PDF version of the entry
Defeasible Reasoning
https://plato.stanford.edu/archives/sum2022/entries/reasoning-defeasible/

from the Summer 2022 Edition of the

Stanford Encyclopedia of Philosophy

Library of Congress ISSN: 1095-5054

Defeasible Reasoning

First published Fri Jan 21, 2005; substantive revision Fri Jun 25, 2021

Reasoning is *defeasible* when the corresponding argument is rationally compelling but not deductively valid. The truth of the premises of a good defeasible argument provide support for the conclusion, even though it is possible for the premises to be true and the conclusion false. In other words, the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information. Philosophers have studied the nature of defeasible reasoning since Aristotle's analysis of *dialectical reasoning* in the *Topics* and the *Posterior Analytics*, but the subject has been studied with unique intensity over the last forty years, largely due to the interest it attracted from the artificial intelligence movement in computer science. There have been two approaches to the study of reasoning: treating it either as a branch of epistemology (the study of knowledge) or as a branch of logic. In recent work, the term *defeasible reasoning* has typically been limited to inferences involving rough-and-ready, exception-permitting generalizations, that is, inferring what has or will happen on the basis of what *normally* happens. This narrower sense of *defeasible reasoning*, which will be the subject of this article, excludes from the topic the study of other forms of non-deductive reasoning, including inference to the best explanation, abduction, analogical reasoning, and scientific induction. This exclusion is to some extent artificial, but it reflects the fact that the formal study of these other forms of non-deductive reasoning remains quite rudimentary.

1. History

Defeasible reasoning has been the subject of study by both philosophers and computer scientists (especially those involved in the field of artificial intelligence). The philosophical history of the subject goes back to Aristotle, while the field of artificial intelligence has greatly intensified interest in it over the last forty years.

1.1 Philosophy

According to Aristotle, deductive logic (especially in the form of the syllogism) plays a central role in the articulation of scientific understanding, deducing observable phenomena from definitions of natures that hold universally and without exception. However, in the practical matters of everyday life, we rely upon generalizations that hold only "for the most part", under normal circumstances, and the application of such common sense generalizations involves merely *dialectical* reasoning, reasoning that is defeasible and falls short of deductive validity. Aristotle lays out a large number and great variety of examples of such reasoning in his work entitled the *Topics*.

Investigations in logic after Aristotle (from later antiquity through the twentieth century) seem to have focused exclusively on deductive logic. This continued to be true as the predicate logic was developed by Peirce, Frege, Russell, Whitehead, and others in the late nineteenth and early twentieth centuries. With the collapse of logical positivism in the mid-twentieth century (and the abandonment of attempts to treat the physical world as a logical construction from facts about sense data), new attention was given to the relationship between sense perception and the external world. Roderick Chisholm (Chisholm 1957; Chisholm 1966) argued that sensory appearances give good, but defeasible, reasons for believing in corresponding facts about the physical world. If I am "appeared to redly" (have the sensory experience as of being in the presence of something red), then, Chisholm argued, I may presume that I really am in the presence of something red. This presumption can, of course, be defeated, if, for example, I learn that my environment is relevantly abnormal (for instance, all the ambient light is red).

H. L. A. Hart (Hart 1951), at a 1949 meeting of the Aristotelian Society, noted the centrality of defeasible reasoning in the law, especially within the Anglo-American common law tradition. Hart pointed out that judges must take into account exceptional circumstances in which a legal principle cannot be applied at all or must be applied in a weakened form. Hart refers explicitly to conditions that can *defeat* (Hart 1951, p. 175) the claim that a contract exists, even when the standard definition of 'contract' is satisfied. A defeasible logic is needed because a judge is required to make a judgment on the basis of an incomplete set of facts: those facts that are presented to the judge by the two parties as germane to the claim.

The idea of defeasibility also showed up in work on the formal theory of argumentation, including Stephen Toulmin's *The Uses of Argument* (Toulmin 1964). Toulmin, building on Hart's observations, argues for the importance of distinguishing between warrants and rebuttals (Toulmin 1964, 101ff, 143ff). The formal theory of argumentation (van Eemeren et al. 2020; Prakken and Vreeswijk 2002) has proved a fruitful ground for the development of models of defeasible reasoning.

John L. Pollock developed Chisholm's idea into a theory of *prima facie reasons* and *defeaters* of those reasons (Pollock 1967, 1970, 1974, 1987, 1995, 2010). Pollock distinguished between two kinds of defeaters of a defeasible inference: *rebutting defeaters* (which give one a prima facie reason for believing the denial of the original conclusion) and *undercutting defeaters* (which give one a reason for doubting that the usual relationship between the premises and the conclusion hold in the given case). According to Pollock, a conclusion is warranted, given all of one's evidence, if it is supported by an ultimately undefeated argument whose premises are drawn from that evidence.

1.2 Artificial Intelligence

As the subdiscipline of artificial intelligence took shape in the 1960s, pioneers like John M. McCarthy and Patrick J. Hayes soon discovered the need to represent and implement the sort of defeasible reasoning that had been identified by Aristotle and Chisholm. McCarthy and Hayes (McCarthy and Hayes 1969) developed a formal language they called the "situation calculus," for use by expert systems attempting to model changes and interactions among a domain of objects and actors. McCarthy and Hayes encountered what they called the *frame problem*: the problem of deciding which conditions will *not* change in the wake of an event. They required a defeasible principle of inertia: the presumption that any given condition will not change, unless required to do so by actual events and dynamic laws. In addition, they encountered the *qualification problem*: the need for a presumption that an action can be successfully performed, once a short list of essential prerequisites have been met. McCarthy (McCarthy 1977, 1038–1044) suggested that the solution lay in a logical principle of *circumscription*: the presumption that the actual situation is as unencumbered with abnormalities and oddities (including unexplained changes and unexpected interferences) as is consistent with our knowledge of it. (McCarthy 1982; McCarthy 1986) In effect, McCarthy suggests that it is warranted to believe whatever is true in all the *minimal* (or otherwise *preferred*) models of one's initial information set.

In the early 1980s, several systems of defeasible reasoning were proposed by others in the field of artificial intelligence: Ray Reiter's default logic (Reiter 1980; Etherington and Reiter 1983, 104–108), McDermott and Doyle's Non-Monotonic Logic I (McDermott and Doyle, 1982), Robert C. Moore's Autoepistemic Logic (Moore 1985), and Hector Levesque's formalization of the "all I know" operator (Levesque 1990). These early proposals involved the search for a kind of *fixed point* or cognitive equilibrium. Special rules (called *default rules* by Reiter) permit drawing certain conclusions so long as these conclusions are consistent with what one knows, including all that one knows on the basis of these very default rules. In some cases, no such fixed point exists, and, in others, there are multiple, mutually inconsistent fixed points. In addition, these systems were procedural or computational in nature, in contrast to the semantic characterization of warranted conclusions (in terms of preferred models) in McCarthy's circumscription system. Later work in artificial intelligence has tended to follow McCarthy's lead in this respect.

2. Applications and Motivation

Philosophers and theorists of artificial intelligence have found a wide variety of applications for defeasible reasoning. In some cases, the defeasibility seems to be grounded in some aspect of the subject or the context of communication, and in other cases in facts about the objective world. The first includes defeasible rules as communicative or representational conventions and *autoepistemic* (reasoning about one's own knowledge and lack of knowledge). The latter, the objective sources of defeasibility, include defeasible obligations, defeasible laws of nature, induction, abduction, and Ockham's razor (the presumption that the world is as uncomplicated as possible).

2.1 Defeasibility as a Convention of Communication

Much of John McCarthy's early work in artificial intelligence concerned the interpretation of stories and puzzles (McCarthy and Hayes 1969; McCarthy 1977). McCarthy found that we often make assumptions based on what is not said. So, for example, in a puzzle about safely crossing a river by canoe, we assume that there are no bridges or other means of conveyance available. Similarly, when using a database to store and convey information, the information that, for example, no flight is scheduled at a certain time is represented simply by *not* listing such a flight. Inferences based on these conventions are defeasible, however, because the conventions can themselves be explicitly abrogated or suspended.

Nicholas Asher and his collaborators (Lascarides and Asher 1993, Asher and Lascarides 2003, Vieu, Bras, Asher, and Aurnague 2005, Txurruka and Asher 2008) have argued that defeasible reasoning is useful in unpacking the pragmatics of conversational implicature.

2.2 Autoepistemic Reasoning

Robert C. Moore (Moore 1985) pointed out that we sometimes infer things about the world based on our *not* knowing certain things. So, for instance, I might infer that I do not have a sister, since, if I did, I would certainly know it, and I do not in fact know that I have a sister. Such an inference is, of course, defeasible, since if I subsequently learn that I have a sister after all, the basis for the original inference is nullified.

2.3 Semantics for Generics and the Progressive

Generic terms (like *birds* in *Birds fly*) are expressed in English by means of bare common noun phrases (without determiner). Adverbs like *normally* and *typically* are also indicators of generic predication. As Asher and Pelletier (Asher and Pelletier 1997) have argued, the semantics for such sentences seems to involve intentionality: a generic sentence can be true even if the majority of the kind, or even all of the kind, fail to conform to the generalization. It can be true that birds fly even if, as a result of a freakish accident, all surviving birds are abnormally flightless. A promising semantic theory for the generic is to represent generic predication by means of a defeasible rule or conditional.

The progressive verb involves a similar kind of intentionality. (Asher 1992) If Jones *is crossing the street*, then it would normally be the case that Jones *will succeed* in crossing the street. However, this inference is clearly defeasible: Jones might be hit by a truck midway across and never complete the crossing.

2.4 Defeasible Reasons

Jonathan Dancy (Dancy 1993, 2004) has developed and defended an anti-Humean conception of practical reasoning, according to which it is the facts themselves, and not our desires, aversions, or other attitudes towards those facts, that constitute *reasons for acting*. These facts consist of particulars' having properties, and those properties provide in each such case some reason for acting--as, for example, someone's need can provide a reason for meeting that need. However, each general property can provide a reason only defeasibly: not only can a reason be overwhelmed by contrary considerations, but a property's valence for action can be completely neutralized or even reversed by further considerations. For example, even if giving pleasure is in general a reason in favor of acting in

a certain way, the fact that some action would give pleasure to those pleased by the suffering of others is a reason *against* and not for so acting. Dancy has introduced (in Dancy 2004) the concepts of *intensifiers* and *attenuators*, applying to facts that strengthen or weaken the force of reasons. In the extreme case, a fact can *disable* a reason altogether, corresponding to what Joseph Raz had described as an *exclusionary reason* (Raz 1975), and to John Pollock's idea of an undercutting defeater.

To the extent that our practical reasoning is guided at all by general rules or principles (something that Dancy explicitly denies), the reasoning must be defeasible, as John Horty has argued (Horty 2007b). From this perspective, Dancy's thesis of *moral particularism* corresponds to the potential defeasibility of all general reasons (see Lance and Little 2004, 2007). Defeasible logic can enable general rules to play an indispensable role despite the *reasons holism* that Dancy has uncovered.

In addition, defeasible reasoning can be used to illuminate moral and legal dilemmas, cases in which general rules come into conflict (see Horty 1994, 2003). This can be done without attributing logical inconsistency to the conflicting rules and without treating the conflict as merely apparent, i.e., as due to an incomplete representation of the rules.

2.5 Defeasible Obligations

Philosophers have, for quite some time, been interested in defeasible obligations, which give rise to defeasible inferences about what we are, all things considered, obliged to do. David Ross, in 1930, discussed the phenomena of *prima facie* obligations (Ross 1930, 1939). The existence of a prima facie obligation gives one good, but defeasible grounds, for believing that one ought to fulfill that obligation. When formal *deontic logic* was developed by Chisholm and others in the 1960s (Chisholm 1963), the use of classical logic gave rise to certain paradoxes, such as

Chisholm's paradox of contrary-to-duty imperatives. These paradoxes can be resolved by recognizing that the inference from imperative to actual duty is a defeasible one (Asher and Bonevac 1996; Nute 1997).

Such defeasible obligations can also appear in the domain of law: see Prakken and Sartor 1995 and 1996.

2.6 Defeasible Laws of Nature and Scientific Programs

Philosophers David M. Armstrong and Nancy Cartwright have argued that the actual laws of nature are *oaken* rather than *iron* (to use Armstrong's terms). (Armstrong 1983; Armstrong 1997, 230–231; Cartwright 1983). Oaken laws admit of exceptions: they have tacit *ceteris paribus* (other things being equal) or *ceteris absentibus* (other things being absent) conditions. As Cartwright points out, an inference based on such a law of nature is always defeasible, since we may discover that additional *phenomenological factors* must be added to the law in question in special cases.

There are several reasons to think that deductive logic is not an adequate tool for dealing with this phenomenon. In order to apply deduction to the laws and the initial conditions, the laws must be represented in a form that admits of no exceptions. This would require explicitly stating each potentially relevant condition in the antecedent of each law-stating conditional. This is impractical, not only because it makes the statement of each and every law extremely cumbersome, but also because we know that there are many exceptional cases that we have not yet encountered and may not be able to imagine. Defeasible laws enable us to express what we really know to be the case, rather than forcing us to pretend that we can make an exhaustive list of all the possible exceptions.

More recently, Tohmé, Delrieux, and Bueno (2011) have argued that defeasible reasoning is crucial to the understanding of scientific research programs.

2.7 Defeasible Principles in Metaphysics and Epistemology

Many classical philosophical arguments, especially those in the perennial philosophy that endured from Plato and Aristotle to the end of scholasticism, can be fruitfully reconstructed by means of defeasible logic. Metaphysical principles, like the laws of nature, may hold in normal cases, while admitting of occasional exceptions. The principle of causality, for example, that plays a central role in the cosmological argument for God's existence, can plausibly construed as a defeasible generalization (Koons 2001).

As discussed above (in section 1.1), prima facie reasons and defeaters of those reasons play a central role in contemporary epistemology, not only in relation to perceptual knowledge, but also in relation to every other source of knowledge: memory, imagination (as an indicator of possibility) and testimony, at the very least. In each cases, an impression or appearance provides good but defeasible evidence of a corresponding reality.

2.8 Occam's Razor and the Assumption of a "Closed World"

Prediction always involves an element of defeasibility. If one predicts what will, or what would, under some hypothesis, happen, one must presume that there are no unknown factors that might interfere with those factors and conditions that are known. Any prediction can be upset by such unanticipated interventions. Prediction thus proceeds from the assumption that the situation as modeled constitutes a *closed world*: that nothing outside that situation could intrude in time to upset one's predictions. In addition, we seem to presume that any factor that is not known to be causally relevant is in fact causally irrelevant, since we are constantly encountering new factors and novel combinations of factors, and it is impossible to verify their causal irrelevance in advance. This closed-world assumption is one of the principal motivations for McCarthy's logic of circumscription (McCarthy 1982; McCarthy 1986).

3. Varieties of Approaches

We can treat the study of defeasible reasoning either (i) as a branch of epistemology (the theory of knowledge), or (ii) as a branch of logic. In the epistemological approach, defeasible reasoning can be studied as a form of inference, that is, as a process by which we add to our stock of knowledge. Alternatively, we could treat *defeat* as a relation between arguments in a disputational discourse. In either version, the epistemological approach is concerned with the obtaining, maintaining, and transmission of *warrant*, with the question of when an inference, starting with justified or warranted beliefs, produces a new belief that is also warranted, given potential defeaters. This approach focuses explicitly on the norms of belief persistence and change.

In contrast, a logical approach to defeasible reasoning fastens on a relationship between propositions or possible bodies of information. Just as deductive logic consists of the study of a certain *consequence relation* between propositions or sets of propositions (the relation of valid implication), so defeasible (or *nonmonotonic*) logic consists of the study of a different kind of consequence relation. Deductive consequence is monotonic: if a set of premises logically entails a conclusion, than any superset (any set of premises that includes all of the first set) will also entail that some conclusion. In contrast, defeasible consequence is nonmonotonic. A conclusion follows defeasibly or nonmonotonically from a set of premises just in case it is true in *nearly all* of the models that verify the premises, or in the *most normal* models that do.

The two approaches are related. In particular, a logical theory of defeasible consequence will have epistemological consequences. It is presumably true that an ideally rational thinker will have a set of beliefs that are closed under defeasible, as well as deductive, consequence. However, a logical theory of defeasible consequence would have a wider scope of application than a merely epistemological theory of inference. Defeasible logic would provide a mechanism for engaging in *hypothetical* reasoning, not just reasoning from actual beliefs.

Conversely, as David Makinson and Peter Gärdenfors have pointed out (Makinson and Gärdenfors 1991, 185–205; Makinson 2005), an epistemological theory of belief change can be used to define a set of nonmonotonic consequence relations (one relation for each initial belief state). We can define the consequence relation $\alpha \mid\sim \beta$, for a given set of beliefs T, as holding just in case the result of adding belief α to T would include belief in β. However, on this approach, there would be many distinct nonmonotonic consequence relations, instead of a single perspective-independent one.

And, as Phan Minh Dung has argued (Dung 1995), formal argumentation can also be used to provide a basis for defining a nonmonotonic consequence relation. A formal argument structure F is an ordered pair $\langle A, B \rangle$, where A is a set of arguments, and B is a binary relation on A (the *attack* relation). Then we can say that an argument structure F has p has a consequence just in case p is the conclusion of some argument in the optimal *extension* of F (which can be defined in a variety of ways--see section 4.4).

4. Epistemological Approaches

There are have been four versions of the epistemological approach, each of which attempts to define how an cognitively ideal agent arrives at warranted conclusions, given an initial input. The first two of these, John L. Pollock's theory of defeasible reasoning and the theory of semantic inheritance networks, are explicitly computational in nature. They take as input a complex, structured state, representing the data available to the agent, and they define a procedure by which new conclusions can be warranted. The third approach, based on the theory of belief change (the AGM model) developed by Alchourrón, Gärdenfors, and Makinson (Alchourrón, Gärdenfors, and Makinson 1982), instead lays down a set of conditions that an ideal process of belief change ought to satisfy. The AGM model can be used to define a nonmonotonic consequence relation that is temporary and local. This can represent reasoning that is hypothetically or counterfactually defeasible, in the sense that what "follows" from a conjunctive proposition (p & q) need not be a superset of what "follows" from p alone. The fourth approach is that of formal argumentation theory, in which *defeat* is treated as a relation between *arguments* within a dialogue

4.1 Formal Epistemology

John Pollock's approach to defeasible reasoning consists of enumerating a set of rules that are constructive and effectively computable, and that aim at describing how an ideal cognitive agent builds up a rich set of beliefs, beginning with a relatively sparse data set (consisting of beliefs about immediate sensory appearances, apparent memories, and such things). The inferences involved are not, for the most part, deductive. Instead, Pollock defines, first, what it is for one belief to be a *prima facie reason* for believing another proposition. In addition, Pollock defines what it is for one belief, say in p, to be a *defeater* for q as a prima facie reason for r. In fact. Pollock distinguishes two kinds of defeaters: *rebutting defeaters*, which are themselves prima facie reasons for believing the negation of the conclusion, and *undercutting defeaters*, which provide a reason for doubting that q provides any support, in the actual circumstances, for r. (Pollock 1987, 484) A belief is *ultimately warranted* in relation to a data set (or *epistemic basis*) just in case it is supported by some ultimately undefeated argument proceeding from that epistemic basis.

In Pollock 1995, Pollock uses a directed graph to represent the structure of an ideal cognitive state. Each directed link in the network represents the first node's being a prima facie reason for the second. The new theory includes an account of *hypothetical*, as well as categorical reasoning, since each node of the graph includes a (possibly empty) set of hypotheses. Somewhat surprisingly, Pollock assumes a principle of monotonicity with respect to hypotheses: a belief that is warranted relative to a set of hypotheses is also warranted with respect to any superset of hypotheses. Pollock also permits conditionalization and reasoning by cases.

An argument is *self-defeating* if it supports a defeater for one of its own defeasible steps. Here is an interesting example: (1) Robert says that the elephant beside him looks pink. (2) Robert's color vision becomes unreliable in the presence of pink elephants. Ordinarily, belief 1 would support the conclusion that the elephant is pink, but this conclusion undercuts the argument, thanks to belief 2. Thus, the argument that the elephant is pink is self-defeating. Pollock argues that all self-defeating arguments should be rejected, and that they should not be allowed to defeat other arguments. In addition, a set of nodes can experience mutual destruction or *collective defeat* if each member of the set is defeated by some other member, and no member of the set is defeated by an undefeated node that is outside the set.

In formalizing the undercutting rebuttal, Pollock introduces a new connective, ⊗, where $p \otimes q$ means that it is not the case that p wouldn't be true unless q were true. Pollock uses rules, rather than conditional propositions, to express the prima facie relation. If he had, instead, introduced a special connective ⇒, with $p \Rightarrow q$ meaning that p would be a prima facie reason for q, then undercutting defeaters could be represented by means of negating this conditional. To express the fact that r is an undercutting defeater of p as a prima facie reason for q, we could state both that $(p \Rightarrow q)$ and $\neg((p \;\&\; r) \Rightarrow q)$.

In the case of conflicting prima facie reasons, Pollock rejects the principle of *specificity*, a widely accepted principle according to which the defeasible rule with the more specific antecedent takes priority over conflicting rules with less specific antecedents. Pollock does, however, accept a special case of specificity in the area of statistical syllogisms with projectible properties. (Pollock 1995, 64–66) So, if I know that most *A*s are *B*s, and the most *AC*s are not *B*s, then I should, upon learning that individual *b* is both *A* and *C*, give priority to the *AC* generalization over the *A* generalization (concluding that *b* is not a *B*).

Pollock's theory of warrant is intended to provide normative rules for belief, of the form: if you have warranted beliefs that are prima facie reasons for some further belief, and you have no ultimately undefeated defeaters for those reasons, then that further belief is warranted and should be believed. For more details of Pollock's theory, see the following supplementary document:

John Pollock's System

Wolfgang Spohn (Spohn 2002) has argued that Pollock's system is *normatively defective* because, in the end, Pollock has no normative standard to appeal to, other than ad hoc intuitions about how a reasonable

person would respond to this or that cognitive situation. Spohn suggests that, with respect to the state of development of the study of defeasible reasoning, Pollock's theory corresponds to C. I. Lewis's early investigations into modal logic. Lewis suggested a number of possible axiom systems, but lacked an adequate semantic theory that could provide an independent check on the correctness or completeness of any given list (of the kind that was later provided by Kripke and Kanger). Analogously, Spohn argues that Pollock's system is in need of a unifying normative standard. This very same criticism can be lodged, with equal justice, against a number of other theories of defeasible reasoning, including semantic inheritance networks and default logic.

4.2 Semantic Inheritance Networks

The system of semantic inheritance networks, developed by Horty, Thomason, and Touretzky (1990), is similar to Pollock's system. Both represent cognitive states by means of directed graphs, with links representing defeasible inferences. The semantic inheritance network theory has a intentionally narrower scope: the initial nodes of the network represent particular individuals, and all non-initial nodes represent kinds, categories or properties. A link from an initial (individual) node to a category node represents simply predication: that Felix (initial node) is a cat (category node), for example. Links between category nodes represent defeasible or generic inclusion: that birds (normally or usually) are flying things. To be more precise, there are both positive ("is a") and negative ("is not a") links. The negative links are usually represented by means of a slash through the body of the arrow.

Semantic inheritance networks differ from Pollock's system in two important ways. First, they cannot represent one fact's constituting an *undercutting* defeater of an inference, although they can represent *rebutting* defeaters. For example, they do not allow an inference from the apparent color of an elephant to its actual color to be undercut by the information that my color vision is unreliable, unless I have information about the actual color of the elephant that contradicts its apparent color. Secondly, they do incorporate the principle of specificity (the principle that rules with more specific antecedents take priority in case of conflict) into the very definition of a warranted conclusion. In fact, in contrast to Pollock, the semantic inheritance approach gives priority to rules whose antecedents are weakly or defeasibly more specific. That is, if the antecedent of one rule is defeasibly linked to the antecedent of a second rule, the first rule gains priority. For example, if Quakers are typically pacifists, then, when reasoning about a Quaker pacifist, rules pertaining to Quakers would override rules pertaining to pacifists. For the details of semantic inheritance theory, see the following supplementary document:

Semantic Inheritance Networks.

David Makinson (Makinson 1994) has pointed out that semantic network theory is very sensitive to the form in which defeasible information is represented. There is a great difference between having a direct link between two nodes and having a path between the two nodes being supported by the graph as a whole. The notion of preemption gives special powers to explicitly given premises over conclusions. Direct links always take priority over longer paths. Consequently, inheritance networks lack two desirable metalogical properties: cut and cautious monotony (which will be covered in more detail in the section on Logical Approaches).

- Cut: If G is a subgraph of G', and every link in G' corresponds to a path supported by G, then every path supported by G is also supported by G'.
- Cautious Monotony: If G is a subgraph of G', and every link in G' corresponds to a path supported by G, then every path supported by G' is also supported by G.

Cumulativity (Cut plus Cautious Monotony) corresponds to reasoning by lemmas or subconclusions. The Horty-Thomason-Touretzky system does satisfy special cases of Cut and Cautious Monotony: if A is an atomic statement (a link from an individual to a category), then if graph G supports A, then for any statement $B, G \cup \{A\}$ supports B if and only if G supports B.

Another form of inference that is not supported by semantic inheritance networks is that of reasoning by cases or by dilemma. In addition, semantic networks do not license modus-tollens-like inferences: from the fact that birds normally fly and Tweety does not fly, we are not licensed to infer that Tweety is not a bird. (This feature is also lacking in Pollock's system.)

4.3 Belief Revision Theory

Alchourrón, Gärdenfors, and Makinson (1982) developed a formal theory of belief revision and contraction, drawing largely on Willard van Orman Quine's model of the *web of belief* (Quine and Ullian 1970). The cognitive agent is modelled as believing a set of propositions that are ordered by their degree of entrenchment. This model provides the basis for a set of normative constraints on belief contraction (subtracting a belief) and belief revision (adding a new belief that is inconsistent with the original set). When a belief is added that is logically consistent with the original belief set, the agent is supposed to believe the logical closure of the original set plus the new belief. When a belief is added that is inconsistent with the original set, the agent retreats to the most entrenched of the maximal subsets of the set that are consistent with the new belief, adding the new proposition to that set and closing under logical consequence. For the axioms of the AGM model, see the following supplementary document:

AGM Postulates

AGM belief revision theory can be used as the basis for a system of defeasible reasoning or nonmonotonic logic, as Gärdenfors and Makinson have recognized (Makinson and Gärdenfors 1991). If K is an epistemic state, then a nonmonotonic consequence relation $\mid\!\sim$ can be defined as follows: $A \mid\!\sim B$ iff $B \in K * A$. Unlike Pollock's system or semantic inheritance networks, this defeasible consequence relation depends upon a background epistemic state. Thus, the belief revision approach gives rise, not to a single nonmonotonic consequence relation, but to family of relations. Each background state K gives rise to its own characteristic consequence relation.

One significant limitation of the belief-revision approach is that there is no representation in the object-language of a defeasible or default rule or conditional (that is, of a conditional of the form *If p, then normally q* or *That p would be a prima facie reason for accepting that q*). In fact, Gärdenfors (Gärdenfors 1978; Gärdenfors 1986) proved that no conditional satisfying the Ramsey test can be added to the AGM system without trivializing the revision relation.[1] (A conditional $\Rightarrow$ satisfies the Ramsey test just in case, for every epistemic state K, K includes $(A \Rightarrow B)$ iff $K * A$ includes B.)

Since the AGM system cannot include conditional beliefs, it cannot elucidate the question of what logical relationships hold between conditional defaults.

The lack of a representation of conditional beliefs is closely connected to another limitation of the AGM system: its inability to model repeated or *iterated* belief revision. The input to a belief change is an epistemic state, consisting both of a set of propositions believed and an entrenchment relation on that set. The output of an AGM revision, in contrast, consists simply of a set of beliefs. The system provides no guidance on the question of what would be the result of revising an epistemic state in two or more steps. If the entrenchment relation could be explicitly represented by means of conditional propositions, then it would be possible to define the new entrenchment relation that would result from a single belief revision, making iterated belief revision representable. A number of proposals along these lines have been made. The difficulty lies in defining exactly what would constitute a *minimal* change in the relative entrenchment or epistemic ranking of a set of beliefs. To this point, no clear consensus has emerged on this question. (See Spohn 1988; Nayak 1994; Wobcke 1995; Bochman 2001.)

On the larger question of the relation between belief revision and defeasible reasoning, there are two possibilities: that a theory of defeasible reasoning should be grounded in a theory of belief revision, and that a theory of belief revision should be grounded in a theory of defeasible reasoning. The second view has been defended by John Pollock (Pollock 1987; Pollock 1995) and by Hans Rott (Rott 1989). On this second view, we must make a sharp distinction between basic or foundational beliefs on the one hand and inferred or derived beliefs on the other. We can then model belief change on the assumption that new beliefs are added to the foundation (and are logically consistent with the existing set of those beliefs). Beliefs can be added which are inconsistent with previous inferred beliefs, and the new belief state consists simply in the closure of the new foundational set under the relation of defeasible consequence. On such an approach, default conditionals can be explicitly represented among the agent's beliefs. Gärdenfors's triviality result is then avoided by rejecting one of the assumptions of the theorem, *preservation*:

> **Preservation**:
> If $\neg A \notin K$,then $K \subseteq K * A$.

From the perspective that uses defeasible reasoning to define belief revision, there is no good reason to accept Preservation. One can add a belief that is consistent with what one already believes and thereby *lose* beliefs, since the new information might be an undercutting defeater to some defeasible inference that had been successful.

4.4 Formal Argumentation Theory

Phan Minh Dung (Dung 1995) initiated a new and fruitful approach to defeasible reasoning, one that focuses on the structure of arguments. Dung defines an *argument structure* as an ordered pair $\langle A, B\rangle$, in which A is a set of arguments and B is a binary relation on A, representing the *attacks* relation. In other words, if $\langle x, y\rangle \in B$, then argument x is representing as attacking argument y in some way. An argument is a sequence of propositions, with the last proposition designated as its conclusion. (The premises of an argument can be null, in which case we can treat the argument as equivalent to the assertion of the single proposition it contains.)

Dung's approach doesn't distinguish between the various ways in which one argument can attack another, such as rebutting, undermining, or undercutting, although this additional information can be added by differentiating several types of attack relations. One argument *rebuts* another when their conclusions are contradictories. An argument *undermines* a second argument when the conclusion of the first contradicts one of the premises of the second. And an argument *undercuts* another when its conclusion provides reason for doubting that the premises of the second are in the actual circumstances reliable indicators of the truth of the conclusion. In many applications, these distinctions can be ignored. However, Henry Prakken (Prakken 2010) makes use of all three forms of attack in his ASPIC+ system.

Central to Dung's approach is the idea of an *admissible* set of arguments relative to an argument structure. A set of arguments A is admissible if and only if it is conflict free (no argument in A attacks another argument in A), and there is every argument that attacks something in A is itself attacked by something in A. In other words, A can defeat everything that defeats one of its members. Dung's approach incorporates the principle: "the one who laughs last, laughs best."

A *preferred extension* of an argument structure is a maximal admissible set of the structure. Every structure possesses at least one preferred extension. A *stable extension* of a structure is a conflict-free set S that attacks each argument that does not belong to S. Every stable extension is a preferred extension, but not vice versa. Some structures do not have stable extensions. Leendert van der Torre and Srdjan Vesic (van der Torre and Vesic 2018) outline the full range of definitions of extensions, along with the principles of rationality they embody.

The characteristic function F_{AS} of an argument structure AF is defined as follows:

$$F_{AS}(S) = \{A\text{: } A \text{ is acceptable with respect to } S\}$$

The *grounded extension* of an argument structure is the least fixed point of its characteristic function. An extension S is *complete* if it contains every argument that is admissible with respect to S. The grounded extension is the minimal complete extension of a structure. If a structure is well-founded (with no infinite regress of attack relations, then the structure has a unique complete extension that is grounded, preferred, and stable (Dung 1995, 331).

These various notions of optimal extension can be used to define when a proposition has been proved or refuted in a particular structure, depending on whether the proposition or its negation belongs to the optimal extension of the structure.

Gerald Vreeswijk (Vreeswijk 1997) has built upon Dung's framework by introducing preference relations among arguments. The relation of relative conclusive force could be determined by such factors as the presence or absence of defeasible rules, the occurrence of a premise of one argument as the conclusion of another, the number of steps in the argument, or the use in the arguments of defeasible rules with varying degrees of reliability. The HERMES system of Karacapilidis and Papadias (Karacapilidis and Papadias 2001) implements numeric weights for reasons for and against a conclusion.

Henry Prakken (Prakken 2010) has expanded Dung's model by including support as well as attack relations between arguments. Bart Verheij's DefLog system (Verheij 2003, 2005) makes use of a conditional to represent support and a negation operator to represent attack. Anthony Hunter, Sylwia Polberg, and Matthias Thimm (Hunter et al. 2020) have recently used epistemic graphs to represent both positive and negative interactions among arguments.

Others have used probabilities to measure the comparative strength of arguments: Dung and Thang (2010), Verheij (2012), and Hunter (2013).

5. Logical Approaches

Logical approaches to defeasible reasoning treat the subject as a part of logic: the study of *nonmonotonic* consequence relations (in contrast to the monotonicity of classical logic). These relations are defined on propositions, not on the beliefs of an agent, so the focus is not on epistemology per se, although a theory of nonmonotonic logic will certainly have implications for epistemology.

5.1 Relations of Logical Consequence

A consequence relation is a mathematical relation that models what follows logically from what. Consequence relations can be defined in a variety of ways, such as Hilbert, Tarski, and Scott relations. A Hilbert consequence relation is a relation between pairs of formulas, a Tarski relation is a relation between sets of formulas (possibly infinite) and individual formulas, and a Scott relation is a relation between two sets of formulas. In the case of Hilbert and Tarski relations, $A \vDash B$ or $\Gamma \vDash B$ mean that the formula B follows from formula A or from set of formulas Γ. In the case of Scott consequence relations, $\Gamma \vDash \Delta$ means that the joint truth of all the members of Γ implies (in some sense) the truth of at least one member of Δ. To this point, studies of nonmonotonic logic have defined nonmonotonic consequence relations in the style of Hilbert or Tarski, rather than Scott.

A (Tarski) consequence relation is *monotonic* just in case it satisfies the following condition, for all formulas p and all sets Γ and Δ:

> **Monotonicity**:
> If $\Gamma \vDash p$, then $\Gamma \cup \Delta \vDash p$.

Any consequence relation that fails this condition is *nonmonotonic*. A relation of defeasible consequence clearly must be nonmonotonic, since a defeasible inference can be defeated by adding additional information that constitutes a rebutting or undercutting defeater.

5.2 Metalogical Desiderata

Once monotonicity is given up, the question arises: why call the relation of defeasible consequence a *logical consequence* relation at all? What properties do defeasible consequence and classical logical consequence have in common, that would justify treating them as sub-classes of the same category? What justifies calling nonmonotonic consequence *logical*?

To count as *logical*, there are certain minimal properties that a relation must satisfy. First, the relation ought to permit reasoning by lemmas or subconclusions. That is, if a proposition p already follows from a set Γ, then it should make no difference to add p to Γ as an additional premise. Relations that satisfy this condition are called *cumulative*. Cumulative relations satisfy the following two conditions (where "$C(\Gamma)$" represents the set of defeasible consequences of Γ):

> **Cut**:
> If $\Gamma \subseteq \Delta \subseteq C(\Gamma)$, then $C(\Delta) \subseteq C(\Gamma)$.

> **Cautious Monotony**:
> If $\Gamma \subseteq \Delta \subseteq C(\Gamma)$, then $C(\Gamma) \subseteq C(\Delta)$.

In addition, a defeasible consequence relation ought to be *supraclassical*: if p follows from q in classical logic, then it ought to be included in the defeasible consequences of q as well. A formula q ought to count as an (at least) defeasible consequence of itself, and anything included in the content of q (any formula p that follows from q in classical logic) ought to count as a defeasible consequence of q as well. Moreover, the defeasible consequences of a set Γ ought to depend only on the content of the formulas in Γ, not in how that content is represented. Consequently, the defeasible consequence relation ought to treat Γ and the classical logical closure of Γ (which we'll represent as "$Cn(\Gamma)$") in exactly the same way. A consequence relation that satisfies these two conditions is said to satisfy *full absorption* (see Makinson 1994, 47).

> **Full Absorption**:
> $Cn(C(\Gamma)) = C(\Gamma) = C(Cn(\Gamma))$

Finally, a genuinely logical consequence relation ought to enable us to reason by cases. So, it should satisfy a principle called distribution: if a formula p follows defeasibly from both q and r, then it ought to follow from their disjunction. (To require the converse principle would be to reinstate monotonicity.) The relevant principle is this:

> **Distribution**:
> $C(\Gamma) \cap C(\Delta) \subseteq C(Cn(\Gamma) \cap Cn(\Delta))$.

Consequence relations that are cumulative, strongly absorptive, and distributive satisfy a number of other desirable properties, including *conditionalization*: If a formula p is a defeasible consequence of $\Gamma \cup \{q\}$, then the material conditional $(q \rightarrow p)$ is a defeasible consequence of Γ alone. In addition, such logics satisfy the property of *loop*: if $p_1 \mid\!\sim p_2 \ldots p_{n-1} \mid\!\sim p_n$ (where " $\mid\!\sim$ " represents the defeasible consequence relation), then the defeasible consequences of p_i and p_j are exactly the same, for any i or j.[2]

There are three further conditions that have been much discussed in the literature, but whose status remains controversial: *disjunctive rationality*, *rational monotony*, and *consistency preservation*.

> **Disjunctive Rationality**:
> If $\Gamma \cup \{p\} \not\mid\!\sim r$, and $\Gamma \cup \{q\} \not\mid\!\sim r$, then $\Gamma \cup \{(\mathrm{p} \vee \mathrm{q})\} \not\mid\!\sim r$.

> **Rational Monotony**:
> If $\Gamma \mid\!\sim A$, then either $\Gamma \cup \{B\} \mid\!\sim A$ or $\Gamma \mid\!\sim \neg B$.

> **Consistency Preservation**:
> If Γ is classically consistent, then so is $C(\Gamma)$ (the set of defeasible consequences of Γ).

All three properties seem desirable, but they set a very hight standard for the defeasible reasoner.

5.3 Default Logic

Ray Reiter's default logic (Reiter 1980; Etherington and Reiter 1983) was part of the first generation of defeasible systems developed in the field of artificial intelligence. The relative ease of computing default extensions has made it one of the more popular systems.

Reiter's system is based on the use of *default rules*. A default rule consists of three formulas: the *prerequisite*, the *justification*, and the *consequent*. If one accepts the prerequisite of a default rule, and the justification is consistent with all one knows (including what one knows on the basis of the default rules themselves), then one is entitled to accept the consequent. The most popular use of default logic relies solely on *normal defaults*, in which the justification and the consequent are identical. Thus, a normal default of the form (p; $q \therefore q$) allows one to infer q from p, so long as q is consistent with one's endpoint (the *extension* of the default theory).

A default theory consists of a set of formulas (the facts), together with a set of default rules. An *extension* of a default theory is a fixed point of a particular inferential process: an extension E must be a consistent theory (a consistent set closed under classical consequence) that contains all of the facts of the default theory T, and, in addition, for each normal default ($p \Rightarrow q$), if p belongs to E, and q is consistent with E, then q must belong to E also.

Since the consequence relation is defined by a fixed-point condition, there are default theories that have no extension at all, and other theories that have multiple, mutually inconsistent extensions. For example, the theory consisting of the fact p and the pair of defaults (p ; (q & r) $\therefore q$) and (q ; $\neg r \therefore \neg r$) has no extension. If the first default is applied, then the second must be, and if the second default is not applied, the first must be. However, the conclusion of the second default contradicts the prerequisite of the first, so the first cannot be applied if the second is. There are many default theories that have multiple extensions. Consider the theory consisting of the facts q and r and the pair of defaults (q ; $p \therefore p$) and (r ; $\neg p \therefore \neg p$). One or the other, but not both, defaults must be applied.

Furthermore, there is no guarantee that if E and E' are both extensions of theory T, then the intersection of E and E' is also an extension (the intersection of two fixed points need not be itself a fixed point). Default logic is usually interpreted as a *credulous* system: as a system of logic that allows the reasoner to select *any* extension of the theory and believe all of the members of that theory, even though many of the resulting beliefs will involve propositions that are missing from other extensions (and may even be contradicted in some of those extensions).

Default logic fails many of the tests for a logical relation that were introduced in the previous section. It satisfied Cut and Full Absorption, but it fails Cautious Monotony (and thus fails to be cumulative). In addition, it fails Distribution, a serious limitation that rules out reasoning by cases. For example, if one knows that Smith is either Amish or Quaker, and both Quakers and Amish are normally pacifists, one cannot infer that Smith is a pacifist. Default logic also fails to represent Pollock's *undercutting defeaters*. Finally, default logic does not incorporate any form of the principle of *Specificity*, the principle that defaults with more specific prerequisites ought, in cases of conflict, to take priority over defaults with less specific prerequisites. Recently, John Horty (Horty 2007a, 2007b) has examined the implications of adding priorities among defaults (in the form of a partial ordering), which would permit the recognition of specificity and other grounds for preferring one default to another. In addition, Horty allows for defeasible reasoning about these priorities (the relative weights of various defaults) by means of higher-order default rules. Such defeasible reasoning about relative weights enables Horty to give an account of Pollock's undercutting defeaters: an undercutting defeater is a triggered default rule that lowers the weight of the undercut rule below some threshold, with the result that the undercut rule can no longer be triggered.

5.4 Nonmonotonic Logic I and Autoepistemic Logic

In both McDermott-Doyle's Nonmonotonic Logic I and Moore's Autoepistemic logic (McDermott and Doyle, 1982; Moore, 1985; Konolige 1994), a modal operator M (representing a kind of epistemic possibility) is used. Default rules take the following form: $((p \,\&\, Mq) \rightarrow q)$, that is, if p is true and q is "possible" (in the relevant sense), then q is also true. In both cases, the extension of a theory is defined, as in Reiter's default logic, by means of a fixed-point operation. Mp represents the fact that $\neg p$ does not belong to the extension. For example, in Moore's case, a set Δ is a *stable expansion* of a theory Γ just in case Δ is the set of classical consequences of the set $\Gamma \cup \{\neg Mp : p \in \Delta\} \cup \{Mp : p \notin \Delta\}$. As in the case of Reiter's default logic, some theories will lack a stable expansion, or have more than one. In addition, these systems fail to incorporate *Specificity*.

5.5 Circumscription

In circumscription (McCarthy 1982; McCarthy 1986; Lifschitz 1988), one or more predicates of the language are selected for minimization (there is, in addition, a further technical question of which predicates to treat as fixed and which to treat as variable). The nonmonotonic consequences of a theory T then consist of all the formulas that are true in every model of T that minimizes the extensions of the selected predicates. One model M of T is preferred to another, M', if and only if, for each designated predicate F, the extension of F in M is a subset of the extension of F in M', and, for some such predicate, the extension in M is a *proper subset* of the extension in M'.

The relation of circumscriptive consequence has all the desirable meta-logical properties. It is cumulative (satisfies Cut and Cautious Monotony), strongly absorptive, and distributive. In addition, it satisfies Consistency Preservation, although not Rational Monotony.

The most critical problem in applying circumscription is that of deciding on what predicates to minimize (there is, in addition, a further technical question about which predicates to treat as fixed and which as variable in extension). Most often what is done is to introduce a family of *abnormality* predicates ab_1, ab_2, etc. A default rule then can be written in the form: $\forall x((F(x) \,\&\, \neg ab_i(x)) \rightarrow G(x))$, where "→" is the ordinary material conditional of classical logic. To derive the consequences of a theory, all of the abnormality predicates are simultaneously minimized. This simple approach fails to satisfy the principle of Specificity, since each default is given its own, independent abnormality predicate, and each is therefore treated with the same priority. It is possible to add special rules for the prioritizing of circumscription, but these are, of necessity, ad hoc and exogenous, rather than a natural result of the definition of the consequence relation.

Circumscription does have the capacity of representing the existence of *undercutting defeaters*. Suppose that satisfying predicate F provides a prima facie reason for supposing something to be a G, and suppose that we use the abnormality predicate ab_1 in representing this default rule. We can state that the predicate H provides an undercutting defeater to this inference by simply adding the rule: $\forall x(H(x) \rightarrow ab_1(x))$, stating that all Hs are abnormal in respect number 1.

5.6 Preferential Logics

Circumscription is a special case of a wider class of defeasible logics, the *preferential* logics (Shoham 1987). In preferential logics, $\Gamma \mid\!\sim p$ iff p is true in all of the *most preferred* models of Γ. In the case of circumscription, the most preferred models are those that minimize the extension of certain predicates, but many other kinds of preference relations can be used instead, so long as the preference relations are transitive and irreflexive (a strict partial order). A structure consisting of a set of models of a propositional or first-order language, together with a preference order on those models, is called a *preferential structure*. The symbol $\prec$ shall represent the preference relation. $M \prec M'$ means that M is strictly preferred to M'. A most preferred model is one that is *minimal* in the ordering.

In order to give rise to a cumulative logic (one that satisfies Cut and Cautious Monotony), we must add an additional condition to the preferential structures, a Limit Assumption (also known as the condition of *stopperedness* or *smoothness*:

> **Limit Assumption**: Given a theory T, and M, a non-minimal model of T, there exists a model M' which is preferred to M and which is a minimal model of T.

The Limit Assumption is satisfied if the preferential structure does not contain any infinite descending chains of more and more preferred models, with no minimal member. This is a difficult condition to motivate as natural, but without it, we can find preferential structures that give rise to nonmonotonic consequence relations that fail to be cumulative.

Once we have added the Limit Assumption, it is easy to show that any consequence relation based upon a preferential model is not only

cumulative but also supraclassical, strongly absorptive, and distributive. Let's call such logics *preferential*. In fact, Kraus, Lehmann, and Magidor (Kraus, Lehmann, and Magidor 1990; Makinson 1994, 77; Makinson 2005, PAGE) proved the following representation theorem for preferential logics:

> **Representation Theorem for Preferential Logics**: if $\mid\!\sim$ is a cumulative, supraclassical, strongly absorptive, and distributive consequence relation (i.e., a preferential relation) then there is a preferential structure $\mathcal{M}$ satisfying the Limit Assumption such that for all *finite* theories T, the set of $\mid\!\sim$ -consequences of T is exactly the set of formulas true in every preferred model of T in $\mathcal{M}$.[3]

There are preferential logics that fail to satisfy consistency preservation, as well as disjunctive rationality and rational monotony:

> **Disjunctive Rationality**:
> If $\Gamma \cup \{p\} \mid\!\not\sim r$, and $\Gamma \cup \{q\} \mid\!\not\sim r$, then $\Gamma \cup \{(p \vee q)\} \mid\!\not\sim r$.
>
> **Rational Monotony**:
> If $\Gamma \mid\!\sim p$, then either $\Gamma \cup \{q\} \mid\!\sim p$ or $\Gamma \mid\!\sim \neg q$.

A very natural condition has been found by Kraus, Lehmann, and Magidor that corresponds to Rational Monotony: that of *ranked models*. (No condition on preference structures has been found that ensures disjunctive rationality without also ensuring rational monotony.) A preferential structure $\mathcal{M}$ satisfies the Ranked Models condition just in case there is a function r that assigns an ordinal number to each model in such a way that $M \prec M'$ iff $r(M) < r(M')$. Let's say that a preferential consequence relation is a *rational* relation just in case it satisfies Rational Monotony, and that a preferential structure is a *rational* structure just in case it satisfies the ranked models condition. Kraus, Lehmann, and Magidor

(Kraus, Lehmann, and Magidor 1990; Makinson 1994, 71–81) also proved the following representation theorem:

> **Representation Theorem for Rational Logics**: if $\mid\!\sim$ is a rational consequence relation (i.e., a preferential relation that satisfies Rational Monotony) then there is a preferential structure $\mathcal{M}$ satisfying the Limit Assumption and the Ranked Models Assumption such that for all finite theories T, the set of $\mid\!\sim$ -consequences of T is exactly the set of formulas true in every preferred model of T in $\mathcal{M}$.

Freund proved an analogous representation result for preferential logics that satisfy *disjunctive rationality*, replacing the ranking condition with a weaker condition of *filtered models*: a filtered model is one such that, for every formula, if two worlds non-minimally satisfy the formula, then there is a world less than both of them that also satisfies the formula (Freund 1993).

5.7 Logics of Extreme Probabilities

Lehmann and Magidor (Lehmann and Magidor 1992) noticed an interesting coincidence: the metalogical conditions for preferential consequence relations correspond exactly to the axioms for a logic of conditionals developed by Ernest W. Adams (Adams 1975).[4] Adams's logic was based on a conditional, ⇒, intended to represent a relation of very high conditional probability: ($p \Rightarrow q$) means that the conditional probability $Pr(q/p)$ is extremely close to 1. Adams used the standard delta-epsilon definition of the calculus to make this idea precise. Let us suppose that a theory T consists of a set of conditional-free formulas (the facts) and a set of probabilistic conditionals. A conclusion p follows defeasibly from T if and only if every probability function satisfies the following condition:

> For every δ, there is an ε such that, if the probability of every fact in T is assigned a probability at least as high as $1 - \varepsilon$, and every conditional in T is assigned a conditional probability at least as high as $1 - \varepsilon$, then the probability of the conclusion p is at least $1 - \delta$.

The resulting defeasible consequence relation is a preferential relation. (It need not, however, be consistency-preserving.) This consequence relation also corresponds to a relation, 0-entailment, defined by Judea Pearl (Pearl 1990), as the common core to all defeasible consequence relations.

Lehmann and Magidor (1992) proposed a variation on Adams's idea. Instead of using the delta-epsilon construction, they made use of nonstandard measure theory, that is, a theory of probability functions that can take values that are *infinitesimals* (infinitely small numbers). In addition, instead of defining the consequence relation by quantifying over *all* probability functions, Lehmann and Magidor assume that we can select a single probability function (representing something like the ideally rational, or objective probability). On their construction, a conclusion p follows from T just in case the probability of p is infinitely close to 1, on the assumption that the probabilities assigned to members of T are infinitely close to 1. Lehmann and Magidor proved that the resulting consequence relation is always not only preferential: it is also *rational*. The logic defined by Lehmann and Magidor also corresponds exactly to the theory of Popper functions, another extension of probability theory designed to handle cases of conditioning on propositions with infinitesimal probability (see Harper 1976; van Fraassen 1995; Hawthorne 1998). For a brief discussion of Popper functions, see the following supplementary document:

> Popper Functions

Arló Costa and Parikh, using van Fraassen's account (van Fraassen, 1995) of primitive conditional probabilities (a variant of Popper functions), proved a representation result for both finite and infinite languages (Arló Costa and Parikh, 2005). For infinite languages, they assumed an axiom of countable additivity for probabilities.

Kraus, Lehmann, and Magidor proved that, for every preferential consequence relation $\mid\!\sim$ that is probabilistically admissible,[5] there is a unique rational consequence relation $\mid\!\sim^*$ that minimally extends it (that is, that the intersection of all the rational consequence relations extending $\mid\!\sim$ is also a rational consequence relation). This relation, $\mid\!\sim^*$, is called the *rational closure* of $\mid\!\sim$. To find the rational closure of a preferential relation, one can perform the following operation on a preferential structure that supports that relation: assign to each model in the structure the smallest number possible, respecting the preference relation. Judea Pearl also proposed the very same idea under the name *1-entailment* or *System Z* (Pearl 1990).

A critical advantage to the Lehmann-Magidor-Pearl 1-entailment system over Adams's epsilon-entailment lies in the way in which 1-entailment handles irrelevant information. Suppose, for example, that we know that birds fly ($B \Rightarrow F$), Tweety is a bird (B), and Nemo is a whale (W). These premises do not epsilon-entail F (that Tweety flies), since there is no guarantee that a probability function assign a high probability to F, given the *conjunction* of B and W. In contrast, 1-entailment does give us the conclusion F.

Moreover, 1-entailment satisfies a condition of *weak independence of defaults*: conditionals with logically unrelated antecedents can "fire" independently of each other: one can warrant a conclusion even though we are given an explicit exception to the other. Consider, for example, the following case: birds fly ($B \Rightarrow F$), Tweety is a bird that doesn't fly (B & $\neg F$), whales are large ($W \Rightarrow L$), and Nemo is a whale (W). These premises 1-entail that Nemo is large (L). In addition, 1-entailment automatically satisfies the principle of Specificity: conditionals with more specific antecedents are always given priority over those with less specific antecedents.

There is another form of independence, *strong independence*, that even 1-entailment fails to satisfy. If we are given one exception to a rule involving a given antecedent, then we are unable to use any conditional with the same antecedent to derive any conclusion whatsoever. Suppose, for example, that we know that birds fly ($B \Rightarrow F$), Tweety is a bird that doesn't fly (B & $\neg F$), and birds lay eggs ($B \Rightarrow E$). Even under 1-entailment, the conclusion that Tweety lays eggs (E) fails to follow. This failure to satisfy Strong Independence is also known as *the Drowning Problem* (since all conditionals with the same antecedent are "drowned" by a single exception).

A consensus is growing that the Drowning Problem should not be "solved" (see Pelletier and Elio 1994; Wobcke 1995, 85; Bonevac, 2003, 461–462). Consider the following variant on the problem: birds fly, Tweety is a bird that doesn't fly, and birds have strong forelimb muscles. Here it seems we should refrain from concluding that Tweety has strong forelimb muscles, since there is reason to doubt that the strength of wing muscles is causally (and hence, probabilistically) independent of capacity for flight. Once we know that Tweety is an exceptional bird, we should refrain from applying other conditionals with *Tweety is a bird* as their antecedents, unless we know that these conditionals are independent of flight, that is, unless we know that the conditional with the stronger antecedent, *Tweety is a non-flying bird*, is also true.

Nonetheless, several proposals have been made for securing strong independence and solving the Drowning Problem. Geffner and Pearl

(Geffner and Pearl 1992) proposed a system of *conditional entailment*, a variant of circumscription, in which the preference relation on models is defined in terms of the sets of defaults that are satisfied. This enables Geffner and Pearl to satisfy both the Specificity principle and Strong Independence. Another proposal is the maximum entropy approach (Pearl 1988, 490–496; Goldszmidt, Morris and Pearl, 1993; Pearl 1990). A theory T, consisting of defaults Δ and facts F, entails p just in case the probability of p, conditional on F, approaches 1 as the probabilities associated with Δ approach 1, using the entropy-maximizing[6] probability function that respects the defaults in Δ. The maximum-entropy approaches satisfies both Specificity and Strong Independence.

Every attempt to solve the drowning problem (including conditional entailment and the maximum-entropy approach) comes at the cost of sacrificing cumulativity. Securing strong independence makes the systems very sensitive to the exact *form* in which the default information is stored. Consider, for example the following case: Swedes are (normally) fair, Swedes are (normally) tall, Jon is a short Swede. Conditional entailment and maximum-entropy entailment would permit the conclusion that Jon is fair in this case. However, if we replace the first two default conditionals by the single default, *Swedes are normally both tall and fair*, then the conclusion no longer follows, despite the fact that the new conditional is logically equivalent to the conjunction of the two original conditionals.

Applying the logic of extreme probabilities to real-world defeasible reasoning generates an obvious problem, however. We know perfectly well that, in the case of the default rules we actually use, the conditional probability of the conclusion on the premises is nowhere near 1. For example, the probability that an arbitrary bird can fly is certainly not infinitely close to 1. This problem resembles that of using idealizations in science, such as frictionless planes and ideal gases. It seems reasonable to think that, in deploying the machinery of defeasible logic, we indulge in the degree of make-believe necessary to make the formal models applicable. Nonetheless, this is clearly a problem warranting further attention.

5.8 Fully Expressive Languages: Conditional Logics and Higher-Order Probabilities

With relatively few exceptions, the logical approaches to defeasible reasoning developed so far put severe restrictions on the logical form of propositions included in a set of premises. In particular, they require the default conditional operator, $\Rightarrow$, to have wide scope in every formula in which it appears. Default conditionals are not allowed to be nested within other default conditionals, or within the scope of the usual Boolean operators of propositional logic (negation, conjunction, disjunction, material conditional). This is a very severe restriction and one that is quite difficult to defend. For example, in representing *undercutting defeaters*, it would be very natural to use a negated default conditional of the form $\neg((p \;\&\; q) \Rightarrow r)$ to signify that q defeats p as a prima facie reason for r. In addition, it seems plausible that one might come gain *disjunctive* default information: for example, that either customers are gullible or salesman are wily.

Asher and Pelletier (Asher and Pelletier 1997) have argued that, when translating generic sentences in natural language, it is essential that we be allowed to nest default conditionals. For example, consider the following English sentences:

> Close friends are (normally) people who (normally) trust one another.
>
> People who (normally) rise early (normally) go to bed early.

In the first case, a conditional is nested within the consequent of another conditional:

$$\forall x \forall y(\mathit{Friend}(x,y) \Rightarrow \forall z(\mathit{Time}(z) \Rightarrow \mathit{Trust}(x,y,z)))$$

In the second case, we seem to have conditionals nested within both the antecedent and the consequent of a third conditional, something like:

$$\forall x(\mathit{Person}(x) \rightarrow \\ (\forall y(\mathit{Day}(y) \Rightarrow \mathit{Rise\text{-}early}(x,y)) \Rightarrow \forall z(\mathit{Day}(z) \Rightarrow \mathit{Bed\text{-}early}(x,z))))$$

This nesting of conditionals can be made possible by borrowing and modifying the semantics of the subjunctive or counterfactual conditional, developed by Robert Stalnaker and David K. Lewis (Lewis 1973). For an axiomatization of Lewis's conditional logic, see the following supplementary document:

David Lewis's Conditional Logic

The only modification that is essential is to drop the condition of Centering (both strong and weak), a condition that makes modus ponens (affirming the antecedent) logically valid. If the conditional ⇒ is to represent a default conditional, we do not want modus ponens to be valid: we do not want $(p \Rightarrow q)$ and p to entail q classically (i.e., monotonically). If Centering is dropped, the resulting logic can be made to correspond exactly to either a preferential or a rational defeasible entailment relation. For example, the condition of Rational Monotony is the exact counterpart of the CV axiom of Lewis's logic:

CV:
$(p \Rightarrow q) \rightarrow [((p \,\&\, r) \Rightarrow q) \vee (p \Rightarrow \neg r)]$

Something like this was proposed first by James Delgrande (Delgrande 1987), and the idea has been most thoroughly developed by Nicholas

Asher and his collaborators (Asher and Morreau 1991; Asher 1995; Asher and Bonevac 1996; Asher and Mao 2001) under the name *Commonsense Entailment*.[7] Commonsense Entailment is a preferential (although not a rational) consequence relation, and it automatically satisfies the Specificity principle. It permits the arbitrary nesting of default conditionals within other logical operators, and it can be used to represent undercutting defeaters, through the use of negated defaults (Asher and Mao 2001).

The models of Commonsense Entailment differ significantly from those of preferential logic and the logic of extreme probabilities. Instead of having structures that contain sets of *models* of a standard, default-free language, a model of the language of Commonsense Entailment includes a set of *possible worlds*, together with a function that assigns standard interpretation (a model of the default-free language) to each world. In addition, to each pair consisting of a world w and a set of worlds (proposition) A, there is a function $*$ that assigns a set of worlds $*(w,A)$ to the pair. The set $*(w,A)$ is the set of most normal A-worlds, from the perspective of w. A default conditional $(p \Rightarrow q)$ is true in a world w (in such a model) just in case all of the most normal p worlds (from w's perspective) are worlds in which q is also true. Since we can assign truth-conditions to each such conditional, we can define the truth of nested conditionals, whether the conditionals are nested within Boolean operators or within other conditionals. Moreover, we can define both a classical, monotonic consequence relation for this class of models and a defeasible, nonmonotonic relation (in fact, the nonmonotonic consequence relation can be defined in a variety of ways). We can then distinguish between a default conditional's following *with logical necessity* from a default theory and its following *defeasibly* from that same theory. Contraposition, for example — inferring $(\neg q \Rightarrow \neg p)$ from $(p \Rightarrow q)$ — is not logically valid for default conditionals, but it might be a defeasibly correct inference.[8]

The one critical drawback to Commonsense Entailment, when compared to the logic of extreme probabilities, is that it lacks a single, clear standard of normativity. The truth-conditions of the default conditional and the definition of nonmonotonic consequence can be fine-tuned to match many of our intuitions, but in the end of the day, the theory of Commonsense Entailment offers no simple answer to the question of what its conditional or its consequence relation are supposed (ideally) to represent.

Logics of extreme probability (beginning with the work of Ernest Adams) did not permit the nesting of default conditionals for this reason: the conditionals were supposed to represent something like subjective conditional probabilities of the agent, to which the agent was supposed to have perfect introspective access. Consequently, it made no sense to nest this conditionals within disjunctions (as though the agent couldn't tell which disjunct represented his actual probability assignment) or within other conditionals (since the subjective probability of a subjective probability is always trivial — either exactly 1 or exactly 0). However, there is no reason why the logic of extreme probabilities couldn't be given a different interpretation, with $(p \Rightarrow q)$ representing something like *the objective probability of q, conditional on p, is infinitely close to 1*. In this case, it makes perfect sense to nest such statements of objective conditional probability within Boolean operators (either the probability of q on p is close to 1, or the probability of r on s is close to 1), or within operators of objective probability (the objective probability that the objective probability of p is close to 1 is itself close to 1). What is required in the latter case is a theory of *higher-order probabilities*.

Fortunately, such a theory of higher-order probabilities is available (see Skyrms 1980; Gaifman 1988). The central principle of this theory is Miller's principle. For a description of the models of the logic of extreme, higher-order probability, see the following supplementary document:

Models of Higher-Order Probability

The following proposition is logically valid in this logic, representing the presence of a defeasible modus ponens rule:

$$((p \,\&\, (p \Rightarrow q)) \Rightarrow q)$$

This system can be the basis for a family of rational nonmonotonic consequence relations that include the Adams ε-entailment system as a proper part (see Koons 2000, 298–319).

5.9 Objections to Nonmonotonic Logic

5.9.1 Confusing Logic and Epistemology?

In an early paper (Israel 1980), David Israel raised a number of objections to the very idea of *nonmonotonic logic*. First, he pointed out that the nonmonotonic consequences of a finite theory are typically not semi-decidable (recursively enumerable). This remains true of most current systems, but it is also true of second-order logic, infinitary logic, and a number of other systems that are now accepted as logical in nature.

Secondly, and more to the point, Israel argued that the concept of *nonmonotonic logic* evinces a confusion between the rules of logic and rules of inference. In other words, Israel accused defenders of nonmonotonic logic of confusing a theory of defeasible inference (a branch of epistemology) with a theory of genuine consequence relations (a branch of logic). Inference is nonmonotonic, but logic (according to Israel) is essentially monotonic.

The best response to Israel is to point out that, like deductive logic, a theory of nonmonotonic or defeasible consequence has a number of applications besides that of guiding actual inference. Defeasible logic can be used as part of a theory of scientific explanation, and it can be used in hypothetical reasoning, as in planning. It can be used to interpret implicit features of stories, even fantastic ones, so long as it is clear which actual default rules to suspend. Thus, defeasible logic extends far beyond the boundaries of the theory of epistemic justification. Moreover, as we have seen, nonmonotonic consequence relations (especially the preferential ones) share a number of very significant formal properties with classical consequence, warranting the inclusion of them all in a larger family of logics. From this perspective, classical deductive logic is simply a special case: the study of indefeasible consequence.

5.9.2 Problems with the Deduction Theorem

In a recent paper, Charles Morgan (Morgan 2000) has argued that nonmonotonic logic is impossible. Morgan offers a series of impossibility proofs. All of Morgan's proofs turn on the fact that nonmonotonic logics cannot support a generalized deduction theorem, i.e., something of the following form:

$$\Gamma \cup \{p\} \mid\!\sim q \text{ iff } \Gamma \mid\!\sim (p \Rightarrow q)$$

Morgan is certainly right about this.

However, there are good grounds for thinking that a system of nonmonotonic logic *should* fail to include a generalized deduction theorem. The very nature of defeasible consequence ensures that it must be so. Consider, for example, the left-to-right direction: suppose that $\Gamma \cup \{p\} \mid\!\sim q$. Should it follow that $\Gamma \mid\!\sim (p \Rightarrow q)$? Not at all. It may be that, normally, if p then $\neg q$, but Γ may contain defaults and information that defeat and override this inference. For instance, it might contain the fact r and the default $((r \,\&\, p) \Rightarrow q)$. Similarly, consider the right-to-left direction: suppose that $\Gamma \mid\!\sim (p \Rightarrow q)$. Should it follow that $\Gamma \cup \{p\} \mid\!\sim q$? Again, clearly not. Γ might contain both r and a default $((p \,\&\, r) \Rightarrow \neg q)$, in which case $\Gamma \cup \{p\} \mid\!\sim \neg q$.

It would be reasonable, however, to demand that a system of nonmonotonic logic satisfy the following *special deduction theorem*:

$$\{p\} \mid\!\sim q \text{ iff } \varnothing \mid\!\sim (p \Rightarrow q)$$

This is certainly possible. The special deduction theorem holds trivially; if we define$\{p\} \mid\!\sim q$ as $\varnothing \vDash (p \Rightarrow q)$; that is, $\{p\}$ defeasibly entails q if and only if (by definition) $(p \Rightarrow q)$ is a theorem of the classical conditional logic.[9]

6. Causation and Defeasible Reasoning

6.1 The Need for Explicit Causal Information

Hanks and McDermott, computer scientists at Yale, demonstrated that the existing systems of nonmonotonic logic were unable to give the right solution to a simple problem about predicting the course of events (Hanks and McDermott 1987). The problem became known as *the Yale shooting problem*. Hanks and McDermott assume that some sort of *law of inertia* can be assumed: that normally properties of things do not change. In the Yale shooting problem, there are two relevant properties: being loaded (a property of a gun) and being alive (a property of the intended victim of the shooting). Let's assume that in the initial situation, s_0, the gun is loaded and the victim is alive, $Loaded(s_0)$ and $Alive(s_0)$, and that two actions are performed in sequence: *Wait* and *Shoot*. Let's call the situation that results from a moment of waiting s_1, and the situation that follows both waiting and then shooting s_2. There are then three instances of the law of inertia that are relevant:

- $Alive(s_0) \Rightarrow Alive(s_1)$

- $Loaded(s_0) \Rightarrow Loaded(s_1)$
- $Alive(s_1) \Rightarrow Alive(s_2)$

We need to make one final assumption: that shooting the victim with a loaded gun results in death (not being alive):

- $((Alive(s_1)\ \&\ Loaded(s_1)) \rightarrow \neg Alive(s_2)$

Intuitively, we should be able to derive the defeasible conclusion that the victim is still alive after waiting, but dead after waiting and shooting: *Alive* $(s_1)\ \&\ \neg Alive(s_2)$. However, none of the nonmonotonic logics described above give us this result, since each of the three instances of the law of inertia can be violated: by the victim's inexplicably dying while we are waiting, by the gun's miraculously becoming unloaded while we are waiting, or by the victim's dying as a result of the shooting. Nothing introduced into nonmonotonic logic up to this point provides us with a basis for preferring the second exception to the law of inertia to the first or third. What's missing is a recognition of the importance of causal structure to defeasible consequence.[10]

There are several even simpler examples that illustrate the need to include explicitly causal information in the input to defeasible reasoning. Consider, for instance, this problem of Judea Pearl's (Pearl 1988): if the sprinkler is on, then normally the sidewalk is wet, and, if the sidewalk is wet, then normally it is raining. However, we should not infer that it is raining from the fact that the sprinkler is on. (See Lifschitz 1990 and Lin and Reiter 1994 for additional examples of this kind.) Similarly, if we also know that if the sidewalk is wet, then it is slippery, we should be able to infer that the sidewalk is slippery if the sprinkler is on and it is *not* raining.

The distinction between causal and evidential rules has been used in the argumentative-narrative model of reasoning with evidence developed by Floris Bex and his colleagues (Bex et al. 2010; Bex 2011).

6.2 Causally Grounded Independence Relations

Hans Reichenbach, in his analysis of the interaction of causality and probability (Reichenbach 1956), observed that the immediate causes of an event probabilistically *screen off* from that event any other event that is not causally posterior to it. This means that, given the immediate causal antecedents of an event, the occurrence of that event is rendered probabilistically independent of any information about non-posterior events. When this insight is applied to the nonmonotonic logic of extreme probabilities, we can use causal information to identify which defaults function independently of others: that is, we can decide when the fact that one default conditional has an exception is irrelevant to the question of whether a second conditional is also violated (see Koons 2000, 320–323). In effect, we have a selective version of Independence of Defaults that is grounded in causal information, enabling us to dissolve the Drowning Problem.

For example, in the case of Pearl's sprinkler, since rain is causally prior to the sidewalk's being wet, the causal structure of the situation does not ensure that the rain is probabilistically independent of whether the sprinkler is on, given the fact that the sidewalk is wet. That is, we have no grounds for thinking that the probability of rain, conditional on the sidewalk's being wet, is identical to the probability of rain, conditional on the sidewalk's being wet and the sprinkler's being on (presumably, the former is higher than the latter). This failure of independence prevents us from using the ($Wet \Rightarrow Rain$) default, in the presence of the additional fact that the sprinkler is on.

In the case of the Yale shooting problem, the state of the gun's being loaded in the aftermath of waiting, $Loaded(s_1)$, has at its only causal antecedent the fact that the gun is loaded in s_0. The fact of $Loaded(s_0)$ screens off the fact that the victim is alive in s_0 from the conclusion

Loaded(s_1). Similarly, the fact that the victim is alive in s_0 screens off the fact that the gun is loaded in s_0 from the conclusion that the victim is still alive in s_1. In contrast, the fact that the victim is alive at s_1 does *not* screen off the fact that the gun is loaded at s_1 from the conclusion that the victim is still alive at s_2. Thus, we can assign higher priority to the law of inertia with respect to both *Load* and *Alive* at s_0, and we can conclude that the victim is alive and the gun is loaded at s_1. The causal law for shooting then gives us the desired conclusion, namely, that the victim is dead at s_2.

6.3 Causal Circumscription

Our knowledge of causal relatedness is itself very partial. In particular, it is difficult for us to verify conclusively that any two randomly selected facts are or are not causally related. It seems that in practice we apply something like Occam's razor, assuming that two randomly selected facts are not causally related unless we have positive reason for thinking otherwise. This invites the use of something like circumscription, minimizing the extension of the predicate *causes*. (This is in fact exactly what Fangzhen Lin does in his 1995 papers [Lin 1995].)

Once we have a set of tentative conclusions about the causal structure of the world, we can use Reichenbach's insight to enable us to localize the problem of reasoning by default in the presence of known abnormality. If a known abnormality is screened off from a default's rule's consequent by constituent of its antecedent, then the rule may legitimately be deployed.

Since circumscription is itself a nonmonotonic logical system, there are at least two independent sources of nonmonotonicity, or defeasibility: the minimization or circumscription of causal relevance, and the application of defeasible causal laws and laws of inertia.

A number of researchers in artificial intelligence have recently deployed one version of circumscription (namely, the *stable models* of Gelfond and Lifschitz [1988]) to problems of causal reasoning, building on an idea of Norman McCain and Hudson Turner's [McCain and Turner 1997]. McCain and Turner employ causal rules that specify when an atomic fact is adequately caused and when it is exogenous and not in need of causal explanation. They then assume a principle of *universal causation*, permitting only those models that provide adequate causal explanations for all non-exempt atomic facts, while in effect circumscribing the extension of the causally explained. This approach has been extended and applied by Giunchiglia, Lee, Lifschitz, McCain and Turner [2004], Ferraris [2007], and Ferraris, Lee, Lierler, Lifschitz and Yang [2012]. Joohyung Lee and Yi Wang (Lee and Wang 2016) have focused on introducing relative weights to the rules.

7. Implementations and Applications

7.1 Applications to Law

Prakken (1997) book provides an extensive treatment of the contributions of techniques from nonmonotonic logic to the formal modeling of legal reasoning. See also Prakken and Sartor (1996, 1998), Hage et al. (1993), Hage (1997), Lodder (1999), and Bench-Capon et al. (2004, 2009). Kevin Ashley's HYPO system (Ashley 1990) employs defeasible reasoning in the study of case-based reasoning in the law.

7.2 Reasoning about Probabilities

Sections 5.7 and 5.8 above discussed probabilistic semantics for defeasible logics. It is also possible to reason defeasibly about propositions that *explicitly* involve numerical probabilities. We can reason defeasibly about propositions that assign specific probabilities to the other propositions or that assert numerical relations (identity, inequalities) between such propositions.

Chitta Baral, Michael Gelfond, and Nelson Rushton (Baral et al. 2009) have developed a declarative language P-Log, which combines defeasible logic with Bayesian probability nets. They use Answer Set Prolog to provide the logical foundations. They make use of a version of the principle of indifference. Belief revision occurs by means of Bayesian conditioning. Baral et al. demonstrate that P-log can reason correctly about the Monty Hall problem and Simpson's paradox. See Gelfond and Kahl 2014, pp. 235–270 for the syntax and semantics of P-log.

Joohyung Lee and Yi Wang (Lee and Wang 2016) take a somewhat different approach, using the log-linear models of Markov Logic (Richardson and Domingos 2006), which they argue is a natural way to add probabilistic information to stable semantics for logic programming languages. A Markov Logic network is a way of finding the probability distribution to a Markov chain that is *stationary*, i.e., stable with respect to updating. Their approach incorporates the ProbLog model of Fierens et al. 2015 as a special case.

Anthony Hunter has developed a strategy for using argumentation theory to reason with incomplete and even inconsistent information about probabilities (Hunter 2020). Once inconsistencies have been eliminated by belief contraction, Hunter relies on the maximum entropy distribution that is consistent with the remaining constraints to define the optimal probability function.

7.3 Software Implementations

ArguMed by Verheij (Verheij 2005) computes a version of stable semantics. Chris Reed and Glenn Rowe (Reed and Rowe 2004) have developed *Araucaria*, an application for analyzing and diagramming legal arguments. Prakken's *ASPIC+* system (Prakken 2010) can be used in analyzing formal argumentative structures.

Michael Gelfond and Yulia Kahl (Gelfond and Kahl 2014, 131–151) discuss how to develop algorithms for efficiently computing answer sets for logic programming. They describe inference engines that can act as answer-set programming solvers.

7.4 Efficiency in Updating

An important practical problem that arises from applying formal models of defeasible reasoning is that of updating in light of new or retracted information. Must we re-compute the nonmonotonic consequences from scratch each time updating is required?

Beishui Liao and his collaborators (Liao et al. 2011) addressed the issue of the computational dynamics of argument systems by investigating under which conditions an argument system can be divided into modules, so that the implications of new information can be efficiently computed by updating only the affected module. They discovered that such modularity is possible for any semantics that has the property of *directionality*. A semantics is directional if and only if, for every argument structure *AS*, the intersection of any extension prescribed for *AS* with an unattacked set /(U/) is identical to one of the extensions prescribed for the restriction of *AS* to *U*, and vice versa. See also Baroni et al. 2018.

Bibliography

Adams, Ernest W., 1975, *The Logic of Conditionals*, Dordrecht: Reidel.

Alchourrón, C., Gärdenfors, P. and Makinson, D., 1982, "On the logic of theory change: contraction functions and their associated revision functions", *Theoria*, 48: 14–37.

Arló Costa, Horacio and Parikh, Rohit, 2005, "Conditional Probability and Defeasible Inference", *Journal of Philosophical Logic*, 34: 97–119.

Armstrong, David M., 1983, *What is a law of nature?*, New York: Cambridge University Press.

–––, 1997, *A world of states of affairs*, Cambridge: Cambridge University Press.

Asher, Nicholas, 1992, "A Truth Conditional, Default Semantics for Progressive", *Linguistics and Philosophy*, 15: 469–508.

–––, 1995, "Commonsense Entailment: a logic for some conditionals", in *Conditionals in Artificial Intelligence*, G. Crocco, L. Farinas del Cerro, and A. Hertzig (eds.), Oxford: Oxford University Press.

Asher, Nicholas and Daniel Bonevac, 1996, "Prima Facie Obligations", *Studia Logica*, 57: 19–45.

Asher, Nicholas, and Alex Lascarides, 2003, *Logics of Conversation*, Cambridge: Cambridge University Press.

Asher, N.. and Y. Mao, 2001, "Negated Defaults in Commonsense Entailment", *Bulletin of the Section of Logic*, 30: 4–60.

Asher, Nicholas, and Michael Morreau, 1991, "Commonsense Entailment: A Modal, Nonmonotonic Theory of Reasoning", in *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, John Mylopoulos and Ray Reiter (eds.), San Mateo, Calif.: Morgan Kaufmann.

Asher, N., and F.J. Pelletier, 1997, "Generics and Defaults", in *Handbook of Logic and Language*, J. van Bentham and A. ter Meulen (eds.), Amsterdam: Elsevier.

Ashley, Kevin D., 1990, *Modeling legal argument: Reasoning with cases and hypotheticals*, Cambridge, MA: MIT Press.

Baker, A. B., 1988, "A simple solution to the Yale shooting problem", in *Proceedings of the First International Conference on Knowledge Representation and Reasoning*, Ronald J. Brachman, Hector Levesque and Ray Reiter (eds.), San Mateo, Calif.: Morgan Kaufmann.

Bamber, Donald, 2000, "Entailment with Near Surety of Scaled Assertions of High Conditional Probability", *Journal of Philosophical Logic*, 29: 1–74.

Baroni, Pietro, Massimiliano Giacomin, and Beishui Liao, 2018, "Locality and Modularity in Abstract Argumentation", in P. Baroni, D. Gabbay, Massimiliano Giacomin, and Leendert van der Torre (eds.), *The Handbook of Formal Argumentation*, London: College Publications, 937–979.

Bench-Capon, T. J. M., H. Prakken, and G Sartor, 2009, "Argumentation in legal reasoning", in I. Rahwan and G. R. Simari (eds.), *Argumentation in Artificial Intelligence*, Dordrecht: Springer, pp. 363–382.

Bex, F. J., 2011, *Arguments, stories and criminal evidence: A formal hybrid theory*, Dordrecht: Springer.

Bex, F. J., P. van Koppen, H. Prakken, and B. Verheij, 2010, "A hybrid formal theory of arguments, stories and criminal evidence", *Artificial Intelligence and Law*, 18(2): 123–152.

Bochman, Alexander, 2001, *A Logical Theory of Nonmonotonic Inference and Belief Change*, Berlin: Springer.

Bodanza, Gustavo A. and F. Tohmé, 2005, "Local Logics, Non-Monotonicity and Defeasible Argumentation", *Journal of Logic, Language and Information*, 14: 1–12.

Bonevac, Daniel, 2003, *Deduction: Introductory Symbolic Logic*, Malden, Mass.: Blackwell, 2nd edition.

Carnap, Rudolf, 1962, *Logical Foundations of Probability*, Chicago: University of Chicago Press.

Carnap, Rudolf and Richard C. Jeffrey, 1980, *Studies in inductive logic and probability*, Berkeley: University of California Press.

Cartwright, Nancy, 1983, *How the laws of physics lie*, Oxford: Clarendon

Press.

Chisholm, Roderick, 1957, *Perceiving*, Princeton: Princeton University Press.

–––, 1963, "Contrary-to-Duty Imperatives and Deontic Logic", *Analysis*, 24: 33–36.

–––, 1966, *Theory of Knowledge*, Englewood Cliffs: Prentice-Hall.

Dancy, Jonathan, 1993, *Moral Reasons*, Malden, MA: Wiley-Blackwell.

–––, 2004, *Ethics without Principles*, Oxford: Clarendon Press.

Delgrande, J. P., 1987, "A first-order conditional logic for prototypical properties", *Artificial Intelligence*, 33: 105–130.

Dung, Phan Minh, 1995, "On the acceptability of arguments and its fundamental role in non-monotonic reasoning logic programming and n-person games", *Artificial Intelligence*, 77: 321–357.

Dung, Phan Minh and Phan Minh Thang, 2010, "Towards (probabilistic) argumentation for jury-based dispute resolution", in P. Baroni, F. Cerutti, M. Giacomin, and G. R. Simari (eds.), *Computational Models of Argument: Proceedings of COMMA 2010*, Amsterdam: Ios Press, 171–182).

–––, 2018, "Fundamental properties of attack relations in structured argumentation with priorities", *Artificial Intelligence*, 255: 1–42.

Etherington, D. W. and R. Reiter, 1983, "On Inheritance Hierarchies and Exceptions", in *Proceedings of the National Conference on Artificial Intelligence*, Los Altos, Calif.: Morgan Kaufmann.

Ferraris, Paolo, 2007, "A Logic Programming Characterization of Causal Theories", *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*, San Francisco, Calif.: Morgan Kaufmann.

Ferraris, Paolo, with J. Lee, Y. Lierler, V. Lifschitz, and F. Yang, 2012, "Representing first-order causal theories by logic programs", *Theory and Practice of Logic Programming*, 12(3): 383–412.

Fierens, D, G. van den Broeck, J. Renkens, D. Shterionov, B. Guttman, I.

Thon, G. Janssens, and L. de Readt, 2015, "Inference and learning in probabilistic logic using weighted Boolean formulas", *Theory and Practice of Logic Programming*, 15(03): 358–401.

Freund, M., with D. Lehmann, and D. Makinson, 1990, "Canonical extensions to the infinite case of finitary nonmonotonic inference relations", in *Proceedings of the Workshop on Nonmonotonic Reasoning*, G. Brewka and H. Freitag (eds.), Sankt Augustin: Gesellschaft für Mathematic und Datenverarbeitung mbH.

Freund, M., 1993, "Injective models and disjunctive relations", *Journal of Logic and Computation*, 3: 231–347.

Gabbay, D. M., 1985, "Theoretical foundations for non-monotonic reasoning in expert systems", in *Logics and Models of Concurrent Systems*, K. R. Apt (ed.), Berlin: Springer-Verlag.

Gaifman, Haim, 1988, "A theory of higher-order probabilities", in *Causation, Chance and Credence*, Brian Skyrms and William Harper (eds.), London, Ontario: University of Western Ontario Press.

Gärdenfors, P., 1978, "Conditionals and Changes of Belief", *Acta Fennica*, 30: 381–404.

–––, 1986, "Belief revisions and the Ramsey test for conditionals", *Philosophical Review*, 95: 81–93.

Geffner, H. A., 1992, *Default Reasoning: Causal and Conditional Theories*, Cambridge, MA: MIT Press.

Geffner, H. A., and J. Pearl, 1992, "Conditional entailment: bridging two approaches to default reasoning", *Artificial Intelligence*, 53: 209–244.

Gelfond, Michael and Yulia Kahl, 2014, *Knowledge Representation, Reasoning, and the Design of Intelligent Agents: The Answer-Set Programming Approach*, Cambridge: Cambridge University Press.

Gelfond, Michael and Leone, N., 2002, "Logic programming and knowledge representation—the A-prolog perspective", *Artificial Intelligence*, 138: 37–38.

Gelfond, Michael and Lifschitz, Vladimir, 1988, "The stable model semantics for logic programming", *Logic Programming: Proceedings of the Fifth International Conference and Symposium*, Robert A. Kowalski and Kenneth A. Bowen (eds.), Cambridge, Mass.: The MIT Press, pp. 1070–1080.

Gilio, Angelo, 2005, "Probabilistic Logic under Coherence, Conditional Interpretations, and Default Reasoning", *Synthese*, 146: 139–152.

Ginsberg, M. L., 1987, *Readings in Nonmonotonic Reasoning*, San Mateo, Calif.: Morgan Kaufmann.

Giunchiglia, E., with J. Lee, V. Lifschitz, N. McCain, and H. Turner, 2004, "Nonmonotonic Causal Theories", *Artificial Intelligence*, 153: 49–104.

Goldszmidt, M. and J. Pearl, 1992, "Rank-Based Systems: A Simple Approach to Belief Revision, Belief Update, and Reasoning about Evidence and Action", in *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, San Mateo, Calif.: Morgan Kaufmann.

Goldszmidt, M., with P. Morris, and J. Pearl, 1993, "A maximum entropy approach to nonmonotonic reasoning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15: 220–232.

Grove, A., 1988, "Two modellings for theory change", *Journal of Philosophical Logic*, 17: 157–170.

Hage, J. C., 1997, *Reasoning with rules: An essay on legal reasoning and its underlying logic*, Dordrecht: Kluwer Academic.

–––, 2000, "Dialectical models in artificial intelligence and law", *Artificial Intelligence and Law*, 8: 137–172.

Hanks, Steve and Drew McDermott, 1987, "Nonmonotonic Logic and Temporal Projection", *Artificial Intelligence*, 33: 379–412.

Hansson, B., 1969, "An analysis of some deontic logics", *Noûs*, 3: 373–398.

Hansson, S. O. and Makinson, D., 1997, "Applying normative rules with restraint", in *Logic and Scientific Methods*, M. Dalla Chiara (ed.), Dordrecht: Kluwer.

Harper, W. L., 1976, "Rational Belief Change, Popper Functions and Counterfactuals", in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Volume I*, Dordrecht: Reidel.

Hart, H. L. A., 1949, "The ascription of responsibility and rights", *Proceedings of the Aristotelian Society*, 49(1): 171–194.

Hawthorne, James, 1998, "On the Logic of Nonmonotonic Conditionals and Conditional Probabilities: Predicate Logic", *Journal of Philosophical Logic*, 27: 1–34.

Horty, J. F., with R.H. Thomason, and D.S. Touretzky, 1990, "A sceptical theory of inheritance in nonmonotonic semantic networks", *Artificial Intelligence*, 42: 311–348.

Horty, John, 1994, "Moral dilemmas and nonmonotonic logic", *Journal of Philosophical Logic*, 23: 35–65.

–––, 2003, "Reasoning with moral conflicts", *Noûs*, 37: 557–605.

–––, 2007a, "Defaults with Priorities", *Journal of Philosophical Logic*, 36: 367–413.

–––, 2007b, "Reasons as defaults", *Philosophers' Imprints*, 7: 1–28.

Hunter, Anthony, 2013, "A probabilistic approach to modelling uncertain logical arguments", *International Journal of Approximate Reasoning*, 54(1): 47–81.

–––, "Reasoning with Inconsistent Knowledge using the Epistemic Approach to Probabilistic Argumentation", *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR'20)*, Palo Alto: AAAI Press.

Israel, David, 1986, "What's Wrong with Non-monotonic Logic", in *Proceedings of the First National Conference on Artificial Intelligence*, Palo Alto: AAAI Press.

Karacapilidis, N., and D. Papadias, 2001, "Computer supported argumentation and collaborative decision making: The HERMES system", *Information Systems*, 26: 259–277.

Konolige, Kurt, 1994, "Autoepistemic Logic", in *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume III: Nonmonotonic Reasoning and Uncertain Reasoning*, D. M. Gabbay, C. J. Hogger, and J. A. Robinson (eds.), Oxford: Clarendon Press.

Koons, Robert C., 2000, *Realism Regained: An Exact Theory of Causation, Teleology and the Mind*, New York: Oxford University Press.

–––, 2001, "Defeasible Reasoning, Special Pleading and the Cosmological Argument: Reply to Oppy", *Faith and Philosophy*, 18: 192–203.

Kraus, S., with D. Lehmann, and M. Magidor, 1990, "Nonmonotonic Reasoning, Preferential Models and Cumulative Logics", *Artificial Intelligence*, 44: 167–207.

Kyburg, Henry E., 1983, *Epistemology and Inference*, Minneapolis: University of Minnesota Press.

–––, 1990, *Knowledge Representation and Defeasible Reasoning*, Dordrecht: Kluwer.

Lance, Mark and Margaret Little, 2004, "Defeasibility and the normative grasp of context", *Erkenntnis*, 61: 435–55.

–––, 2007, "Where the laws are", *Oxford Studies in Metaethics*, 2: 149–171.

Lascarides, Alex and Nicholas Asher, 1993, "Temporal Interpretation, Discourse Relations and Commonsense Entailment", *Linguistics and Philosophy*, 16: 437–493.

Lee, Joohyung and Yi Wang, 2016, "Weighted Rules under the Stable Model Semantics", in *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR 2016)*, Palo Alto: AAAI Press, pp. 145–154.

Lehmann, D., and M. Magidor, 1992, "What does a conditional knowledge base entail?", *Artificial Intelligence*, 55: 1–60.

Levesque, H., 1990, "A study in autoepistemic logic", *Artificial Intelligence*, 42: 263–309.

Lewis, David K., 1973, *Counterfactuals*, Cambridge, Mass.: Harvard University Press.

Liao, Beishui, Li Jin, and Robert C. Koons, 2011, "Dynamics of argumentation systems: A division-based method", *Artificial Intelligence*, 175(11): 1790–1814.

Liao, Beishui, Nir Oren, Leendert van der Torre, and Serena Villata, 2019, "Prioritized Norms in Formal Argumentation", *Journal of Logic and Computation*, 29(2): 215–240.

Lifschitz, V., 1988, "Circumscriptive theories: a logic-based framework for knowledge representation", *Journal of Philosophical Logic*, 17: 391–441.

–––, 1989, "Benchmark Problems for Formal Nonmonotonic Reasoning", in *Non-Monotonic Reasoning*, M. Reinfrank, J. de Kleer, M. L. Ginsberg and E. Sandewall (eds.), Berlin: Springer-Verlag.

–––, 1990, "Frames in the space of situations", *Artificial Intelligence*, 46: 365–376.

Lin, Fangzhen, 1995, "Embracing causality in specifying the indirect effects of actions", *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, San Mateo, Calif.: Morgan Kaufmann, pp. 1985–1993.

Lin, Fangzhen, and Robert Reiter, 1994, "State constraints revisited", *Journal of Logic and Computation*, 4: 655–678.

Lukasiewicz, Thomas, 2005, "Nonmonotonic Probabilistic Reasoning under Variable-Strength Inheritance with Overriding", *Synthese*, 146: 153–169.

McCain, Norman and Hudson Turner, 1997, "Causal theories of action and change", in *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI)*, Cambridge, Mass.: The MIT Press, pp. 460–5.

McCarthy, John M. and Patrick J. Hayes, 1969, "Some Philosophical

Problems from the Standpoint of Artificial Intelligence”, in *Machine Intelligence 4*, B. Meltzer and D. Mitchie (eds.), Edinburgh: Edinburgh University Press.

–––, 1977, “Epistemological Problems of Artificial Intelligence”, in *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, Pittsburgh: Computer Science Department, Carnegie-Mellon University.

–––, 1982, “Circumscription — A Form of Non-Monotonic Reasoning”, *Artificial Intelligence*, 13: 27–39, 171–177.

–––, 1986, “Application of Circumscription to Formalizing Common-Sense Knowledge”, *Artificial Intelligence*, 28: 89–111.

McDermott, Drew and Jon Doyle, 1982, “Non-Monotonic Logic I”, *Artificial Intelligence*, 13: 41–72.

Makinson, David, 1994, “General Patterns in Nonmonotonic Reasoning”, in *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume III: Nonmonotonic Reasoning and Uncertain Reasoning*, D. M. Gabbay, C. J. Hogger, and J. A. Robinson (eds.), Oxford: Clarendon Press.

–––, 2005, *Bridges from Classical to Nonmonotonic Logic*, London: King’s College Publications.

Makinson, David and Gärdenfors, Peter, 1991, “Relations between the logic of theory change and Nonmonotonic Logic”, in *Logic of Theory Change*, A. Fuhrmann and M. Morreau (eds.), Berlin: Springer-Verlag.

Makinson, David and van der Torre, L., 2000, “Input/output logics”, *Journal of Philosophical Logic*, 29: 155–85.

Morgan, Charles, 2000, “The Nature of Nonmonotonic Reasoning”, *Minds and Machines*, 10: 321–360.

Moore, Robert C., 1985, “Semantic Considerations on Nonmonotonic Logic”, *Artificial Intelligence*, 25: 75–94.

Morreau, M., and N. Asher, 1995, “What some generic sentences mean”,

in *The Generic Book*, J. Pelletier (ed.), Chicago: University of Chicago Press.

Nayak, A. C., 1994, “Iterated belief change based on epistemic entrenchment”, *Erkenntnis*, 41: 353–390.

Nute, Donald, 1988, “Conditional Logic”, in *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, D. Gabbay and F. Guenthner (eds.), Dordrecht: D. Reidel.

–––, 1997, *Defeasible Deontic Logic*, Dordrecht: Kluwer.

Pearl, Judea, 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, Calif.: Morgan Kaufmann.

–––, 1990, “System Z: A Natural Ordering of Defaults with Tractable Applications to Default Reasoning”, Proceedings of the Third Conference on Theoretical Aspects of Reasoning about Knowledge, Rohit Parikh (ed.), San Mateo, Calif.: Morgan Kaufmann.

Pelletier, F. J. and R. Elio, “On Relevance in Nonmonotonic Reasoning: Some Empirical Studies”, in R. Greiner & D. Subramanian (eds) *Relevance: AAAI 1994 Fall Symposium Series*, Palo Alto: AAAI Press.

Pollock, John L., 1967, “Criteria and our knowledge of the material world”, *Philosophical Review*, 76: 28–62.

–––, 1970, “The structure of epistemic justification”, *American Philosophical Quarterly* (Monograph Series), 4: 62–78.

–––, 1974, *Knowledge and Justification*, Princeton: Princeton University Press.

–––, 1987, “Defeasible Reasoning”, *Cognitive Science*, 11: 481–518.

–––, 1995, *Cognitive Carpentry*, Cambridge, Mass.: MIT Press.

–––, 2010, “Defeasible reasoning and degrees of justification”, *Argument & Computation*, 1(1): 7–22.

Prakken, Henry, 2010, “An abstract framework for argumentation with structured arguments”, *Argument and Computation*, 1: 93–124.

Prakken, Henry and Giovanni Sartor, 1995, “On the relation between legal language and legal argument: assumptions, applicability, and dynamic priorities”, in *Proceedings of the Fifth International Conference on Artificial Intelligence and the Law (ICAIL-95)*, New York: The ACM Press.

–––, 1996, “A dialectical model of assessing conflicting arguments in legal reasoning”, *Artificial Intelligence and the Law*, 4: 331–368.

Prakken, H., and Vreeswijk, G. A. W., 2002, “Logics for defeasible argumentation”, in D. Gabbay and F. Guenthner (eds.), *Handbook of philosophical logic* (2nd edition, Volume 4), Dordrecht: Kluwer, pp. 219–318.

Quine, Willard van Orman, and J.S. Ullian, 1982, *The Web of Belief*, New York: Random House.

Raz, Joseph, 1975, *Practical Reasoning and Norms*, London: Hutchinson and Company.

Reed, Christopher A. and Rowe, G. W. A., 2004, “Araucaria: Software for argument analysis, diagramming and representation”, *International Journal on Artificial Intelligence Tools*, 13(4): 961–979.

Reiter, Ray, 1980, “A logic for default reasoning”, *Artificial Intelligence*, 13: 81–137.

Richardson, M. and P. Domingos, 2006, “Markov logic networks”, *Machine Learning*, 62(1–3): 107–136.

Ross, David, 1930, *The Right and the Good*, Oxford: Oxford University Press.

–––, 1939, *Foundations of Ethics*, Oxford: Clarendon Press.

Rott, Hans, 1989, “Conditionals and Theory Change: Revisions, Expansions and Additions”, *Synthese*, 81: 91–113.

Schlechta, Karl, 1997, *Nonmonotonic Logics: Basic Concepts, Results and Techniques*, Berlin: Springer-Verlag.

Shoham, Yoav, 1987, “A Semantical Approach to Nonmonotonic Logic”, in *Proceedings of the Tenth International Conference on Artificial Intelligence*, John McDermott (ed.), Los Altos, Calif.: Morgan Kaufmann.

Skyrms, Brian, 1980, “Higher order degrees of belief”, in *Prospects for Pragmatism*, Hugh Mellor (ed.), Cambridge: Cambridge University Press.

Spohn, Wolfgang, 1988, “Ordinal Conditional Functions”, in *Causation, Decision, Belief Change and Statistics, Volume III*, W. L. Harper and B. Skyrms (eds.), Dordrecht: Kluwer.

–––, 2002, “A Brief Comparison of Pollock’s Defeasible Reasoning and Ranking Functions”, *Synthese*, 13: 39–56.

Tohmé, Fernando, with Claudio Delrieux and Otávio Bueno, 2011, “Defeasible Reasoning + Partial Models: A Formal Framework for the Methodology of Research Programs”, *Foundations of Science*, 16: 47–65.

Toulmin, Stephen E., 1964, *The Uses of Argument*, Cambridge: Cambridge University Press.

Txurruka, I. and N. Asher, 2008, “A discourse-based approach to Natural Language Disjunction (revisited)”, in M. Aunargue, K. Korta and J. Lazzarabal (eds.), *Language, Representation and Reasoning*, University of the Basque Country Press.

van der Torre, Leendert and Srdjan Vesic, 2018, “The Principle-Based Approach to Abstract Argumentation Semantics”, in P. Baroni, D. Gabbay, Massimiliano Giacomin, and Leendert van der Torre (eds.), *The Handbook of Formal Argumentation*, London: College Publications, 797–838.

van Eemeron, Frans H., Bart Garssen, Erik C. W. Krabbe, A. Francisca Snoeck Henkemans, Bart Verheij, and Jean H. M. Wagemans, 2020, *Handbook of Argumentation Theory*, Dordrecht: Springer Netherlands.

van Fraassen, Bas, 1973, “Values and the heart’s command”, *The Journal of Philosophy*, 70: 5–19.

–––, 1995, “Fine-grained opinion, probability, and the logic of folk belief”, *Journal of Philosophical Logic*, 24: 349–377.

Verheij, B., 2003, “DefLog: On the logical interpretation of prima facie justified assumptions”, *Journal of Logic and Computation*, 13(3): 319–346.

–––, 2005, *Virtual arguments: On the design of argument assistants for lawyers and other arguers*, The Hague: T. M. C. Asser Press.

–––, 2012, “Jumping to conclusions: A logico-probabilistic foundation for defeasible rule-based arguments”, in L. Fariñas del Cerro, A. Herzig & J. Mengin (eds.), *Logics in Artificial Intelligence. 13th European conference, JELIA 2012*, Dordrecht: Springer, 411–423.

–––, 2017, “Proof with and without probabilities: Correct evidential reasoning with presumptive arguments, coherent hypotheses and degrees of uncertainty”, *Artificial Intelligence and Law*, 25 (1): 127–154.

Vieu, L., with M. Bras, N. Asher, and M. Aurnague, 2005, “Locating adverbials in discourse”, *Journal of French Language Studies*, 15(2): 173–193.

Vreeswijk, Gerard, 1997, “Abstract argumentation systems”, *Artificial Intelligence*, 90: 225–27.

Wobcke, Wayne, 1995, “Belief Revision, Conditional Logic and Nonmonotonic Reasoning”, *Notre Dame Journal of Formal Logic*, 36: 55–103.

Academic Tools

- How to cite this entry.
- Preview the PDF version of this entry at the Friends of the SEP Society.
- Look up topics and thinkers related to this entry at the Internet Philosophy Ontology Project (InPhO).
- Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

- Online papers, Cognitive Systems Laboratory, UCLA Computer Science Department
- Daniel Lehmann’s home page, Hebrew University

Related Entries

artificial intelligence: logic and | causation: probabilistic | epistemology: Bayesian | logic: modal | logic: non-monotonic | logic: of belief revision | moral particularism | probability, interpretations of

John Pollock’s System

In defining what beliefs are warranted, given a directed graph representing a cognitive state, Pollock first defines a *partial status assignment* that assigns the statuses “defeated” or “undefeated” to some of the nodes of the graph.

An assignment σ is a partial status assignment iff:

1. σ assigns “undefeated” to all nodes that are such that neither they nor any of their ancestors are defeated by any nodes in the graph.
2. σ assigns “undefeated” to α iff σ assigns “undefeated” to all the immediate ancestors of α, and all nodes defeating α are assigned “defeated”.
3. σ assigns “defeated” to a node α iff either α has an immediate ancester that is assigned “defeated”, or there is a node β that defeats α and that is assigned “undefeated”.

An assignment σ is a *status assignment* iff σ is a *maximal* partial status assignment. A node is *defeated outright* iff no status assignment assigns “undefeated” to it. If some status assignments assign “defeated” to it, and some assign “undefeated” to it, then it is *provisionally defeated*. A node is warranted if it is neither defeated outright nor provisionally defeated; that is, if every status assignment assigns “undefeated” to it. Here are some of the consequences of Pollock’s definitions:

- A node α is undefeated iff all its immediate ancestors are undefeated and all nodes defeating α are defeated.
- If some immediate ancestor of α is defeated outright, then α is defeated outright.
- If some node defeating α is undefeated, then α is defeated outright.
- If α is self-defeating, then α is defeated outright.

Semantic Inheritance Networks

A path is a sequence of links in a graph G, with the final node of each link being the initial node of the next, where all the links, with the possible exception of the last one, are positive. A *generalized* path is a sequence of links that can contain negative links anywhere, and more than one. Each path has both an initial node and a final node. A path can be taken as representing an assertion about an individual: that the individual corresponding to the initial node belongs to the category corresponding to the final node. The *degree* of a path is the length of the longest generalized path connecting the path’s initial node to its final node.

Horty, Thomason, and Touretzky define the relation of *support* between graphs (cognitive states) and paths (assertions) by mathematical induction on the degree of the path. Direct links (paths of length one) are always supported by the graph.

1. If σ is a positive path, $x \to \sigma^1 \to u \to y$, then G supports σ iff:
    1. G supports path $x \to \sigma^1 \to u$.
    2. $u \to y$ is a direct link in G.
    3. The negative link $x \not\to y$ does not belong to G.
    4. for all v, τ such that G supports $x \to \tau \to v$, with the negative link $v \not\to y$ in G, there exist z, τ^1, τ^2 such that $z \to y$ is in G, and either $z = x$, or G supports the path $x \to \tau^1 \to z \to \tau^2 \to$ v.
2. If σ is a negative path, $x \to \sigma^1 \to u \not\to y$, then G supports σ iff:
    1. G supports path $x \to \sigma^1 \to u$.
    2. $u \not\to y$ is a direct negative link in G.
    3. The positive link $x \to y$ does not belong to G.
    4. for all v, τ such that G supports $x \to \tau \to v$, with the positive link $v \to y$ in G, there exist z, τ^1, τ^2 such that $z \not\to y$ is in G, and either $z = x$, or G supports the path $x \to \tau^1 \to z \to \tau^2 \to v$.

The definition ensures that each potentially conflicting path be preempted by a path with a specificity-based priority.

AGM Postulates

Where K is a belief state, $K * A$ represents the set of beliefs resulting from revising K with new belief A.

($K * 1$) $K * A$ is closed under logical consequence.
($K * 2$) A belongs to $K * A$.
($K * 3$) $K * A$ is a subset of the logical closure of $K \cup \{A\}$.
($K * 4$) If $\neg A$ does not belong to K, then the closure of $K \cup \{A\}$ is a subset of $K * A$.
($K * 5$) If $K * A$ is logically inconsistent, then either K is inconsistent, or A is.
($K * 6$) If A and B are logically equivalent, then $K{*}A = K{*}B$.
($K * 7$) $K * (A \,\&\, B)$ is a subset of the logical closure of $K * A \cup \{B\}$.

($K * 8$) If $\neg B$ does not belong to $K * A$, then the logical closure of $K * A \cup B$ is a subset of $K * (A \,\&\, B)$.

Popper Functions

A Popper function is a function from pairs of propositions to real numbers that satisfies the following conditions:

1. For some $D, E, P[D \mid E] \neq 1$.
2. $P[A \mid A] = 1$.
3. $P[A \mid (C \,\&\, B)] = P[A \mid (B \,\&\, C)]$.
4. $P[(B \,\&\, A) \mid C] = P[(A \,\&\, B) \mid C]$.
5. $P[A \mid B] + P[\neg A \mid B] = 1$, or $P[C \mid B] = 1$.
6. $P[(A \,\&\, B) \mid C] = P[A \mid (B \,\&\, C)] \times P[B \mid C]$.

David Lewis's Conditional Logic

This is Donald Nute's axiom system (Nute 1984, 396-399) for David K. Lewis's preferred logic for the counterfactual conditional, *VC* (Lewis 1973, 132):

Rules:

1. Modus ponens.
2. Deduction within the consequent of conditionals: if $\chi_1 \ldots \chi_n$ logically entails ψ, then the conditionals $(\phi \Rightarrow \chi_1) \ldots (\phi \Rightarrow \chi_n)$ logically entail $(\phi \Rightarrow \psi)$.
3. Interchange of logical equivalents.

Axioms:

1. All truth-functional tautologies.
2. *ID*: $(\phi \Rightarrow \phi)$

3. *MOD*: $(\neg\phi \Rightarrow \phi) \rightarrow (\psi \Rightarrow \phi)$
4. *CSO*: $[(\phi \Rightarrow \psi) \,\&\, (\psi \Rightarrow \phi)] \rightarrow [(\phi \Rightarrow \chi) \leftrightarrow (\psi \Rightarrow \chi)]$
5. *CV*: $(\phi \Rightarrow \psi) \rightarrow [((\phi \,\&\, \chi) \Rightarrow \psi) \vee (\phi \Rightarrow \neg\chi)]$
6. *MP*: $(\phi \Rightarrow \psi) \rightarrow (\phi \rightarrow \psi)$
7. *CS*: $(\phi \,\&\, \psi) \rightarrow (\phi \Rightarrow \psi)$

The last two axioms, *MP* and *CS*, correspond to weak and strong centering, respectively (in effect, the stipulation that the actual world is one of the most, or uniquely the most, normal of all worlds). For nonmonotonic logic, these conditions, and these two axioms, must be dropped. The fifth axiom, *CV*, is the object-language correlate of Rational Monotony.

Models of Higher-Order Probability

A model of higher-order probability consists of W, a set of possible worlds, together with a function that assigns to each world w a probability function μ_w. Let's assume that these probability functions are non-standard, that is, they may assign infinitesimal probabilities to some sets of worlds. Let $[W]$ be the partition of W in terms of probabilistic agreement: worlds w and w' belong to the same cell of $[W]$ just in case they are assigned exactly the same probability function.

Let A be a subset of $[W]$. The set-theoretic union of A, $\bigcup A$, is a proposition expressible entirely in terms of probabilities (that is, entirely in terms of Boolean combinations of $\Rightarrow$ conditionals). In the finite case, Miller's principle states that, for all worlds w and all propositions B and C:

$$\mu_w(C/B \cap \bigcup A) = \Sigma_{w' \in \cup A}\, \mu_{w'}(C/B) \times \mu_w(\{w\})$$

The logic of extreme higher-order probabilities consists of Lewis's *VC* conditional logic, minus Strong Centering, and plus the following two axiom schemata, which I call Skyrms's axioms (Koons 2000, Appendix B):

$$(p \Rightarrow (q \Rightarrow r)) \leftrightarrow ((p \,\&\, q) \Rightarrow r)$$

$$[(p \Rightarrow \neg(q \Rightarrow r)) \,\&\, \neg((p \,\&\, q) \Rightarrow \bot)] \leftrightarrow \neg((p \,\&\, q) \Rightarrow r)$$

In both cases, "p" must be a Boolean combination of $\Rightarrow$-conditionals. The variables "q" and "r" may be replaced by any two formulas. The formula $\neg((p \,\&\, q) \Rightarrow \bot)$ expresses the joint possibility of p and q (in the sense that they don't defeasibly imply a logical absurdity, $\bot$).

Notes to Defeasible Reasoning

1. Gärdernfors proved that no non-trivial belief revision system satisfies the following five conditions:

1. The set of belief states is closed under expansions.
2. Success: A belongs to $K * A$.
3. Consistency: if K and A are both consistent, then so is $K * A$.
4. Ramsey test: $(A \Rightarrow B) \in K$ iff $B \in K * A$.
5. Preservation: if $\neg A \notin K$, then $K \subseteq K * A$.

2. There is another desirable property that is independent of the ones discussed here: *infinite conditionalization*: $C(\Gamma \cup \Delta) \subseteq Cn(\Gamma \cup C(\Delta))$. The study of nonmonotonic Tarski relations (in which the set of premises can be infinite) is still relatively undeveloped (although see Makinson 1994 and Freund, Lehmann, and Makinson 1990).

3. Karl Schlechta (Schlechta 1997) has proved that the restriction of these theorem to *finite* models is a necessary one. In order to cover the case of infinite theories, it is necessary to add yet another condition to the preferential models: a condition Schlechta calls "definability preservation" (Schlechta 1997, 76).

4. Adams's logic of high conditional probability also corresponds closely to the logic of subjunctive or counterfactual conditionals developed by Robert Stalnaker and David K. Lewis (Lewis 1973). By deleting both weak centering and the *CV* axiom (which corresponds to Rational Montonoty) from Lewis's favored logic, *VC*, Adams's logic results. Adding axiom *CV* produces a logic corresponding to the Lehmann-Magidor-Pearl 1-entailment (see below). The correspondence between counterfactual logics and preferential logics was first noted by Adam Grove (Grove 1988).

5. Admissibility has also known as "p-consistency" (Adams 1975) and "ε-consistency" (Pearl 1988). A preferential consequence relation is *admissible* if and only if, for every $\varepsilon > 0$, there exists a probability function P such that, for all propositions p and $q, p \mathrel{|\!\sim} q$ iff $P(q \mid p) \geq 1 - \varepsilon$.

6. The *entropy* of a probability function is a measure of the expectation of information. To simplify, suppose that there are finitely many models. The entropy of a function P is the sum, over all the models $\mathcal{M}$ of the product of the probability of $\mathcal{M}$ and the negative logarithm (base 2) of that probability.

7. Wayne Wobcke's (Wobcke 1995) system is an alternative to Commonsense Entailment that allows a limited amount of nesting. Both Delgrande and Asher use conditional logics that are significantly weaker than either Ernest Adam's or David Lewis's *VC* (minus Centering).

8. Under an extension of 1-entailment or the maximum entropy approach, we could defeasibly infer that $\neg q$ is not exceptional, that is, that $\neg(\top \Rightarrow q)$, unless the exceptional character of $\neg q$ follows monotonically

from our theory. (The symbol $\top$ represents any logical tautology.) From $(p \Rightarrow q)$ and $\neg(\top \Rightarrow q)$, the contraposed conditional $(\neg q \Rightarrow \neg p)$ follows monotonically (in *VC* minus centering), as the following derivation shows:

$(p \Rightarrow q)$	[Assumption]
$\neg(\top \Rightarrow q)$	[Assumption]
$(p \Rightarrow (\neg p \vee q))$	[1, right weakening]
$(\neg p \Rightarrow \neg p)$	[Reflexivity]
$(\neg p \Rightarrow (\neg p \vee q))$	[4, right weakening]
$((p \vee \neg p) \Rightarrow (\neg p \vee q))$	[3, 5, Or]
$(\top \Rightarrow (\neg p \vee q))$	[6, Left equivalence]
$((\top \,\&\, \neg q) \Rightarrow (\neg p \vee q))$	[2, 7, CV]
$(\neg q \Rightarrow (\neg p \vee q))$	[8, Left equivalence]
$(\neg q \Rightarrow \neg q)$	[Reflexivity]
$(\neg q \Rightarrow \neg p)$	[9, 10, right conjunction, right weakening]

9. Morgan also gives us good reason to reject the principle of absorption for conditionals:

$$(p \Rightarrow (q \Rightarrow r)) \leftrightarrow ((p \,\&\, q) \Rightarrow r)$$

This principle of absorption holds in the logic of extreme higher-order probabilities *only* in the special case where p consists entirely of Boolean combinations of $\Rightarrow$-conditionals.

10. Andrew Baker (Baker 1988) offers a solution to the Yale shooting problem with the resources of circumscription, by altering which predicates are allowed to vary and which are held fixed. Baker's solution is not an exception to my claim, however, since his solution involves making a critical distinction between the role of causal and non-causal information.