on his own whether a statement is warranted should take both positions, both as the proponent of the thesis and as its critical opponent.⁵⁴

I cannot provide here a deeper analysis of this logical model. What matters for our purposes is that accepting defeasible reasoning in the legal domain does not entail abandoning logical rigour. On the contrary, it means to provide logical models which match to a greater extent the structures of legal knowledge,

⁵¹vedi H. Prakken and G. Sartor, “Argument-based Extended Logic Programming with Defeasible Priorities” (1997) 7 *Journal of Applied Non-classical Logics*, 25–75 and Sartor, *Legal Reasoning: A Cognitive Approach to the Law*.

⁵²See for instance, T. F. Gordon, *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice* (Dordrecht, Kluwer, 1995); , J. C. Hage, *Reasoning with Rules: An Essay on Legal Reasoning and Its Underlying Logic* (Dordrecht, Kluwer, 1997), T. J. M. Bench-Capon, T. Geldard, and P. Leng, “A Method for the Computational Modelling of Dialectical Argument with Dialogue Games” (2000) 8 *Artificial Intelligence and Law*, 233–54, D. N. Walton, *Argumentation Methods for Artificial Intelligence in Law* (Berlin, Springer, 2005), T. F. Gordon, “Constructing Arguments with a Computational Model of an Argumentation Scheme for Legal Rules”, in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law (ICAIL-2007)* (ACM, 2007), 117–21.

⁵³Recently, the model just mentioned has been extended in such a way as to model also the dialectical allocation of the burden of proof: the party which is shown to be burdened can validly attack an argument of the other party only by providing arguments which are stronger than the argument of the other party, up to the level required by the applicable proof standard. See (Prakken and Sartor, “Formalising Arguments about the Burden of Persuasion”).

⁵⁴One form that such a game can take is the one of H. Prakken and G. Sartor, eds., “Logical Models of Legal Argumentation (Special Issue)” (1996) 5 *Artificial Intelligence and Law*, 157–372. Several games are possible, reflecting different notions of defeasible consequence; cf. e.g. H. Prakken and G. A. W. Vreeswijk, “Logical Systems for Defeasible Argumentation”, in *Handbook of Philosophical Logic*, ed. by D. Gabbay and F. Günthner (Dordrecht, Kluwer, 2002), 218–319. For present purposes their differences do not matter.

the patterns of legal reasoning and the dialectics of legal interaction.

References

- Sartor, G., *Legal Reasoning: A Cognitive Approach to the Law*, vol. 5, *Treatise on Legal Philosophy and General Jurisprudence* (Berlin, Springer, 2005).
- Pollock, J. L., *Cognitive Carpentry: A Blueprint for How to Build a Person* (New York, N. Y., MIT Press, 1995).
- Pollock, J. L. and J. Cruz, *Contemporary Theories of Knowledge* (Totowa, N. Y., Rowman and Littlefield, 1999).
- Raz, J., *Practical Reason and Norms* (London, Hutchinson, 1975).
- Fodor, J., *The Modularity of Mind* (Cambridge, Mass., MIT Press, 1983).
- Wróblewski, J., “Legal Syllogism and Rationality of Judicial Decision” (1974) 5 *Rechtstheorie*, 33–45.
- Sainsbury, M., *Logical Forms: An Introduction to Philosophical Logic* (Oxford, Blackwell, 2001).
- Hodges, W., “Elementary Predicate Logic”, in *Handbook of Philosophical Logic. Volume I: Elements of classical logic*, ed. by D. Gabbay and F. Günthner (Dordrecht, Kluwer, 1983), 1–131.
- Ginzberg, M. L., ed., *Readings in Nonmonotonic Reasoning* (Los Altos, Cal., Morgan Kaufmann, 1987).
- Brewka, G., *Nonmonotonic Reasoning: Logical Foundations of Commonsense* (Cambridge, Cambridge University Press, 1991).
- Peczenik, A., “Scientia Juris”, in *Treatise of Legal Philosophy and General Jurisprudence - Volume 2* (Berlin, Springer, 2006).
- Buridan, J., *Quaestiones super decem libros Ethicorum Aristotelis ad Nicomachum* (Frankfurt am Main, Minerva, 1968).
- Pollock, J. L., “Perceiving and Reasoning about a Changing World” (1998) 14 *Computational Intelligence*, 489–562.
- Popper, K. R., *The Logic of Scientific Discovery* (London, Hutchinson, 1959).
- Alchourrón, C. E., P. Gärdenfors, and D. Makinson, “On the Logic of Theory Change: Partial Meet Functions for Contractions and Revisions” (1985) 50 *Journal of Symbolic Logic*, 510–30.
- Gärdenfors, P., *Knowledge in Flux* (Cambridge, Mass., MIT Press, 1987).
- Perelman, C. and L. Olbrechts-Tyteca, *Traité de l’argumentation: la nouvelle rhétorique*, 5th ed. (Brussels, Éditions de l’université de Bruxelles, 1988).
- Moore, R. C., “Possible-World Semantics for Autoepistemic Logic”, in *Readings in Nonmonotonic Reasoning*, ed. by M. L. Ginsberg (Los Altos, Cal., Morgan Kaufmann, 1987), 137–42.
- Clark, K. L., “Negation as Failure”, in *Readings in Nonmonotonic Reasoning*, ed. by M. L. Ginsberg (Los Altos, Cal., Morgan Kaufmann, 1987), 311–25.
- Sartor, G., “Defeasibility in Legal Reasoning”, in *Informatics and the Foundations of Legal Reasoning*, ed. by Z. Bankowski, I. White, and U. Hahn (Dordrecht, Kluwer Academic, 1995), 119–57.
- Russell, S. J. and P. Norvig, *Artificial Intelligence. A Modern Approach*, 2nd ed. (Englewood Cliffs, N. J., Prentice Hall, 2003).
- Shafer, G. and J. Pearl, eds., *Readings in Uncertain Reasoning* (Los Altos, Cal., Morgan Kaufmann, 1990).
- Luhmann, N., *Rechtssystem und Rechtsdogmatik* (Stuttgart, Kohlhammer, 1974).
- Haack, S., “Truth and Justice, Inquiry and Advocacy, Science and Law” (2003) 7 *Associations*, 103–14.
- *Defending Science within Reason* (Amherst, N. Y., Prometheus, 2003).
- Pollock, J. L., *Knowledge and Justification* (Princeton, N. J., Princeton University Press, 1974).
- Chisholm, R. M., *Theory of Knowledge*, 2nd ed. (Englewood Cliffs, N. J., Prentice Hall, 1977).
- Gordon, T. F., “The Importance of Nonmonotonicity for Legal Reasoning”, in *Expert Systems in Law: Impacts on Legal Theory and Computer Law*, ed. by H. Fiedler, F. Haft, and R. Traunmüller (Tübingen, Attempto, 1988), 111–26.
- MacCormick, D. N., “Defeasibility in Law and Logic”, in *Informatics and the Foundations of Legal Reasoning*, ed. by Z. Bankowski, I. White, and U. Hahn (Dordrecht, Kluwer Academic, 1995), 99–117.
- Aristotle, *Nicomachean Ethics* (Oxford, Oxford University Press, 1954).
- Aquinas, T., *Summa Theologiae* (Allen, Texas, Benzinger Bros, 1947).
- Ross, W. D., *The Right and the Good* (Oxford, Clarendon, 1930).
- *Foundations of Ethics* (Oxford, Clarendon, 1939).
- Hart, H. L. A., “The Ascription of Responsibility and Rights”, in *Logic and Language*, ed. by A. Flew (Oxford, Blackwell, 1951), 145–66.
- Prakken, H. and G. Sartor, “Formalising Arguments about the Burden of Persuasion”, in *Proceedings of the 11th international conference on Artificial intelligence and law (ICAIL-2007)* (New York, N. Y., ACM, 2007).

BIBLIOGRAPHY

- “More on Presumptions and Burdens of Proof”, in *Proceedings of JURIX 2008*, ed. by E. Francesconi, G. Sartor, and D. Tiscornia (Amsterdam, IOS, 2008).
- Alexy, R., *A Theory of Legal Argumentation: The Theory of Rational Discourse as Theory of Legal Justification* (Oxford, Clarendon, 1989).
- Peczenik, A., *On Law and Reason* (Dordrecht, Kluwer, 1989).
- Alexy, R., *Begriff und Geltung des Rechts* (Freiburg, Steiner, 1992).
- Perelman, C. and L. Olbrechts-Tyteca, *The New Rhetoric: A Treaty on Argumentation*. (Notre Dame, Ind., University of Notre Dame Press, 1969).
- Aristotle, *Topics* (New York, N. Y., Random House, 1941).
- Günther, K., *The Sense of Appropriateness* (Albany, N. Y., State University of New York Press, 1993).
- “A Normative Conception of Coherence for a Discursive Theory of Legal Justification” (1989) 2 *Ratio Juris*, 155–66.
- Mill, J. S., “On Liberty”, in *On Liberty and Other Essays*, ed. by J. Gray (Oxford, Oxford University Press, 1991), 5–130.
- Rescher, N., *Dialectics: A Controversy-oriented Approach to the Theory of Knowledge* (Albany, N. Y., State University of New York Press, 1977).
- Alchourrón, C. E. and E. Bulygin, *Normative Systems* (Vienna, Springer, 1971).
- Alchourrón, C. E., “On Law and Logic” (1996) 9 *Ratio Juris*, 331–48.
- Hempel, C. G., *Philosophy of Natural Sciences* (Englewood Cliffs, N. J., Prentice-Hall, 1966).
- Kowalski, R. A., *Logic for Problem Solving* (New York, N. Y., North Holland, 1979).
- Prakken, H., *Logical Tools for Modelling Legal Argument: A Study of Defeasible Reasoning in Law* (Dordrecht, Kluwer, 1997).
- Prakken, H. and G. Sartor, “Argument-based Extended Logic Programming with Defeasible Priorities” (1997) 7 *Journal of Applied Non-classical Logics*, 25–75.
- Gordon, T. F., *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice* (Dordrecht, Kluwer, 1995).
- Hage, J. C., *Reasoning with Rules: An Essay on Legal Reasoning and Its Underlying Logic* (Dordrecht, Kluwer, 1997).
- Bench-Capon, T. J. M., T. Geldard, and P. Leng, “A Method for the Computational Modelling of Dialectical Argument with Dialogue Games” (2000) 8 *Artificial Intelligence and Law*, 233–54.
- Walton, D. N., *Argumentation Methods for Artificial Intelligence in Law* (Berlin, Springer, 2005).
- Gordon, T. F., “Constructing Arguments with a Computational Model of an Argumentation Scheme for Legal Rules”, in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law (ICAIL-2007)* (ACM, 2007), 117–21.
- Prakken, H. and G. Sartor, eds., “Logical Models of Legal Argumentation (Special Issue)” (1996) 5 *Artificial Intelligence and Law*, 157–372.
- Prakken, H. and G. A. W. Vreeswijk, “Logical Systems for Defeasible Argumentation”, in *Handbook of Philosophical Logic*, ed. by D. Gabbay and F. Günthner (Dordrecht, Kluwer, 2002), 218–319.