Department of Computer Science,
University of Otago

Technical Report OUCS-2007-01

ENTAILMENT, DUALITY, AND THE FORMS OF REASONING

Authors:

K. BRITZ
SCHOOL OF COMPUTING, UNIVERSITY OF SOUTH AFRICA, PRETORIA, SOUTH AFRICA

J. HEIDEMA
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTH AFRICA, PRETORIA, SOUTH AFRICA

W. LABUSCHAGNE
Department of Computer Science, University of Otago

Department of Computer Science,
University of Otago, PO Box 56, Dunedin, Otago, New Zealand

http://www.cs.otago.ac.nz/research/techreports.html

ENTAILMENT, DUALITY, AND THE FORMS OF REASONING

K. BRITZ, J. HEIDEMA, AND W. LABUSCHAGNE

Abstract. We explore the notion of duality for defeasible entailment relations induced by preference orderings on states. We then show that such preferential entailment relations may be used to characterise Peircean inductive and abductive reasoning. Interpreting the preference relations as accessibility relations establishes modular Gödel-Löb logic as the modal logic of inductive and abductive reasoning.

§1. Introduction. The heart of logic is *entailment* – a relation between information-bearers X and Y according to which Y in some sense follows from X. We take X and Y to be discrete strings (sentences) in some object language. The entailment relation that may hold between premiss X and consequence Y is induced by the choice of a relation E between possibly independent representations $P(X)$ and $Q(Y)$ of X and Y, where the criterion E may vary between opposing poles on at least three qualitative dimensions: syntactic to semantic; conservative to liberal; and ontological to epistemological. Our primary purpose in this paper is to discuss two kinds of entailment, $\mid\!\sim$ and $\mid\!\sim^*$, induced by semantic, liberal, epistemological criteria. In order to set the scene, we briefly indicate what we understand by the opposite poles of each dimension. Next, as a reference point, classical entailment is situated in this landscape. We are then ready to consider $\mid\!\sim$, the sort of entailment relation encountered in nonmonotonic logic, and to introduce a notion of duality relative to which every preferential entailment relation $\mid\!\sim$ has an associated dual $\mid\!\sim^*$. After examining the relationship between the new $\mid\!\sim^*$ and the familiar $\mid\!\sim$ from several points of view, we shall argue that $\mid\!\sim$ constrains inductive reasoning, and $\mid\!\sim^*$ abductive reasoning.

Let us consider in more detail the ways in which entailment criteria may vary.

The *syntactic* approach takes $P(X)$ and $Q(Y)$ to be symbol-strings derived from X and Y, while E denotes formal deduction according to a set of inference rules that pattern-match the shapes of strings, so that the criterion for "X syntactically entails Y" addresses form rather than content. A *semantic* approach, on the other hand, associates with the object language some class of extralinguistic structures; establishes a relationship between sentences and structures indicating whether a sentence is satisfied by a structure; designates certain structures as the representations $P(X)$ and $Q(Y)$; and finally nominates a relationship E between the representations, such as $\subseteq$, to induce a semantic

Key words and phrases. abduction, induction, defeasible entailment, nonmonotonic logic, preference orders.

entailment relation. In this case the syntactic forms of X and Y play no role, and it is the meanings provided by the representational structures of X and Y that are crucial.

The entailment between premiss X and consequence Y is *conservative* if a rigid constraint of truth-preservation is applied, and *liberal* if a more flexible constraint is imposed, permitting the consequence Y to be false under at least some circumstances that make the premiss X true. This simple opposition is complicated by the expressiveness of the object language to which premiss and consequence belong. A criterion giving a liberal entailment relation in the context of a nonmodal propositional language may induce a conservative entailment relation when the language is enriched by appropriate modal operators. (Indeed, we shall show that two kinds of liberal entailment relation $\mid\!\sim$ and $\mid\!\sim^*$ defined for nonmodal languages become special cases of a conservative entailment relation $\models$ in a suitable modal language.)

E embodies an *ontological* criterion if it leaves no room to take account of a reasoning agent's understanding of or information about the state of the world, i.e. if "there is no knowing or reasoning subject, even an ideal one, appealed to or legislated for" [42, p.493]. Instead, the objective structure of such states imposes a constraint such as that the same state of the world making this sentence true makes that sentence true too. For example, classical entailment (discussed in the next section) is concerned with conservative entailments sanctioned by an ontological criterion. The ontological criterion may further presuppose a metaphysical reality determining the meaning of entailment and the nature of truth. We shall return to the metaphysical vs. epistemological distinction in Section 6.2, and to the position of classical logic in this landscape in Section 2.

An *epistemological* criterion would address the question of whether a rational agent's believing X constitutes a reason for believing Y. The growth of research in artificial intelligence has firmly established the relevance of the epistemological perspective. The challenge is to design an artificial agent capable of reasoning adaptively, i.e. capable of functioning effectively in the everyday world of humans, houses, automobiles, and business. Higher-level agents (e.g. belief-desire-intention (BDI) agents) are typically seen as planning and acting on the basis of beliefs about their environment (and other agents). The agent is generally able to extract only partial information from its environment. Since, in order to achieve their goals, agents need to take decisions based on partial information, the epistemological perspective is naturally associated with liberal inferences and the risk of error. We shall explore relations E that characterise two such liberal forms of reasoning: induction and abduction.

The taxonomy above provides the setting rather than the focus of the discussion to follow, and remains largely informal. For a detailed taxonomy of entailment relations, which however does not take the ontological vs. epistemological distinction into account, the reader is referred to [5]. We note that both the conservative-liberal and the ontological-epistemological dimensions of entailment presuppose the syntactic-semantic dimension, since both involve the syntactic strings X and Y as well as their meanings and truth.

§2. Classical entailment. What do we understand by 'classical' entailment? Historically, the first formal treatment of entailment was in terms of the syntactic (as well as ontological and conservative) relation of formal deduction. Thus the classical, syntactic approach takes $P(X) = X$ and $Q(Y) = Y$ to be the symbol-strings X and Y themselves, while $E = \vdash$, deduction according to a set of syntactic inference rules. As Bertrand Russell put it a century ago, an argument "is valid in virtue of its *form*, not in virtue of the particular terms occurring in it" [79, p.197], and more generally:

> The language of mathematical logic, if it were perfected, would ... have symbols for variables, such as 'x' and 'R' and 'y', arranged in various ways; and the way of arrangement would indicate that something was being said to be true of all values or some values of the variables. We should not need to know any words, because they would only be needed for giving values to the variables, which is the business of the applied mathematician, not of the pure mathematician or logician [79, p.200].

Thus for Russell entailment, or valid argument, was concerned with the transformation of one string of symbols into another by the application of carefully selected rules. The rules were selected in order to preserve truth, thus ensuring that the entailment relation was conservative, but the meanings that might be attributed to the strings by a comprehending subject (i.e., the semantics) had no bearing on whether an argument was valid. Many different sets of inference rules have been proposed [86, 87], so that in a sense there are many syntactic entailment relations $\vdash$.

A contrasting semantic view was articulated by Charles Sanders Peirce, who wrote [65, see 2:778]:

> The third elementary way of reasoning is *deduction*, of which the warrant is that the facts presented in the premisses could not under any imaginable circumstances be true without involving the truth of the conclusion ... The process is as follows, at least in many cases: We form in the imagination some sort of diagrammatic, that is, iconic, representation of the facts, as skeletonized as possible. The impression of the present writer is that with ordinary persons this is always a visual image, or mixed visual and muscular; but this is an opinion not founded on any systematic examination ... This diagram, which has been constructed to represent intuitively or semi-intuitively the same relations which are abstractly expressed in the premisses, is then observed, and a hypothesis suggests itself that there is a certain relation between some of its parts – or perhaps this hypothesis has already been suggested.

And again [65, 5:162]:

> All necessary reasoning without exception is diagrammatic. That is, we construct an icon of our hypothetical state of things and proceed to observe it. This observation leads us to suspect that something is true, which we may or may not be able to formulate with precision, and we proceed to enquire whether it is true or not.

Following Peirce, we shall use 'deduction' for conservative entailment, but without making assumptions about the remaining qualities. Note that Peirce's semantic approach incorporates the epistemological perspective of a comprehending agent, for he explicitly refers to imagining and constructing, and the semantics involves iconic representations, which are psychological phenomena. As is well known, Peirce did not use 'symbol' as a synonym for 'anything that represents' but instead distinguished between various kinds of representation. An iconic representation differs from a symbolic representation in abstractness: symbols acquire their denotations by convention whereas an iconic representation such as a photograph, map, or blueprint resembles in some analog fashion that which it represents. For example, the image formed on the retina when we look at our friend Alice is an iconic representation of her, as is also the subsequent topographically organised pattern of neural excitation in the brain that we subjectively experience as an image of Alice; on the other hand, the string 'Alice' is a symbolic representation. Perception produces iconic representations (images) in the early sensory cortices of organisms [18, p.91]. From these an analog-to-discrete transformation, perhaps involving a form of filtering that picks out invariant features, produces a categorical representation, from which in turn may arise a symbolic representation in the form of a label associated with the categorical representation [38, 29]. We note in passing that a major challenge for artificial intelligence researchers is to design situated agents in which some version of this analog-to-discrete transformation will convert inputs from sensors (or as Harnad so memorably puts it: "the proximal projection of the distal stimulus object on the device's transducer surfaces") into sentences intended for communication or processing by a syntactic algorithm [16]. In the case of human reasoning, it was Peirce's view that our (partly unconscious) manipulation of iconic representations guides our thoughts from one symbolic representation (sentence) to another – a foreshadowing of Johnson-Laird's theory of cognition based on mental models [44, 45]. Peirce's concept of entailment was thus epistemological, semantic, and (in virtue of his warrant for deduction) conservative.

A subsequent, and better-known, semantic view of entailment was developed in a sequence of papers by Alfred Tarski [88]. Here abstract set-theoretic models take the place of the iconic representations involved in actual human thought, facilitating the formal analysis of semantic notions. This model-theoretic approach associates with the object language a class $\mathbf{W}$ of semantic structures (states, interpretations, worlds) each consisting of a domain of discourse together with functions and relations on the domain. Denotations are then specified, so that it is known what elements, functions, and relations are denoted by the constants, function symbols, and predicate symbols of the object language. A definition of what it takes for a sentence to be satisfied by such a semantic structure is given. Now every sentence X is true relative to a subset of $\mathbf{W}$, the set $Mod(X)$ of models of X, and is false relative to the remainder, making it possible to define that sentence X semantically entails sentence Y, written $X \models Y$, by taking the representations of X and Y to be $P(X) = Mod(X)$ and $Q(Y) = Mod(Y)$ and by taking the inducing relation $E = \subseteq$, set-theoretical inclusion. Peirce's warrant for deduction is again realised: $\models$ preserves truth in the sense that if X is true in a state (interpretation, world) and $X \models Y$, then Y is also true in that state (interpretation, world).

Peirce was interested in science and the discovery of truths about the real world; Tarski focused on applications in mathematics. In this latter context $\models$ has a fixed ontological foundation: to the (first-order) object language is associated the class of all semantic structures whose signature fits the vocabulary of the object language, subclasses that may be of interest then being demarcated by axiomatisation. Furthermore, it is a purely set-theoretic question whether $X \models Y$, with no room for a reasoning agent to intrude any influence. For first-order logic, there are, in fact, multiple versions of $\models$ because the notion of model can be defined either locally (model = interpretation + assignment of values to variables) or more globally (model = interpretation) [6], but for sentences the local and global notions coincide. In this limiting case, $\models$ is the same binary relation as (sensibly defined versions of) the syntactic $\vdash$, the latter being sound and complete with regard to the former. Since propositional logic affords a similar coincidence of $\models$ and $\vdash$ relative to the fixed ontology in which all assignments of truth values to atoms are included, one might argue for the existence of a single 'classical' entailment relation capable of being generated either semantically or syntactically. This classical entailment relation, in either of its incarnations, serves to characterise deduction by providing a mathematically precise static constraint against which deductive lunges from premiss to consequence may be tested.

The ontology, by which we understand the class $\mathbf{W}$ of states associated with the object language, need not be fixed at the class of all semantic structures mathematically compatible with the signature of the language (in the propositional case, need not be the class of all truth assignments). If a propositional language is used to represent knowledge about a simple system such as the 3 Card Game (in which there are three different cards and three different players, so that there are nine elementary ideas of the form "This player has that card" to express in the object language), the realisable states (legal deals) correspond to just 6 of the 512 ways to assign truth values to the nine atomic sentences. It is convenient to exclude from $\mathbf{W}$ the 506 spurious states, for example the truth assignment making all atoms true simultaneously, which would represent the impossible deal in which every player gets all three cards. Allowing flexibility in the choice of $\mathbf{W}$ may in effect inject agent-orientation, and the supply of potential models, i.e. $\mathbf{W}$, affects the semantic entailment relation $\models$, for if $\mathbf{W} \subseteq \mathbf{W}'$ then $\models_{\mathbf{W}'} \subseteq \models_{\mathbf{W}}$. A relation defined in terms of set-theoretic models may therefore be construed epistemologically even though the semantics is formal (set theoretic) rather than psychological (Peircean images).

The agent-orientation characteristic of epistemological criteria does not reduce logic to psychology. The distinction is well drawn by John Pollock, who says: "Psychologists study human thought when it goes wrong as well as when it goes right, but we want to know *what it is* for it to go right" [68, p.1]. If by *reasoning* we understand patterns of thought that are rational in the pragmatic sense of tending to go right, in the everyday world and not just in the universe of mathematics, then it would seem prudent to give Peirce a respectful hearing without necessarily discarding the technical conveniences of Tarskian model theory. An insight frequently re-iterated by Peirce was that there are three kinds of reasoning. In addition to deduction, there are induction and abduction [65, e.g. 2:774, 5:161, 5:274]. This invites the question: What entailment relations constrain induction and abduction?

In what follows we shall put forward two kinds of semantic, liberal, epistemological entailment, where 'liberal' is made precise as supraclassicality. In *supraclassical* entailment more pairs (X, Y) are allowed into the entailment relation than can be justified by the classical entailment relations $\vdash$ or $\models$. This is done by suitable choices, determined by meta-information beyond that carried by X and Y, of $P(X)$, $Q(Y)$ and E, and then stipulating "X entails Y" to mean $P(X)EQ(Y)$. The truth of premiss X does not guarantee that consequence Y is true, but knowing X constitutes a reason for believing Y at least tentatively and provisionally. The entailment is therefore *defeasible*; it represents entailment from premiss to consequence that may have a counterexample, an (unexpected) interpretation under which X is true but Y is false. The reasoning captured by such supraclassical entailment is *ampliative* inasmuch as the consequence may go beyond the information contained in the premiss X. Frameworks for the explication of ampliative reasoning are provided by *nonmonotonic logic*, in one form of which a preference order on states determines a defeasible entailment relation $\mid\!\sim$. We call this *preferential semantics*, for short [81, 49]. An overview of this and other approaches can be found in [20], while a more abstract treatment is given in [5].

We shall argue below that the relation $\mid\!\sim$ given by preferential semantics constrains inductive reasoning. For abduction, we introduce the dual relation $\mid\!\sim^*$, which embodies the idea of being a *potential partial explanation.* But before forging the links to induction and abduction, we explore the relationship between the new $\mid\!\sim^*$ and the familiar $\mid\!\sim$ from the perspectives of Boolean algebra, inference rules, and modal axiomatisation respectively.

§3. Preferential semantics. Assume that we have a propositional object language L, with which is associated a set $\mathbf{W}$ of states or possible worlds. For simplicity we shall identify states with assignments of truth values to atomic sentences, although in general the connection between states and truth assignments is more loose [49]. In a preferential semantics for L one assumes an order relation on $\mathbf{W}$ to be given. The intuitive idea captured by the order relation on states is that states higher up (greater) in the order are more preferred, more normal, more likely to occur in the context under consideration, than those lower down. For historical reasons the order is often inverted in the literature, i.e., states lower down in the order are preferred to those higher up (see, for example, [1, 19, 59]), but we will follow Shoham in taking upwards as the direction of increased preference [81, p.74], which respects the direction of accessibility relations in modal logic and observes the informal intuition that 'up' is 'better'. Indeed, this is in accordance "with a long-standing philosophical tradition" [36].

To ensure that the entailment relation eventually generated is what is often called a *rational* consequence relation [25, 51, 24], we assume the preference order to be a *modular partial order*, i.e. a reflexive, transitive relation such that, for all u, v, w in $\mathbf{W}$, if u and v are incomparable and u is strictly below w, then v is also strictly below w. Equivalently, we may take the preference order to be a total preorder, i.e. a reflexive, transitive relation such that, for all u, v in $\mathbf{W}$, u and v are comparable. Both modular partial orders and total preorders have the effect of stratifying the set $\mathbf{W}$ into layers (levels), thus facilitating visualisation, and for purposes of selecting maximally preferred models it does not matter which is used (although subsidiary concepts such as entrenchment, which are not pertinent to the present discussion, would be affected [57, 56]). We further assume that the order relation is *Noetherian* (and hence, in Shoham's terminology [81, p.75], bounded, which is the dual of well-founded, which in turn implies, in the terminology of Kraus, Lehmann, and Magidor [49], that the order relation is smooth), i.e., there is no infinite strictly ascending chain of states. In the presence of transitivity, the Noetherian property is equivalent to the following condition: For every nonempty subset $\mathbf{U}$ of $\mathbf{W}$ and $u \in \mathbf{U}$ there is an element $v \in \mathbf{U}$, maximal in $\mathbf{U}$, with v greater than or equal to u. In particular, any nonempty set $Mod(X)$ has a subset of maximal elements $MaxMod(X)$ – those models w of X for which there is *no* model of X strictly higher up in the preference order $\leq$ than w, i.e. no model strictly preferred to w.

The preference order $\leq$ on $\mathbf{W}$ permits us to expand the classical entailment relation $X \models Y$, i.e. $Mod(X) \subseteq Mod(Y)$, to a supraclassical set of pairs (X, Y) by shrinking $Mod(X)$ to a smaller set, the set $P(X)$ of most preferred models of X, capable of fitting into more different sets $Mod(Y)$. We define the defeasible entailment relation $\mid\!\sim$ by

$$X \mid\!\sim Y \text{ iff } P(X) \subseteq Mod(Y),$$

where

$$\begin{aligned} P(X) &= MaxMod(X) \\ &= \{w \in Mod(X) \mid \text{for no } w' \in Mod(X) \text{ is } w \leq w' \text{ but } w' \not\leq w\}. \end{aligned}$$

The entailment relation $\mid\!\sim$ is defeasible inasmuch as a model of X may exist which is a counterexample to $X \models Y$ but not to $X \mid\!\sim Y$. Such a model of X which is not a model of Y cannot be among the most preferred models of X, and must be somewhat "abnormal" – for the cognoscenti: like a state of the world in which Tweety is a bird but is flightless.

By way of fully formalised example, consider the Light-Fan System consisting of two components, a light and a fan, each of which may be either on or off. As knowledge representation language we take the propositional language generated by the set $\{p, q\}$ of atoms, where p is intended to express the atomic fact that the light is on and q that the fan is on. The usual connectives are assumed: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional), and $\leftrightarrow$ (biconditional). The set of states is $\mathbf{W} = \{11, 10, 01, 00\}$, where the string 10 is an abbreviation for the truth assignment making p true and q false, and so on.

As an aid to intuition we may regard the Light-Fan system as a metaphor for (say) a nuclear powerplant, with the light standing for the atomic pile and the fan for the cooling system.

The following modular partial order depicts the heuristic information that the utility company selling the electricity generated by the powerplant tries as far as possible to keep both the light and the fan on; sometimes has to switch off the light for maintenance but tries then to keep the fan going lest shutting down and starting up again cause problems; more rarely has to switch off both the light and the fan (say, if the fan needs maintenance); and only very exceptionally would have the light on while the fan is off (for example, during a serious malfunction on the order of the Three Mile Island meltdown):

11	
01	↑ more preferred
00	
10	

The defeasible entailment relation $\mid\!\sim$ induced by this order is such that, for example, $\neg p \mid\!\sim q$ since $MaxMod(\neg p) = \{01\} \subseteq \{11, 01\} = Mod(q)$. This reflects the intuition that when the light is off, it is more usual for the fan to be on than to be off. Since $\neg p \not\models q$, the relation $\mid\!\sim$ is clearly supraclassical.

It should be remarked that the relation $\mid\!\sim$ does not model statistical inference. Statistical data may provide the basis for a default rule, and thus probabilities may be the source of the original preference order, but nonmonotonic logic then proceeds in a manner subtly different from probabilistic inductive logic [39]. Suppose the order relation depicted above were derived from statistical information about the likelihood that the system is in a given state. The statistical information would be in the form of probabilities such as (for example) $\Pr(10) = \frac{1}{10}$, $\Pr(00) = \frac{2}{10}$, $\Pr(01) = \frac{3}{10}$, and $\Pr(11) = \frac{4}{10}$. In order to generalise classical entailment $\models$ to probabilistic entailment, these numbers would be used to calculate conditional probabilities [39]. For example, to determine whether $\neg p$ probabilistically entails q, one would calculate $\Pr(\{01, 11\} \mid \{00, 01\})$ where $\{01, 11\} = Mod(q)$ and $\{00, 01\} = Mod(\neg p)$. The calculation of this conditional probability in effect restricts consideration to the models of $\neg p$ and counts the proportion of these that satisfy q. In contrast, the defeasible entailment relation $\mid\!\sim$ picks out only the most preferred models of $\neg p$ and checks whether (all of) these satisfy q, a procedure requiring no arithmetical calculation and sometimes leading to different results.

§4. Duality. As noted above, classical entailment $X \models Y$ is defined by requiring that $Mod(X) \subseteq Mod(Y)$, and in preferential semantics the defeasible entailment relation $X \mid\!\sim Y$ shrinks $Mod(X)$ to $P(X)$ and requires merely that $P(X) \subseteq Mod(Y)$. Instead of shrinking $Mod(X)$ it is possible to expand the relation $\models$ by dilating $Mod(Y)$ to $Q(Y)$, for some appropriate choice of $Q(Y)$. Since $P(X)$ is the set comprising only the most preferred models of X, we take $Q(Y)$ to be the dual notion, i.e. the set of all states except for the most preferred models of $\neg Y$.

Formally, we define

$$Q(Y) = \mathbf{W} - P(\neg Y),$$

and since $\mathbf{W} - P(\neg Y) = Mod(Y) \cup [Mod(\neg Y) - P(\neg Y)]$, we may think of $Q(Y)$ as obtained by adding to $Mod(Y)$ those models of $\neg Y$ which are not maximally preferred. Now we can define the dual defeasible entailment relation $\mid\!\sim^*$ by

$$X \mid\!\sim^* Y \text{ iff } Mod(X) \subseteq Q(Y).$$

The intuition underlying this form of entailment is that, should X (against expectations) have a model not in $Mod(Y)$ (i.e. a counterexample to $X \models Y$), then this model (counterexample) is "abnormal", being not amongst the most preferred models of $\neg Y$.

Returning to the Light-Fan System for concrete illustrations, we see that $\mid\!\sim^*$ is different from $\mid\!\sim$, since $\neg(p \leftrightarrow q) \mid\!\sim^* \neg q$ whereas it is not the case that $\neg(p \leftrightarrow q) \mid\!\sim \neg q$. To see that $\neg(p \leftrightarrow q)$ does not defeasibly entail $\neg q$ under $\mid\!\sim$, it suffices to note that the models of $\neg(p \leftrightarrow q)$ are 10 and 01 of which the preferred model 01 fails to be a model of $\neg q$. That $\neg(p \leftrightarrow q) \mid\!\sim^* \neg q$ follows since $\{01, 10\} \subseteq \{00, 10\} \cup \{01\}$, where we have dilated the models of $\neg q$ by adding the model of q that is not maximally preferred. Furthermore $\mid\!\sim^*$ is not the converse of $\mid\!\sim$, for it is not the case that $\neg q \mid\!\sim \neg(p \leftrightarrow q)$, as the most preferred model of $\neg q$ is 00, which is not a model of $\neg(p \leftrightarrow q)$.

Let us now examine the extent to which $\mid\!\sim^*$ deserves to be called the dual of $\mid\!\sim$. Duality in a context such as Boolean algebra has to do with swapping 'up' and 'down' (e.g. join and meet) under complementation. Classically, $X \models Y$ iff $\neg Y \models \neg X$, where $\models$ is the order relation in the Lindenbaum-Tarski algebra of propositions, and negation is complementation. This property does not generally hold for $\mid\!\sim$, as may be seen from the counterexample $\neg p \mid\!\sim q$ in the Light-Fan System above – it is not the case that we have the contrapositive entailment $\neg q \mid\!\sim p$, since the most preferred model of $\neg q$ is 00, which fails to satisfy p. However, note that

$$\begin{aligned} \neg Y \mid\!\sim \neg X \ &\text{iff}\ \ P(\neg Y) \subseteq \mathbf{W} - Mod(X) \\ &\text{iff}\ \ Mod(X) \subseteq \mathbf{W} - P(\neg Y) \\ &\text{iff}\ \ X \mid\!\sim^* Y. \end{aligned}$$

Thus $\mid\!\sim^*$ is the dual of $\mid\!\sim$ where the operation $()^*$ on binary relations of sentences is given by $(X, Y) \in R^*$ iff $(\neg Y, \neg X) \in R$. As one would expect of a duality operation, $\neg\neg X \mid\!\sim^{**} \neg\neg Y$ iff $X \mid\!\sim Y$.

From an algebraic perspective one has that, for a fixed premiss X, the set $\{Y \mid X \mid\!\sim Y\}$ of consequences under $\mid\!\sim$ is a filter in the Lindenbaum-Tarski algebra of the language, i.e., is closed under conjunction $\wedge$ and classical entailment $\models$. (It is a principal filter if $P(X)$ is the set of models of a single sentence, as always happens for a finitely generated propositional language.) But, for a fixed consequence Y, the set of its premisses $\{X \mid X \mid\!\sim Y\}$ under $\mid\!\sim$ merits no acclamation: it is not an ideal, amongst others because $\mid\!\sim$ is *nonmonotonic*; $X \mid\!\sim Y$ does not always ensure that $X \wedge X' \mid\!\sim Y$, so the set of premisses of Y under $\mid\!\sim$ is not closed downward in the Lindenbaum-Tarski algebra.

In sharp contrast, when we change to $\mid\!\sim^*$ we find that it is the consequences $\{Y \mid X \mid\!\sim^* Y\}$ of a fixed X under $\mid\!\sim^*$ that are collectively disappointing since they do not constitute a filter in the Lindenbaum-Tarski algebra, but it is easily verified that now the set of premisses $\{X \mid X \mid\!\sim^* Y\}$ of a fixed consequence Y *do* form an ideal, in large part because $\mid\!\sim^*$ is *monotonic*. In sum, whereas the consequences under $\mathrel{|\!\sim}$ form a filter, the premisses under $\mathrel{|\!\sim}^*$ form an ideal in the Boolean algebra of propositions.

We now turn to properties of entailment that can be expressed in the form of inference rules. It was mentioned in passing that although $\mathrel{|\!\sim}$ is nonmonotonic, $\mathrel{|\!\sim}^*$ is monotonic. Monotonicity is one of the characteristic properties of classical entailment, and reflects that we may blithely strengthen the premiss on the left hand side of an entailment: if $X \models Z$ then, for any sentence Y, it is also the case that $X \wedge Y \models Z$. Since the premisses under $\mathrel{|\!\sim}$ do not form an ideal, the best we can hope for is that some more careful form of left strengthening may hold, in which some safety criterion filters out unsuitable additional premisses Y. Two such versions are given below, cautious monotonicity and rational monotonicity respectively.

The following are a sample of the properties known from [49, 51] to typify $\mathrel{|\!\sim}$.

(1) Reflexivity: $X \mathrel{|\!\sim} X$

(2) And: $$\frac{X \mathrel{|\!\sim} Y \qquad X \mathrel{|\!\sim} Z}{X \mathrel{|\!\sim} Y \wedge Z}$$

(3) Or: $$\frac{X \mathrel{|\!\sim} Z \qquad Y \mathrel{|\!\sim} Z}{X \vee Y \mathrel{|\!\sim} Z}$$

(4) Left defeasible equivalence: $$\frac{X \mathrel{|\!\sim} Y \quad Y \mathrel{|\!\sim} X \quad X \mathrel{|\!\sim} Z}{Y \mathrel{|\!\sim} Z}$$

(4′) Left logical equivalence: $$\frac{\models X \leftrightarrow Y \qquad X \mathrel{|\!\sim} Z}{Y \mathrel{|\!\sim} Z}$$

(5) Right weakening: $$\frac{X \mathrel{|\!\sim} Y \qquad \models Y \rightarrow Z}{X \mathrel{|\!\sim} Z}$$

(6) Cautious monotonicity:
(Cautious left strengthening) $$\frac{X \mathrel{|\!\sim} Y \qquad X \mathrel{|\!\sim} Z}{X \wedge Y \mathrel{|\!\sim} Z}$$

(7) Rational monotonicity:
(Rational left strengthening) $$\frac{X \mathrel{|\!\sim} Y \qquad X \not\mathrel{|\!\sim} \neg Z}{X \wedge Z \mathrel{|\!\sim} Y}$$

(8) Cut:
(Cautious left weakening) $$\frac{X \wedge Y \mathrel{|\!\sim} Z \qquad X \mathrel{|\!\sim} Y}{X \mathrel{|\!\sim} Z}$$

Since $\mathrel{|\!\sim}$ is supraclassical, the premisses for left defeasible equivalence are weaker than those for left logical equivalence.

The duality between ideals and filters in Boolean algebra leads us to expect that the inference rules that hold for $\mathrel{|\!\sim}^*$ ought to form matching pairs with those for $\mathrel{|\!\sim}$, differing in whether it is the left or the right that is altered and whether the alteration amounts to weakening or strengthening (in the sense of respectively dilating or contracting the set of models). The property of right weakening for $\mathrel{|\!\sim}$ ought to be replaced by left strengthening (monotonicity) for $\mathrel{|\!\sim}^*$, while the careful forms of monotonicity for $\mathrel{|\!\sim}$ have as their analogs careful forms of right weakening for $\mathrel{|\!\sim}^*$, and the cut rule for $\mathrel{|\!\sim}$, being a cautious form of left weakening, ought to have as its analog a cautious form of right strengthening. And indeed, each of the following inference rules holds for $\mathrel{|\!\sim}^*$, since it is (via the definition of $\mathrel{|\!\sim}^*$) equivalent to the correspondingly numbered rule for $\mathrel{|\!\sim}$.

(1*) Reflexivity: $X \mathrel{|\!\!\sim}^* X$

(2*) Or: $\dfrac{X \mathrel{|\!\!\sim}^* Z \qquad Y \mathrel{|\!\!\sim}^* Z}{X \vee Y \mathrel{|\!\!\sim}^* Z}$

(3*) And: $\dfrac{X \mathrel{|\!\!\sim}^* Y \qquad X \mathrel{|\!\!\sim}^* Z}{X \mathrel{|\!\!\sim}^* Y \wedge Z}$

(4*) Right defeasible equivalence: $\dfrac{X \mathrel{|\!\!\sim}^* Y \quad Y \mathrel{|\!\!\sim}^* X \quad Z \mathrel{|\!\!\sim}^* X}{Z \mathrel{|\!\!\sim}^* Y}$

(4*′) Right logical equivalence: $\dfrac{\models X \leftrightarrow Y \qquad Z \mathrel{|\!\!\sim}^* X}{Z \mathrel{|\!\!\sim}^* Y}$

(5*) Monotonicity:
(Left strengthening) $\dfrac{Y \mathrel{|\!\!\sim}^* Z \qquad \models X \to Y}{X \mathrel{|\!\!\sim}^* Z}$

(6*) Cautious right weakening: $\dfrac{X \mathrel{|\!\!\sim}^* Y \qquad Z \mathrel{|\!\!\sim}^* Y}{X \mathrel{|\!\!\sim}^* Y \vee Z}$

(7*) Rational right weakening: $\dfrac{X \mathrel{|\!\!\sim}^* Y \qquad \neg Z \not\mathrel{|\!\!\sim}^* Y}{X \mathrel{|\!\!\sim}^* Y \vee Z}$

(8*) Cautious right strengthening: $\dfrac{X \mathrel{|\!\!\sim}^* Y \vee Z \qquad Z \mathrel{|\!\!\sim}^* Y}{X \mathrel{|\!\!\sim}^* Y}$

Since $\mathrel{|\!\!\sim}^*$ is supraclassical, the premisses for right defeasible equivalence are weaker than those for right logical equivalence. The form in which monotonicity for $\mathrel{|\!\!\sim}^*$ is given has been chosen to rhyme with the form in which right weakening for $\mathrel{|\!\!\sim}$ was given previously. We could of course recast monotonicity for $\mathrel{|\!\!\sim}^*$ in the equivalent but more familiar garb whereby the entailment $X \mathrel{|\!\!\sim}^* Z$ is blithely strengthened on the left to $X \wedge Y \mathrel{|\!\!\sim}^* Z$. In a similar vein, it may help to understand the two careful forms of right weakening for $\mathrel{|\!\!\sim}^*$ if the reader first recasts the (blithely unguarded) right weakening rule for $\mathrel{|\!\!\sim}$ as the equivalent assertion that if $X \mathrel{|\!\!\sim} Y$ then, for any Z whatever, $X \mathrel{|\!\!\sim} Y \vee Z$.

Since $\mathrel{|\!\!\sim}$ and $\mathrel{|\!\!\sim}^*$ are respectively associated with filters and ideals of the Lindenbaum-Tarski algebra of propositions and since the inference rules governing the behaviour of $\mathrel{|\!\!\sim}$ have intuitively natural dual versions that apply to $\mathrel{|\!\!\sim}^*$, it is clear that $\mathrel{|\!\!\sim}^*$ may justifiably be termed the dual of $\mathrel{|\!\!\sim}$. Further evidence is found in modal logic, where rational monotonicity for $\mathrel{|\!\!\sim}$ and rational right weakening for $\mathrel{|\!\!\sim}^*$ require a tense logic with bidirectional frames.

§5. Modal characterisation. Our relation $\mathrel{|\!\!\sim}$ is produced by a preference order on $\mathbf{W}$ with very specific properties – a modular partial order (or, equivalently, a total preorder). In contrast, an accessibility relation for a modal logic can, in general, be *any* binary relation R on $\mathbf{W}$. This invites the conjecture that the defeasible entailment relation $\mathrel{|\!\!\sim}$ can be characterised modally by constructing an apt accessibility relation R from the preference order $\leq$ and then formulating a modal sentence $P(X)$ which, with semantics relative to R, describes the preferred models (relative to $\leq$) of premiss X:

$$Mod(P(X)) = MaxMod(X).$$

5.1. Modal formulation. Unlike in the previous sections, where $P(X)$ was defined semantically as a set of interpretations, here we want a syntactic representation of X that matches a given semantic construction. There are several ways in which this can be done, as for example in [17, 11, 4, 32, 93]. The proposal of [32], also characterised semantically in [93], is particularly natural and elegant, and we adopt it below. Note that the modal characterisation of default reasoning presented in [4] is not compatible with our proposal. For example, seriality (for every u there is some v such that $u < v$, axiomatised by $\Box X \to \neg\Box\neg X$) is argued to be mandatory in a minimal modal logic for default logic, whereas it never holds in Modular **GL** as defined below. (Seriality is not compatible with a Noetherian strict partial order as accessibility relation.)

As a first step, let us expand the (nonmodal propositional) object language L by including the sentences formed by application of a modal operator $\Box$. Call the resulting language $L^{\Box}$. Given a preference order $\leq$ on the set of states $\mathbf{W}$, which we assume to be at least a preorder (reflexive, transitive relation), take the accessibility relation R to be $<$, the strict partial order (irreflexive, transitive relation) corresponding to $\leq$:

$$(\forall w)(\forall w')[w < w' \text{ iff } w \leq w' \text{ and } w' \not\leq w].$$

Then define

$$P(X) = X \wedge \Box\neg X,$$

which is logically stronger than X. According to the semantics induced by R, the sentence $\Box\neg X$ is true in state w if and only if X is false in all w' such that $w < w'$. Hence $P(X)$ is true in all the maximal models of X, and false in all other states (non-models of X as well as non-maximal models of X):

$$Mod(P(X)) = MaxMod(X); \text{ and } X \mathrel{|\!\!\sim} Y \text{ iff } P(X) \models Y.$$

For the dual $\mathrel{|\!\!\sim}^*$ of $\mathrel{|\!\!\sim}$ we see that

$$\begin{aligned} X \mathrel{|\!\!\sim}^* Y \ &\text{iff } \neg Y \mathrel{|\!\!\sim} \neg X \\ &\text{iff } \neg Y \wedge \Box Y \models \neg X \\ &\text{iff } X \models Q(Y), \end{aligned}$$

where

$$Q(Y) = \Box Y \to Y.$$

$Q(Y)$ is not in general a tautology, since $<$ is irreflexive, but $Q(Y)$ is logically weaker than Y.

This shows that the defeasible entailments $X \mathrel{|\!\!\sim} Y$ and $X \mathrel{|\!\!\sim}^* Y$ are equivalent to entailments $P(X) \models Y$ and $X \models Q(Y)$, with $P(X) = X \wedge \Box\neg X$ and $Q(Y) = Y \vee \neg\Box Y$ in the modal language $L^{\Box}$ with semantics induced by that accessibility relation which is the strict variant of the preference order. Note that in the non-modal language L for preferential logic in Section 3 the filter $\{Y \mid X \mathrel{|\!\!\sim} Y\}$ need not be principal, since $MaxMod(X)$ need not be axiomatisable by a single sentence; similarly the ideal $\{X \mid X \mathrel{|\!\!\sim}^* Y\}$ need not be principal. However, in the more expressive modal language $L^{\Box}$ of this section the corresponding filter and ideal are principal, generated respectively by the single modal sentences $P(X)$ and $Q(Y)$.

5.2. Modal axiomatisation. We now turn to the axiomatisation of the accessibility relation $<$ which was used to give an appropriate semantics to the sentence $P(X) = X \wedge \Box\neg X$. In its most general form, $<$ is a Noetherian strict partial order.

Gödel-Löb logic **GL** (a provability logic) is obtained from the minimal modal logic **K** by adding the transitivity and Löb axioms (see, for example, [10]):

$$\mathbf{GL} = \mathbf{K} \ \oplus \ \Box X \to \Box\Box X \ \oplus \ \Box(\Box X \to X) \to \Box X.$$

GL is determined by the class of all Noetherian strict partial orders [80]. This makes **GL** the appropriate logic to reason syntactically about $\mid\!\sim$.

The emphasis in [32] is on the development of a tableau proof system for preferential semantics. Our concern here is not with the development of a decision procedure, but rather with the modal representation of modularity (or connectedness, if we are using a total preorder). In fact, we only need to consider modularity, as any total preorder can be converted to an associated modular partial order via its associated strict partial order: Let preorders S and T be *order-equivalent* iff they have the same associated strict partial orders. It is then not difficult to show that, for any modular partial order S, there is a total preorder T such that S and T are order-equivalent, and conversely, for every total preorder T there is a modular partial order S such that S and T are order-equivalent. So one can move without loss of information (about maximal elements) from a total preorder to a modular partial order via the modal representation of information in terms of a strict accessibility relation.

Linearity, and trichotomy, cannot be axiomatised directly in modal logic, and neither can modularity. However, just as linearity may be weakened to prohibit branching to the right (see, for example, [8, p.193]), modularity can be weakened to prohibit upward-branching over more than one level. Let an order relation be *weakly modular* if

$$(\forall w)(\forall u)(\forall v)[\text{If } w < v \text{ and } w < u \text{ then } v < u \text{ or } u < v \text{ or } \ \uparrow v = \ \uparrow u],$$

where $\uparrow v = \{v' \mid v < v'\}$ is the strict upclosure of v.

The axiomatisation of weak modularity in Lemma 5.2 below requires bolstering the syntax of the non-modal language L of Section 3 to allow conjunctions and disjunctions of sets of sentences with arbitrary cardinality [47, 21], and then adding the modal operator $\Box$. Of course, this is superfluous in a finitely generated propositional language. It is not clear whether Lemma 5.2 can be proved without referring to either of these conditions.

LEMMA 5.1. *For any* $\boldsymbol{U} \subseteq \boldsymbol{W}$ *there exists a sentence* X *such that* $\boldsymbol{U} = Mod(X)$.

PROOF. Suppose $\mathbf{U} = \{u_i \mid i \in I\}$. Define the "diagram" of u_i:

$$D(u_i) = \bigwedge\{l \mid l \text{ is a literal true in } u_i\},$$

where a literal is a propositional atom or its negation. Then it is easy to see that

$$X = \bigvee\{D(u_i) \mid i \in I\}$$

suffices. ⊣

The axiomatisation of weak modularity is then as follows:

Lemma 5.2. *Let $(\mathbf{W}, <)$ be a **GL**-frame (i.e. a Noetherian strict partial order). $(\mathbf{W}, <)$ is weakly modular if and only if any sentence of the form*

$$\Box(\Box X \to Y) \vee \Box(\Box Y \to \Box X)$$

is valid in the frame.

Proof. Left to right: Suppose $\Box(\Box X \to Y) \vee \Box(\Box Y \to \Box X)$ is false in world w. Then

(i) there exists u such that $w < u$ and $\Box X \to Y$ is false in u, i.e. $\Box X$ is true in u and Y is false in u, and

(ii) there exists v such that $w < v$ and $\Box Y \to \Box X$ is false in v, i.e. $\Box Y$ is true in v and $\Box X$ is false in v.

Suppose also that $<$ is weakly modular. Then $v < u$ or $u < v$ or $\uparrow v = \uparrow u$. If $v < u$, then $\Box Y$ is true in v, so Y is true in u, contradicting (i). If $u < v$, then $\Box X$ is true in u, so $\Box X$ is true in v, contradicting (ii). If $\uparrow v = \uparrow u$, then $\Box X$ is true in u, so $\Box X$ is true in v, contradicting (ii). Therefore weak modularity of the frame implies validity of $\Box(\Box X \to Y) \vee \Box(\Box Y \to \Box X)$.

Conversely, suppose there exist worlds u, v, w such that $w < v$ and $w < u$ and not($v < u$ or $u < v$ or $\uparrow v = \uparrow u$). Without loss of generality, we can assume that there exists a world z such that $v < z$ and $u \not< z$.

Let Y be any sentence true in $\uparrow v$ and false in u, and let X be any sentence true in $\uparrow u$ and false in z. The existence of X and Y is guaranteed by Lemma 5.1, since we can for instance choose X and Y to satisfy $\uparrow u = Mod(X)$ and $\uparrow v = Mod(Y)$. We see that $\Box X$ is true in u, Y is false in u, $\Box Y$ is true in v and $\Box X$ is false in v. Therefore $\Box X \to Y$ is false in u and $\Box Y \to \Box X$ is false in v. So $\Box(\Box X \to Y) \vee \Box(\Box Y \to \Box X)$ is not true in w, and hence not valid. Therefore validity of the axiom implies weak modularity. ⊣

Let *Weakly Modular **GL*** be the logic obtained from **GL** by adding the weak modularity axiom (and, if necessary, infinitary conjunctions and disjunctions in the underlying purely propositional language). We then have the following result:

Theorem 5.3. *Weakly Modular **GL** is determined by the class of all Noetherian weakly modular strict partial orders.*

Weak modularity of a **GL**-frame does not suffice to ensure rational monotonicity of its associated defeasible entailment relation $\mid\!\sim$, nor does it ensure rational right weakening of $\mid\!\sim^*$. In order to obtain such a result, we turn Weakly Modular **GL** into a *tense logic*. The class of frames under consideration here is the class of bidirectional frames of the form $(\mathbf{W}, <, >)$, where $<$ is a Noetherian weakly modular strict partial order with weakly modular converse $>$.

Let *Modular **GL*** be the tense logic with modal operators $\Box$ and $\Box^c$ obtained from the minimal tense logic $\mathbf{K}_t$ [8, p.205] by adding the transitivity, Löb, and weak modularity axioms for $\Box$, as well as the weak modularity axiom for $\Box^c$. We then have the following result, which is proved similarly to Lemma 5.2:

Theorem 5.4. *Modular **GL** is determined by the class of all Noetherian weakly modular strict partial orders with weakly modular converse.*

Any Noetherian modular strict partial order is weakly modular with weakly modular converse, and hence a Modular **GL**-frame as determined by Theorem 5.4. However, the converse statement is not true in general. Weak modularity of $<$ prohibits upward branching over more than one level. Similarly, weak modularity of $>$ prohibits downward branching over more than one level. Together, the weak modularity axioms ensure that, for any two worlds x and y, either $x < y$ or $y < x$ or $\uparrow x = \uparrow y$ or x and y have no common upper or lower bound. That is, either both x and y are elements of the same modular strict partial order, or they belong to two disjoint modular strict partial orders. We use this observation to obtain the following representation result:

Theorem 5.5. *Modular **GL** is determined by the class of all Noetherian modular strict partial orders.*

Proof. Suppose X is valid in Modular **GL**. Then X is true in all Noetherian weakly modular strict partial orders with weakly modular converse, and hence also in all Noetherian modular strict partial orders. Conversely, suppose X is true in all Noetherian modular strict partial orders. Let $(\mathbf{W}, <, >)$ be a Noetherian weakly modular strict partial order with weakly modular converse. $(\mathbf{W}, <, >)$ consists of one or more disjoint Noetherian modular strict partial orders. X is true in each of these subframes, and hence also in $(\mathbf{W}, <, >)$. So X is valid in Modular **GL**. $\dashv$

Thus Modular **GL** is the tense logic of the forms of reasoning constrained by $\mid\!\sim$ and $\mid\!\sim^*$. We now proceed to a closer examination of these forms of reasoning, and argue that $\mid\!\sim$ characterises inductive reasoning while $\mid\!\sim^*$ characterises abductive reasoning.

§6. Inductive reasoning.

Peirce claimed that reasoning was of three kinds: deductive, inductive, and abductive. Deductive reasoning is comparatively well understood. But what is inductive reasoning? Popper argued that there is, in fact, no such thing [71]. Is it possible that this contradiction between the Peircean and the Popperian views can be resolved with the help of $\mid\!\sim$? Such is our contention.

At its simplest, induction surely has to do with making predictions on the basis of past experience, for example predicting on the basis of experience summarised by the premiss "The bread which I formerly eat, nourished me" that "Other bread must also nourish me", to quote Hume [43, from para.29]. Hume drew attention to the fact that such predictive inferences are not truth-preserving, i.e. are not instances of deduction: "That there are no demonstrative arguments in the case seems evident; since it implies no contradiction that the course of nature may change, and that an object, seemingly like those which we have experienced, may be attended with different or contrary effects."

Popper's subsequent explicit formulation of the problem of induction was very broad [71, p.110]:

> Are we rationally justified in reasoning from instances or from counterinstances of which we have had experience to the truth or falsity of the corresponding laws, or to instances of which we have had no experience?

By including counterinstances as well as instances, and directing the inference toward laws as well as novel instances, Popper could rely on Hume's arguments against any claim that the inference from instances is truth-preserving, while putting forward falsification as the closest thing to a solution, because a counterinstance suffices to falsify a law, at least if that law is expressed as a universally quantified sentence. (Actually, the matter is not as simple as Popper thought, because a contingent sentence can only be falsified in some specific interpretation of the object language. Despite his respect for Tarski's model-theoretic initiatives, Popper never quite adopted the corresponding methodology. But this is not crucial in what follows.)

6.1. Peircean induction. Peirce would have granted no Popperian broadening of the notion of induction, for laws (theories) would in Peirce's view result from the process of hypothesis-formation he termed abduction. According to Peirce, induction *uses* laws rather than *produces* laws [65, 5:170]:

> Induction consists in starting from a theory, deducing from it predictions of phenomena, and observing those phenomena in order to see how nearly they agree with the theory.

Here the word "deducing" should be read as a careless synonym for 'inferring', for Peirce was quite clear about deductive reasoning being necessary (i.e. apodictic) reasoning whereas induction is associated with predictions such as might be used to design experiments by which to test theories. Induction thus involves defeasible, not necessary, reasoning (although Peirce used the term 'probable' rather than the modern 'defeasible') [65, 5:270, 5:272]:

> A complete, simple, and valid argument, or syllogism, is either *apodictic* or *probable.* An apodictic or deductive syllogism is one whose validity depends unconditionally upon the relation of the fact inferred to the facts posited in the premisses ... But a syllogism whose validity depends partly upon the *non-existence* of some other knowledge, is a *probable* syllogism ... The absence of knowledge which is essential to the validity of any probable argument relates to some question which is determined by the argument itself. This question, like every other, is whether certain objects have certain characters [i.e. features]. Hence, the absence of knowledge is either whether besides the objects which, according to the premisses, possess certain characters, any other objects possess them; or, whether besides the characters which, according to the premisses, belong to certain objects, any other characters not necessarily involved in these belong to the same objects. In the former case, the reasoning proceeds as though all the objects which have certain characters were known, and this is *induction*; in the latter case, the inference proceeds as though all the characters requisite to the determination of a certain object or class were known, and this is *hypothesis* [i.e. abduction].

Peirce explicitly rejected the commonly held view that induction led from instances to generalisations [65, 2:775]:

> *Induction* takes place when the reasoner already holds a theory more or less problematically (ranging from a pure interrogative apprehension to a strong leaning mixed with ever so little doubt); and having reflected that if that theory be true, then under certain conditions certain phenomena ought to appear (the stranger and less antecedently credible the better), proceeds to *experiment*, that is, to realize those conditions and watch for the predicted phenomena ... But all of the above is at variance with the doctrines of almost all logicians; and, in particular, they commonly teach that the inductive conclusion approximates to the truth because of the uniformity of nature. They only contemplate as inductive reasoning cases in which, from finding that certain individuals of a class have certain characters the reasoner concludes that every single individual individual of the class has the same character.

From Peirce's remarks it is clear that his view of induction encompasses, though perhaps is not confined to, the kind of thought process that cognitive psychologists have studied as 'categorical induction' since the publication of Rips's seminal 1975 paper [73, 64, 61, 62, 82]. The Rips paradigm, on which many of these studies are based, involves describing a property that applies to one or more category members and then asking subjects to make judgments about whether that property is also true of another category member. For example, the premiss might be "Sparrows use serotonin as a neurotransmitter" and the prediction might be "Crows use serotonin as a neurotransmitter". One of the surprising discoveries has been that natural categories have a typicality structure: a mathematical class is defined by necessary and sufficient conditions and thus all members have equivalent status; in contrast, the concept of a natural category is acquired by exposure to instances from which a prototype is abstracted, and now some members are more typical members than others because they have a closer similarity to the prototype [76]. This typicality structure determines the readiness with which subjects make predictive inferences. For example, in the category of birds (i.e. of bird genera), sparrows are more typical than penguins [75]. And subjects are much more willing to infer from "Sparrows use serotonin as a neurotransmitter" that some other kind of bird, such as crows, do likewise than if the premiss were "Penguins use serotonin as a neurotransmitter" [64].

Research on categorical induction in psychology provides experimental evidence that people do use a form of reasoning that matches what Peirce called inductive reasoning. Furthermore, the engine that drives the reasoning is the typicality structure of the relevant category, which may be construed as a *default rule*: If the prototypical members of the category have a property, and if a member of the category *which is not known to be atypical* is presented, then it is reasonable to believe that this member will also have the property.

We suggest that it is this typicality structure of categories, or equally the default rules expressing the structure, that provide the 'medium' Hume sought [43, para.29]:

> The bread which I formerly eat, nourished me ... but does it follow, that other bread must also nourish me at another time[?] ... There

> is required a medium, which may enable the mind to draw such an inference, if indeed it be drawn by reasoning and argument. What that medium is, I must confess, passes my comprehension ...

The normative, as opposed to descriptive, study of reasoning based on default rules is nonmonotonic logic, which grew out of work in artificial intelligence during the 1970s and in which the canonical example involves categorical induction [31, Chapter 1]:

> It is not the case that *all* birds fly, but only that *typically* birds fly. Now if I tell you that Tweety is a bird, it is reasonable for you to conclude that Tweety can fly, since you have no evidence that Tweety is atypical. But there is a crucial difference here: You based your conclusion that Tweety could fly on an *absence* of information about Tweety's atypicality. On learning something new about Tweety (such as that he has a foot set in concrete), you will need to flexibly revise your conclusion that he can fly.

The comprehensive survey of nonmonotonic logic (as it stood in 1986) from which this passage was excerpted cites no references to the psychological literature, presumably because interest in the field of categorical induction lay dormant after Rips's 1975 paper until developmental studies by Gelman and Markman [30] showed that inductive reasoning was not something children mastered immediately, but something more complex and interesting [61, p.247]. It is remarkable that the study of inductive reasoning became an active research area simultaneously but independently in two different disciplines.

The early roots of nonmonotonic logic exemplified different ways to represent default rules. The preferential semantics proposed by Shoham [81] generalises several of these, and in the developed form discussed by Kraus, Lehmann, and Magidor [49, 51] may be taken to be the most general model-theoretic framework for nonmonotonic logic that is known. As indicated earlier, preferential semantics represents default rules as order relations on a set of states. The defeasible entailment relation $\mid\!\sim$ generated by the ordering is designed to capture precisely the sort of prediction that constitutes categorical induction. We propose, therefore, that inductive reasoning is characterised by $\mid\!\sim$ in the sense that $\mid\!\sim$ provides a static constraint against which inductive lunges from premiss to conclusion may be tested.

6.2. Popperian skepticism. Inductive reasoning is a psychological fact, thus Popper's denial must have been aimed at claims that such reasoning could be formalised in a logical system. Hume's argument's sufficed to show that classical entailment did not go far enough, and so Popper's chief concern was to deny the explication of induction via Carnap's probabilistic inductive logic. A decades-long controversy resulted [58, 50, 53, 48].

In his address to the Karl Popper Centenary Congress, Vienna 2002, Musgrave resolved the Carnap-Popper controversy by drawing a distinction between metaphysical and epistemological entailment [63, p.28]. This distinction is also relevant to another formal approach to induction – that of Solomonoff [84, 85], who uses probability distributions induced by a notion of complexity based on program lengths, and appeals to the metaphysical principle that Nature has a preference for simple programs.

A metaphysical principle such as the uniformity of Nature, or her preference for simple programs, is required in order to underpin the assertion that a premiss, summarising our limited experience, renders *probable* a prediction about a new instance. Such metaphysical principles may be contrasted with epistemic principles such as CR: The fact that a hypothesis is well-corroborated is a good reason to adopt it, tentatively, as true. To demonstrate that CR is not a metaphysical principle, and that it is reasonable to adopt CR, Musgrave argues as follows:

> But epistemology is one thing, metaphysics is another. Call CR an 'epistemic inductive principle', if you like. This principle neither implies nor assumes that well-corroborated theories are true, or more likely to be true than not. It says only that it is reasonable to adopt such theories as true. Can we separate epistemology and metaphysics? ... Can it be reasonable to believe falsely? If it can, then the (epistemic) reason for believing falsely cannot be a conclusive (metaphysical) reason for what is believed. Everybody will agree that any reasonable theory of reasonable belief must make room for reasonable yet false belief. Everybody will agree, too, that if the state of the critical discussion changes, and we find a reason to think something false, then it is no longer reasonable to adopt it as true. What we say in such cases is that what we reasonably believed turned out to be wrong, not that it was wrong or unreasonable for us to have believed it ... Surely, we must show that the method of corroboration is a reliable method, which produces more truth beliefs than false ones, if it is to be a rational method? No. [J]ust as a belief need not be true, or shown true, to be reasonable, so also a method of forming beliefs need not be reliable, or shown reliable, to be reasonable.

Do the probabilistic entailment relations of inductive logic incorporate any metaphysical criteria? There are many views of probability – one thinks of Good's paper titled "46656 varieties of Bayesians" [35] – but at least some approaches to probabilistic inference do exemplify a system of liberal inferences sanctioned by metaphysical criteria. However, if the only Popperian objection to inductive logic were that it rests on a metaphysical assumption, we could simply point out that the objection does not carry over to the explication of inductive reasoning offered by nonmonotonic logic in general and $\mid\!\sim$ in particular, for $\mid\!\sim$ is unabashedly epistemological and acquires its peculiar force from the agent's default rule, which is something "well-corroborated" that the agent has learnt from the environment. The epistemological roots of preferential semantics become even clearer when we note that these can be traced back to a paper by John McCarthy tellingly titled "Epistemological problems of artificial intelligence" [54].

Still, there is a second criticism to consider. Popper and Miller wrote dismissively: "All probabilistic support is purely deductive" [72]. Can the accusation be levelled at $\mid\!\sim$ that it is merely deduction in disguise? The answer depends on whether only *flat* deduction is meant, i.e. deduction in a nonmodal context, where no additional semantic structure on the set $\mathbf{W}$ of states plays a role in the definition of satisfaction. We claim that inductive reasoning, which we have explicated in terms of $\mid\!\sim$, cannot be fully explicated in terms of flat deduction. Preferential semantics allows a proper generalisation of flat deduction and is indispensable for the modelling of rational belief formation and change. To show this, we first relate $\models$, $\mid\!\sim$, and $\mid\!\sim^*$.

Consider information expressed by a sentence or theory Z in the (nonmodal propositional) object language L. Assume that Z is non-trivial in the sense of being neither contradictory nor tautological. Then the sentence Z all by itself induces a preference order on $\mathbf{W}$ which stratifies $\mathbf{W}$ into two levels, as follows: the models of Z form the upper layer, and the nonmodels of Z form the lower layer, i.e., every model of Z is strictly preferred to every nonmodel. This trivial preference order induces entailment relations $\mid\!\sim_Z$ and $\mid\!\sim^*_Z$. Now:

THEOREM 6.1. *For any sentences X and Y such that X is consistent with Z and $\neg Y$ is also consistent with Z (i.e., Y is not classically entailed by Z alone), the following are equivalent:*

1. $Z \wedge X \models Y$
2. $X \mid\!\sim_Z Y$
3. $X \mid\!\sim^*_Z Y$

PROOF. $1 \Rightarrow 2$: Suppose $Z \wedge X \models Y$. Then $P(X) = Mod(Z \wedge X) \subseteq Mod(Y)$, so $X \mid\!\sim_Z Y$.

$2 \Rightarrow 1$: Simply reverse the steps above.

Now 1 is equivalent to $1'$. $Z \wedge \neg Y \models \neg X$. We shall show that $1' \Leftrightarrow 3$:

$1'$ says that $Mod(Z \wedge \neg Y) \subseteq Mod(\neg X)$, i.e. that $Mod(X)$ has no element in common with $Mod(Z \wedge \neg Y)$. But since $Mod(Z \wedge \neg Y)$ is not empty, it is $P(\neg Y)$. That $Mod(X)$ has no element in common with $P(\neg Y)$ is precisely what 3 claims. ⊣

Thus $\mid\!\sim$ and $\mid\!\sim^*$ are very natural generalisations of the limiting case $\models$, and it should be evident that these generalisations are proper. For whereas the information in any sentence Z corresponds semantically to the trivial dichotomous preference order that prefers every model of Z to every nonmodel of Z, preference orders that stratify $\mathbf{W}$ into more than two layers cannot in general be replaced by sentences. This follows easily from a consideration of the cardinalities involved – there are far more preference orders than (sets of) sentences.

Of course, we concede that if one expands the object language L to the more expressive modal language $L^\Box$ of Section 5 (or the subsequent tense logic) with the concomitant characterisations of the preferential entailments, then all instances of $X \mid\!\sim Y$ or $X \mid\!\sim^* Y$ with respect to any preference order whatsoever are equivalent to the corresponding 'classical' $P(X) \models Y$ or $X \models Q(Y)$, respectively. However, this is just preferential semantics in a different guise – we are again making use of meta-information represented by an ordering on states, and in this context deduction is not flat. Popper may well have been willing to extend his notion of deduction to the non-flat entailment relation $\models$ in the context of $L^\Box$ or tense logic, since he seems to have regarded possible worlds semantics as a natural extension of Tarski's ideas. For example, his definition of "naturally or physically necessary" statements is formulated in terms of "satisfied in all worlds that differ from our world, if at all, only with respect to initial conditions" [69, p.433]. It may be that the modal languages of the previous section bridge the conceptual gap between the Popperian conviction that deduction renders induction unnecessary, on one side, and, on the other, the explication of inductive reasoning rendered possible by preferential semantics and $\mid\!\sim$. Induction may indeed be seen as one form of non-flat deduction.

Whether expressed in a modal language or not, we claim that the meta-information made available by preferential semantics is needed for any adequate treatment of belief formation and change. While belief formation (via reasoning) may be characterised by suitable entailment relations, belief change goes beyond entailment. Beliefs are not surrendered with equal readiness in the face of new information; some beliefs are more entrenched than others [1, 57]. The entrenchment relations that may hold between beliefs derive from preference orderings, and if these stratify the set of states into multiple layers then the nuances are reflected in the entrenchment relation. Hence flat deduction (and dichotomous orderings) cannot suffice.

But can $\mid\!\sim$ not simply be reduced to the $\models$ of flat deduction by augmenting the premiss in a suitable manner, say by representing the default rule as a sentence in the object language?

The first thing to note is that in our everyday decision-making (and even in science [13]), we make predictions by applying rules that cannot baldly be expressed by universal sentences. We drive to work every morning by relying on a default rule of the form "Drivers normally stop at red lights". Without this rule, one would hardly dare to cross an intersection. And yet we all know that exceptions exist, such as drunk drivers. Therefore it is not the case that we predict the behaviour of an oncoming driver by deducing from the universal sentence "Drivers always stop at red lights" that this particular driver will stop when the light turns red.

Perhaps it is possible to replace the quantifier "For all" by a quantifier like "For most", which tolerates exceptions? Mostowski developed the semantics of such generalised quantifiers [60, 28], but in their original form these do not deliver satisfactory default rules. In simple terms, if the domain of an interpretation has n objects, then a sentence of the form "For most x, such-and-such" is satisfied if the formula "such-and-such" with free variable x is satisfied by more than $\frac{n}{2}$ of the possible assignments of values in the domain to the variable x. But consider the sentence "Most cats have three legs", which would be formalised along the lines of "For most x, if x is a cat then x has three legs". And consider an interpretation whose domain has ten objects of which, say, four are perfectly ordinary four-legged cats and the remaining six objects are sheep. Due to the peculiarities of the material conditional, the formula "if x is a cat then x has three legs" is satisfied by more than half the objects (the six sheep rendering the antecedent false). Counter-intuitively, the supposed default rule "Most cats have three legs" would be true in a domain (possible world) having several cats all of which have four legs. We conclude that the material conditional is an unpromising vehicle for expressing default rules. However, the generalisation to Lindström quantifiers eliminates reliance on the material conditional [22]. Nevertheless, the more general quantifiers still rely on cardinalities of subsets, and this introduces something not present in the original reasoning of the agent. Drivers do not learn a default rule by keeping actuarial tables of frequencies and consulting these (e.g. adding up the number of exceptions) when decisions need to be made about crossing an intersection. The same consideration rules out explications based on non-subjective probabilities, and even any subjective probabilities that require conscious arithmetical calculations, as conditional probabilities would.

Perhaps a default rule such as "If I turn the key, then normally the car will start" might be included as a premiss in a (flat) deduction by rewriting it in the form "If I turn the key and the battery is not flat and ... and there is petrol in the tank, then the car will start". While the exhaustive spelling out of exceptions may be feasible in some cases, it is by no means clear that it is even theoretically possible in the general case, for there is no way to be sure that all exceptions have been mentioned. (This is the familiar *qualification problem* in artificial intelligence, closely related to the *frame problem* and the *closed world assumption.*) And even if it were possible to spell out all the exceptions, a person (or artificial agent) does not first verify that the exceptions fail to apply before drawing the inference, so that this trick can only succeed at the cost of failing to formalise the reasoning actually used in such cases.

Perhaps the agent simply assumes that the car will start, or that the driver will stop at a red light, as an hypothesis? After all, Popper wrote: "The actual procedure of science is to operate with conjectures: to jump to conclusions" [70, p.53]. But such an hypothesis is not a random shot in the dark. Rational decision-making takes account of risks and is not only far from random but also guards against believing whatever the agent may want to believe (i.e. is not wishful thinking). Somehow the adoption of the hypothesis is rendered reasonable by the past experience of the agent, by what has been learnt. And what has been learnt is not that this car will start on this occasion or that this driver will stop at this red light, for we are now talking about the grounds for assuming these outcomes. What the agent has learnt is something along the lines of "This is the way things usually work, though not always". Might the Popperian rejoinder not be that logic has no business concerning itself with whatever mysterious process underlies the formation of hypotheses? But why should such defeatism be taken seriously? If it is possible to give a semantic account of this process, there is surely nothing wrong with doing so. And nonmonotonic logic does give such a semantic account, by representing what the agent has learnt as a relation on states. It is entirely compatible with the epistemological perspective to include this agent-orientation in the logical modelling process.

Our justification of the explication of inductive reasoning by $\mid\!\sim$ has brought us face to face with the question of hypothesis formation, to which we now give full attention.

§7. Abductive reasoning. Popperian (and popular) conceptions of inductive reasoning conflate singular predictive inference and hypothesis formation, whereas Peirce was at pains to distinguish them. Hypothesis formation itself has two aspects: it may involve a search for an object language sentence X (singular or general) that explains the observed facts (as expressed by another object language sentence Y), or it may involve the abstraction of a law or rule by which to guide singular predictive inferences. Peirce did not distinguish between these aspects of hypothesis formation, but we are in a position to do so. The former process of forming an explanatory hypothesis at the level of the object language we shall continue to call *abduction*, and our purpose shall be to relate this process to the entailment relation $\mid\!\sim^*$. But the second process, that of abstracting a rule by which to guide predictions, appears to us to be substantially different. We have argued above that rule abstraction does not necessarily result in something that can be expressed by a sentence of the object language. It follows that the process of rule abstraction is not amenable to characterisation by a relation on object-level sentences. Accordingly we take rule abstraction to be an important type of *learning* rather than a type of reasoning, for reasoning starts with a sentence and ends with a sentence. As psychological research on child development has shown, categories, the most widely-useful kind of rule for guiding predictions, are gradually learnt from instances [9, 30, 7].

Since the difference between reasoning and rule abstraction is important, we note that in computer science, and in particular in artificial intelligence, programs such as artificial neural networks have been created that learn from examples to classify new instances or make predictions with a fair degree of accuracy [77, 55]. There are two levels of granularity at which the operation of such a network may be analysed. The programmer is concerned with the finer granularity at which a probabilistic or other learning algorithm is implemented, while the ultimate user may be interested solely in the predictions made by the network. To appreciate the difference, consider the example of a network which, upon being told that some room contains, say, a ceiling and an oven, then makes predictions about what else is likely to be in the room, and infers that the room contains a coffee cup but no fireplace, and a coffee pot but no computer [78, 83]. At the fine-grained level, the system doesn't know about bathrooms or offices, as its nodes correspond to low-level descriptors such as 'has-windows', but its search leads it to local maxima that we recognise as prototypical kitchens or bathrooms, and so its inferences have all the hallmarks of categorical induction. The default rule is learnt by a subsymbolic process at the fine-grained level; the predictions manifest themselves at the coarse-grained level where logical judgments can be made, for instance that the predictions are defeasible (since errors occur from time to time). Since our concern is with reasoning and entailment, rather than the subsymbolic processes involved in learning, we shall say no more about rule abstraction.

7.1. Views on abduction. What, then, is abduction? As a starting point, we turn to Peirce, who writes [65, 5:189]:

> Long before I first classed abduction as an inference it was recognized by logicians that the operation of adopting an explanatory hypothesis – which is just what abduction is – was subject to certain conditions. Namely, the hypothesis cannot be admitted, even as a hypothesis,

> unless it be supposed that it would account for the facts or some of them. The form of inference, therefore, is this:
>
> The surprising fact, C, is observed;
> But if A were true, C would be a matter of course,
> Hence, there is reason to suspect that A is true.

A voluminous literature on abduction has grown from Peirce's writings. Among the questions that deserve attention are: What constitutes an explanation? Should we concern ourselves only with finding the best explanation? Can abduction be characterised logically? What is the epistemic status of an explanation?

What makes X an explanation of Y? To judge by the quoted extract, Peirce takes the test for whether X constitutes an explanation of Y to be $X \models Y$. This very natural idea has been termed 'classical abduction' [91] and its supporters appear to include Helft and Flach. Helft (who did not distinguish between abduction and induction) wrote [40]:

> ... a system is presented with information concerning a domain; its task is to infer hypotheses that allow it to 'explain' what it observes. From a logical standpoint, what we informally call here 'explain' is in fact 'deduce'.

Flach [23] arrived at a similar point via a different route; he defines a 'strong explanatory consequence relation' $Y \mid\!< X$ relating an explanandum Y to an explanans X in terms of a set of postulates, and then shows that $Y \mid\!< X$ iff $X \models Y$ for some choice of the set of states.

There have however been dissenting voices, who have argued that "abduction is not deduction-in-reverse" [15, 90]. A natural alternative is to take X to be an explanation of Y if $X \mid\!\sim Y$. Suppose the fact we want to explain is $Y =$ "Tweety flies". An obvious sort of explanation is $X =$ "Tweety is a bird". It is not the case that $X \models Y$, since penguins cannot fly, but if we have an appropriate representation of the typicality structure of the category of birds, we can exploit the fact that the most typical birds fly to get that $X \mid\!\sim Y$. This is the notion of epistemic explanation [12], according to which X is an epistemic explanation of Y if knowing X would be a reason to believe Y. Similarly, Walton considers explanation to be the transmission of understanding, and seems implicitly to equate understanding something with being able to defeasibly infer it [92]. However, as pointed out by a reviewer [41]: " ... ability to infer the occurrence of a phenomenon is neither necessary nor sufficient for understanding it; for example, one can understand why an atom of a radioactive isotope decays without being able to infer it from the data, and one can infer from hearing thunder that lightning just struck without understanding why lightning struck".

7.2. Abduction and explanation. We argue that neither $\models$ nor $\mid\!\sim$ constitutes a necessary criterion of explanation, at least not of the notion that X is a *potential partial explanation* of Y. We introduce this notion via a murder mystery, and show that its constraining criterion is given by $\mid\!\sim^*$.

The lifeless body of the Squire is discovered in his library. The famous detective Sherlock Holmes is brought in to solve the case. Ultimately, Sherlock hopes to discover a complete explanation of events. But he knows that this complete explanation is a puzzle to be built piece by piece, and he begins by focusing his attention on such matters as fingerprints, footprints in the flowerbed, and (in the modern day) DNA evidence. What is the significance that ought to be attached to the discovery, on an empty brandy glass in the library, of fingerprints belonging to the Squire's neighbour, Colonel Pepper? If Colonel Pepper were a frequent visitor, often to be found chatting with the Squire over a convivial spot of brandy, then the fingerprints would have very little significance. But suppose Sherlock discovered the Colonel and the Squire to have been estranged, so that a sociable visit would be highly unusual. Now the fingerprints become very significant indeed. What model-theoretic criterion reflects this intuitive notion of significance?

Suppose $Y =$ "The Squire is murdered in his own house" and $X =$ "Colonel Pepper visits the house". What gives X significance and makes it a piece of the puzzle is not that $X \models Y$, for this is far too stringent a test of significance – that the Colonel visited the house does not necessarily imply that he murdered the Squire. Nor is the criterion of significance that $X \mid\!\sim Y$, for one could hardly claim that if the Colonel visited the Squire's house then one would normally expect him to murder the Squire. But what one can say is that, in view of their estrangement, a visit by the Colonel would not be a normal event for the Squire. That is, the models of X have empty intersection with the most normal models of "The Squire is not murdered", i.e. $\neg Y$. But the requirement that $Mod(X) \cap P(\neg Y) = \varnothing$ is precisely equivalent to $X \mid\!\sim^* Y$.

For a fully formalised example contrasting $\models$, $\mid\!\sim$, and $\mid\!\sim^*$, we return to the Light-Fan System (metaphorically construed as a nuclear powerplant). Recall from Section 2 that the object language is generated by the atoms p and q, with $p =$ "The light is on" and $q =$ "The fan is on". The truth assignments, written as binary strings giving the truth values assigned to p first and then q, are ordered as follows:

11	
01	↑ more preferred
00	
10	

Suppose the fact to be explained is q. Perhaps the control room agent has observed, on coming to work that morning, that the cooling system (i.e. 'fan') is operating at full blast. What candidates are permitted by the three possible criteria for explanation? If we require strictly that $X \models q$, then $X = p \wedge q$ and $X = \neg p \wedge q$ are the only explanations (apart from the contradictory explanation $p \wedge \neg p$ and the trivial explanation q). These would make sense as ultimate solutions or completed puzzles, but are less reasonable as preliminary pieces of the puzzle. If we desire candidate explanations that are small pieces of the puzzle, then we could relax the requirement to $X \mid\!\sim q$, in which case the possibilities include all of the preceding plus $X = p$ and $X = \neg p$ and $X = \neg(p \leftrightarrow q)$. As simple pieces of the puzzle, $X = p$ and $X = \neg p$ stand out. But now the main defect of $\mid\!\sim$ as criterion for explanation becomes evident. It is possible to have $X \mid\!\sim Y$ in cases where all the (perhaps very many) models of X except for the (perhaps very few) maximally preferred ones are in fact typical models of

$\neg Y$. In our example, $X = \neg p$ has one maximally preferred model, namely 01, which is a model of $Y = q$, while all the other models of $X = \neg p$ (by chance, there happens to be only a single other model) are models of $\neg Y = \neg q$, and in fact happen to be the most preferred model(s) of $\neg Y = \neg q$. Can one really consider as a potential explanation of $Y = q$ a feature which is associated with typical cases of $\neg Y = \neg q$? Unfortunately, the criterion $\mid\!\sim$ cannot discriminate between sensible candidate p and suspect candidate $\neg p$. However, if we require that $X \mid\!\sim^* q$ then $X = p$ remains as a candidate but $X = \neg p$ is eliminated, since $Mod(\neg p) \nsubseteq \mathbf{W} - P(\neg q)$.

In fairness to Peirce, it should be noted that he did not identify abduction with deduction-in-reverse, i.e. claim that X explains Y if and only if $X \models Y$. While he seems content with $X \models Y$ as a sufficient condition for X to be an explanation of Y, it is not a necessary condition, for he stipulated that a hypothesis X cannot be admitted unless it would account for "the facts *or some of them*" (emphasis added). The necessary condition is thus that from X we should be able to deduce some weakening of Y, in other words that $X \models Q(Y)$, where $Q(Y)$ is logically weaker than Y. But recall Section 5, where we showed that, for a sensible choice of $Q(Y)$, the condition $X \models Q(Y)$ is equivalent to $X \mid\!\sim^* Y$.

Is there any other evidence that Peirce may have been willing to accept, as a necessary condition for explanation, some criterion resembling $X \mid\!\sim^* Y$? To some extent there is. He distinguishes between theories (explanations) on grounds of plausibility [66, p.167]:

> By Plausible, I mean that a theory that has not yet been subjected to any test, although more or less surprising phenomena have occurred which it would explain if it were true, is in itself of such a character as to recommend it for further examination or, if it be *highly* plausible, justify us in seriously inclining toward belief in it, as long as the phenomena be inexplicable otherwise.

Although Peirce never defines plausibility more formally, he gives examples to illustrate its absence. Consider the first of these [66, p.166]:

> Suppose a particularly symmetrical larch tree near the house of a great lover of such trees had been struck by lightning and badly broken, and that as he was looking sorrowfully out of the window at it, he should have happened to say, 'I wonder why that particular tree should have been struck, when there are so many about the place that seem more exposed!' Suppose, then, that his wife should reply, 'Perhaps there may be an eagle's eyrie on some of the hills in the neighborhood, and perhaps a male bird in building it may have used some stick that had a nail in it; and one of the eaglets may have scratched itself against the nail; so that the mother may have reproached the male for using such a dangerous stick; and he, being vexed with her teasing, may have determined to carry the piece to a great distance; it may have been while he was doing this that the explosion of lightning took place, and the electricity may have been deflected by the iron in such a way as to strike this tree. Mind, I do not say that this is what did happen; but if you want to find out why that tree was struck, I think you had

> better search for an eyrie, and see whether any of the eaglets have been scratched.'

If Y is "Lightning struck the symmetric larch" and the proposed explanation is $X =$ "An eagle flew by carrying a stick with a nail in it", then what is it that makes X a poor explanation? It is surely that we can conceive of the eagle thus carrying the stick in typical states of the world in which the larch was untouched by lightning. The implausible theory is shown to be suspect when tested by $\mathrel{|\!\!\sim}^*$, since $Mod(X) \cap P(\neg Y) \neq \varnothing$. We cannot recruit Peirce as a supporter of $\mathrel{|\!\!\sim}^*$, of course, but we may at least aver that as far as may be judged from Peirce's own examples, $\mathrel{|\!\!\sim}^*$ is compatible with his intuition about the character that would recommend an explanation for further examination.

To summarise, we suggest that abduction should be thought of according to the detective scenario, i.e. as a process of solving a puzzle one piece at a time rather than as a single leap to an unique best explanation. The pieces of the puzzle are potential partial explanations, and to be worthy of attention such a potential partial explanation X need meet only a simple initial test, namely that $X \mathrel{|\!\!\sim}^* Y$, where Y is the thing to be explained. Here $\mathrel{|\!\!\sim}^*$ is to be seen as the initial criterion for arriving at a set of candidates, not as a means of singling out a 'best' explanation. As others have put it in a related context: " ... we feel it is better to study inference procedures which represent the set of remaining theories ... rather than inference procedures which are constrained to return a single answer" [74]. It is to be expected that the abductive process will continue by culling candidates that for various reasons are unsuitable. We shall not discuss this culling further, as it is unclear whether the criteria for culling would be logical or pragmatic, and in any case there is an enormous literature devoted to abduction as inference to the best explanation in which such criteria are proposed [37, 89, 46].

7.3. The direction of abduction. If one element of our proposal is that emphasis should initially be placed on the set of potential partial explanations rather than on the best (complete) explanation, then the other element is that our criterion for a potential partial explanation is a relation that runs from explanans to explanandum rather than vice versa. That is, although abductive reasoning may be conceived as a process in which one starts with the fact Y to be explained and eventually arrives at a potential partial explanation X, our criterion is a relation such that $X \mathrel{|\!\!\sim}^* Y$, read as "$X$ explains Y". This is in contrast to, for example, Flach's notion of an explanatory consequence relation $Y \mathrel{|\!<} X$ which is to be read as "Y is explained by X". In a mathematical sense, it doesn't matter whether one works with a relation or, reversing all the ordered pairs of the relation, with its converse. Is $X \mathrel{|\!\!\sim}^* Y$ just the converse of $Y \mathrel{|\!<} X$? No, for Flach's $\mathrel{|\!<}$ satisfies a collection of properties not all of which are satisfied by the converse of $\mathrel{|\!\!\sim}^*$.

Consider the postulate of Predictive Convergence [23, p.96]: if $Y \wedge X \models Z$ and $Y \mathrel{|\!<} X$, then $Z \mathrel{|\!<} X$. Rewritten in the form that would be appropriate if the converse of $\mathrel{|\!<}$ were $\mathrel{|\!\!\sim}^*$, the postulate becomes: if $Y \wedge X \models Z$ and $X \mathrel{|\!\!\sim}^* Y$, then $X \mathrel{|\!\!\sim}^* Z$. But consider the Light-Fan example, with the four states ordered as previously. Let $X = \neg(p \leftrightarrow q)$, $Y = q$, and $Z = (\neg p \vee q)$. Then certainly

$Y \wedge X \models Z$, for $Mod(Y \wedge X) = \{01\} \subseteq Mod(Z)$. Also $X \mid\!\sim^* Y$, for the most preferred model of $\neg Y$ is 00, which is not in $Mod(X) = \{10, 01\}$. But it is not the case that $X \mid\!\sim^* Z$, since $\neg Z$ is equivalent to $p \wedge \neg q$, and the most preferred model of $\neg Z$ is 10, which is in $Mod(X)$.

The counter-example shows directly that $\mid\!\sim^*$ is not the converse of Flach's $\mid\!<$, but we could also infer this indirectly, since he shows that $\mid\!<$ is (more or less) deduction-in-reverse. The question remains whether the converse of $\mid\!\sim^*$ ought to be used, so as to mimic the direction of Flach's $\mid\!<$. We think not, in view of the elegant duality between $\mid\!\sim^*$ and $\mid\!\sim$. Moreover, the tendency, for those accustomed to Western left-to-right script, to focus attention on the lefthand side of expressions, has the effect of magnifying the salience of properties such as the monotonicity of $\mid\!\sim^*$ (left strengthening). This is convenient, because if we think of abduction as a process of puzzle-building, then it is very important that pieces of the puzzle should be able to be combined to give a larger portion of the puzzle. Flach does include the counterpart of monotonicity for $\mid\!\sim^*$, which becomes the postulate of (admissible) right strengthening, but the importance of this property for abductive reasoning is underplayed. In contrast, those who adopt the explanandum → explanans direction tend to emphasise what they call the nonmonotonicity of abduction (see e.g. [52]), by which they mean that an explanation X of Y need not be an explanation of $Y \wedge Z$. From the puzzle-building perspective, this is not at all important, for the puzzle begins with Y, and the challenge is to find the set of potential partial explanations from which, eventually, a complete solution can be selected.

7.4. Questions on abduction. Can the process of abduction as a whole be characterised logically? And what is the epistemic status of an explanation? For a general discussion of the last two questions, we refer the reader to recent work by Gabbay and Woods on formal pragmatics, agendas, and abduction [27]. Gabbay follows Aliseda [2, 3] in regarding the broad abductive process as being constrained by a schema in which logical relationships play a narrow role (narrower for Gabbay than for Aliseda). Without denying the importance of pragmatics for the broad abductive process, we shall confine ourselves to the logical relationship that holds between explanans and explanandum. As far as the epistemic status of explanations is concerned, this appears impossible to resolve without careful consideration of the actions that might be based on such explanations, but we grant Gabbay's claim to the effect that an explanation ought not to be adopted ipso facto as a belief and thus as a basis for planning to achieve goals unrelated to testing of the explanatory hypothesis.

The epistemic status of abductively derived explanations (prior to a deliberate culling process) remains unclear, and we are mindful that having reason to *suspect* that an explanation is true (as Peirce put it) is not the same as having reason to *believe* the explanation. But we conjecture that the epistemic architecture of agents ought to allow for the representation of disbeliefs as well as beliefs, with disbeliefs acting as hedges against the formation of some new beliefs and, in some cases, contracting the belief set. It may be that, in terms of such an architecture, an abductive explanation X could be represented, not as a belief, but as a tentative *rejection* of, or *disbelief* in, $\neg X$ [33, 34, 14], which is closely related to *contracting* a theory by removing or withdrawing $\neg X$ [26, 56]. In the contexts of rejection, disbelief, and contraction it is acknowledged that the negative stance towards the sentence, in this case $\neg X$, should extend to all logically stronger sentences that would classically, or even defeasibly, entail $\neg X$. Here the fact that the set $\{X \mid X \mathrel{|\!\!\sim}^* Y\}$ forms an ideal in the Lindenbaum-Tarski algebra may prove useful. We shall not explore this idea further, leaving it for future research.

§8. Conclusions. We introduced the notion of duality into preferential semantics, showing that to every defeasible entailment relation $\mathrel{|\!\!\sim}$ there may be associated a dual relation $\mathrel{|\!\!\sim}^*$ which is not the converse of $\mathrel{|\!\!\sim}$, but rather its contrapositive, and which is not characterised by the same properties as $\mathrel{|\!\!\sim}$. Whereas preferential semantics and $\mathrel{|\!\!\sim}$ have together created the impression that the key feature of ampliative reasoning, and therefore of defeasible entailment, is nonmonotonicity, it turns out that $\mathrel{|\!\!\sim}^*$, while supraclassical, is always monotonic.

Following a comparison of $\mathrel{|\!\!\sim}$ and $\mathrel{|\!\!\sim}^*$, we argued in favour of Peirce's tripartite classification of reasoning as deductive, inductive, or abductive, and that the three forms of reasoning can be semantically characterised by the constraints embodied in the entailment relations $\models$, $\mathrel{|\!\!\sim}$, and $\mathrel{|\!\!\sim}^*$ respectively. Popperian skepticism of induction was countered by Musgrave's distinction between metaphysical and epistemological criteria for entailment as well as arguments to the effect that $\mathrel{|\!\!\sim}$ cannot be reduced to (flat) deduction. We demonstrated that the semantic constructions behind $\mathrel{|\!\!\sim}$ and $\mathrel{|\!\!\sim}^*$ can be syntactically simulated in appropriate modal languages, so that it is indeed the case that $\mathrel{|\!\!\sim}$ and $\mathrel{|\!\!\sim}^*$ could be reduced to $\models$ in the context of a tense logic whose bidirectional frames incorporate the preference ordering relation on states (and its converse).

For the explication of abductive reasoning, we distinguished between rule abstraction and explanatory reasoning. For the latter we proposed a puzzle-building metaphor, and used $X \mathrel{|\!\!\sim}^* Y$ as the criterion for X to be a potential partial explanation of Y. Certain of Peirce's discussions on the nature of hypotheses suggest that our criterion is compatible with (our reconstruction of) his views. While conceding that the logical characterisation we offer pertains only to a part of the overal abductive process, we find aesthetic gratification in the fact that the duality between $\mathrel{|\!\!\sim}$ and $\mathrel{|\!\!\sim}^*$ mirrors the intuitive impression that inductive and abductive reasoning proceed in opposite directions.

REFERENCES

[1] C.E. Alchourrón, P. Gärdenfors, and D. Makinson, *On the logic of theory change: Partial meet functions for contraction and revision*, ***The Journal of Symbolic Logic***, vol. 50 (1985), pp. 510–530.

[2] A. Aliseda, *A unified framework for abductive and inductive reasoning in philosophy and AI*, ***ECAI'96 Workshop on Abductive and Inductive Reasoning*** (M. Denecker, L. De Raedt, P. Flach, and T. Kakas, editors), 1996.

[3] ———, ***Abductive Reasoning: Logical Investigations into Discovery and Explanation***, Springer, Dordrecht, 2006.

[4] G. Amati, L. Carlucci, D. Gabbay, and F. Pirri, *A structural property on modal frames characterizing default logic*, ***Logic Journal of the IGPL***, vol. 4 (1996), no. 1, pp. 7–22.

[5] O. Arieli and A. Avron, *General patterns for nonmonotonic reasoning: From basic entailments to plausible relations*, ***Logic Journal of the IGPL***, vol. 8 (2000), pp. 119–148.

[6] A. Avron, *Simple consequence relations*, ***Information and Computation***, vol. 92 (1991), pp. 105–139.

[7] R. Baillargeon, *Infants' understanding of the physical world*, ***Advances in Psychological Science*** (M. Sabourin, F. Craik, and M. Roberts, editors), vol. 2: *Biological and Cognitive Aspects*, Psychology Press, London, 1998, pp. 503–529.

[8] P. Blackburn, M. de Rijke, and Y. Venema, ***Modal Logic***, Cambridge Tracts in Theoretical Computer Science, vol. 53, Cambridge University Press, Cambridge, UK, 2002.

[9] P.C. Bomba and E.R. Siqueland, *The nature and structure of infant form categories*, ***Journal of Experimental Child Psychology***, vol. 35 (1983), pp. 294–328.

[10] G. Boolos, ***The Logic of Provability***, Cambridge University Press, Cambridge, UK, 1993.

[11] C. Boutilier, *Conditional logics of normality: A modal approach*, ***Artificial Intelligence***, vol. 68 (1994), no. 1, pp. 87–154.

[12] C. Boutilier and V. Becher, *Abduction as belief revision*, ***Artificial Intelligence***, vol. 77 (1995), pp. 43–94.

[13] P.D. Bruza, D. Widdows, and J. Woods, *A quantum logic of down below*, ***Handbook of Quantum Logic, Quantum Structure, and Quantum Computation*** (K. Engesser, D. Gabbay, and D. Lehmann, editors), Elsevier, Forthcoming. URL: arXiv:quant-ph/0612051.

[14] S. Chopra, J. Heidema, and T. Meyer, *Some logics of belief and disbelief*, ***AI 2003: Advances in Artificial Intelligence*** (T.D. Gedeon and L.C.C. Fung, editors), Lecture Notes in Artificial Intelligence, vol. 2903, Springer-Verlag, Berlin, 2003, pp. 364–376.

[15] M. Cialdea Mayer and F. Pirri, *Abduction is not deduction-in-reverse*, ***Logic Journal of the IGPL***, vol. 4 (1996), no. 1, pp. 95–108.

[16] W.J. Clancey, ***Situated Cognition: On Human Knowledge and Computer Representations***, Cambridge University Press, Cambridge, 1997.

[17] G. Crocco and P. Lamarre, *On the connections between nonmonotonic inference systems and conditional logics*, ***Principles of Knowledge Representation and Reasoning (KR'92)*** (R. Nebel, C. Rich, and W. Swartout, editors), Morgan Kaufmann Publishers, San Mateo, CA, 1992, pp. 565–571.

[18] A.R. Damasio, ***Descartes' Error: Emotion, Reason, and the Human Brain***, HarperCollins, New York, 1994.

[19] E. Davis, ***Representations of Commonsense Knowledge***, Morgan Kaufmann, San Mateo, CA, 1990.

[20] J. Delgrande, T. Schaub, H. Tourpits, and K. Wang, *A classification and survey of preference handling approaches in non-monotonic reasoning*, ***Computational Intelligence***, vol. 20 (2004), no. 2, pp. 308–334.

[21] M.A. Dickman, ***Large Infinitary Languages***, North Holland, Amsterdam, 1975.

[22] H.-D. Ebbinghaus, *Extended logics: The general framework*, ***Model-Theoretic Logics*** (J. Barwise and S. Feferman, editors), Springer-Verlag, New York, 1985, pp. 25–76.

[23] P.A. Flach, *On the logic of hypothesis generation*, ***Abduction and Induction: Essays on their Relation and Integration*** (P.A. Flach and A.C. Kakas, editors), Kluwer, Dordrecht, 2000, pp. 89–106.

[24] M. Freund and D. Lehmann, *Belief revision and rational inference*, ***Technical Report TR–94–16***, Leibniz Center for Research in Computer Science, Hebrew University, Jerusalem, 1994, URL: arXiv:cs/0204032v1 [cs.AI].

[25] M. Freund, D. Lehmann, and P. Morris, *Rationality, transitivity and contraposition*, ***Artificial Intelligence***, vol. 52 (1991), no. 2, pp. 191–203.

[26] A. Fuhrmann, *Theory contraction through base contraction*, ***Journal of Philosophical Logic***, vol. 20 (2000), pp. 175–203.

[27] D. Gabbay and J. Woods, *Advice on abductive logic*, ***Logic Journal of the IGPL***, vol. 14 (2006), pp. 189–219.

[28] P. Gärdenfors (editor), ***Generalized Quantifiers: Linguistic and Logical Approaches***, D. Reidel, Dordrecht, 1987.

[29] ———, ***Conceptual Spaces: The Geometry of Thought***, MIT Press, Cambridge, MA, 2000.

[30] S.A. Gelman and E.M. Markman, *Categories and induction in young children*, ***Cognition***, vol. 23 (1986), pp. 183–209.

[31] M.L. Ginsberg (editor), ***Readings in Nonmonotonic Reasoning***, Morgan Kaufmann, San Mateo, CA, 1987.

[32] L. Giordano, V. Gliozzi, N. Olivetti, and G.L. Pozzato, *Analytic tableaux for KLM preferential and cumulative logics*, ***LPAR 2005*** (G. Sutcliffe and A. Voronkov, editors), Lecture Notes in Artificial Intelligence, vol. 3835, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 666–681.

[33] A. Gomolińska, *On the logic of acceptance and rejection*, ***Studia Logica***, vol. 60 (1998), pp. 233–251.

[34] A. Gomolińska and D. Pearce, *Disbelief change*, ***Electronic essays on the occasion of the fiftieth birthday of Peter Gärdenfors***, 2001.

[35] I.J. Good, *46656 Varieties of Bayesians*, ***American Statistician***, vol. 25 (1971), pp. 62–63.

[36] S.O. Hansson and T. Grüne-Yanoff, *Preferences*, ***The Stanford Encyclopedia of Philosophy*** (E.N. Zalta, editor), Center for Logic and Information, Stanford University, winter 2006 ed., 2006, URL: plato.stanford.edu/archives/win2006/entries/preferences.

[37] G.H. Harman, *The inference to the best explanation*, ***Philosophical Review***, vol. 71 (1978), pp. 88–95.

[38] S. Harnad (editor), ***Categorical Perception: The Groundwork of Cognition***, Cambridge University Press, Cambridge, 1987.

[39] J. Hawthorne, *Inductive logic*, ***The Stanford Encyclopedia of Philosophy*** (E.N. Zalta, editor), Center for Logic and Information, Stanford University, summer 2007 ed., 2007, Forthcoming URL: plato.stanford.edu/archives/sum2007/entries/logic-inductive.

[40] N. Helft, *Induction as nonmonotonic inference*, ***Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*** (R.J. Brachman, H.J. Levesque, and R. Reiter, editors), Morgan Kaufmann, San Francisco, CA, 1989, pp. 149–156.

[41] D. Hitchcock, *Review of Walton's Abductive Reasoning*, ***University of Toronto Quarterly***, vol. 75 (2006), no. 1, pp. 155–156.

[42] C. Howson, *Logic with numbers*, ***Synthese***, vol. 156 (2007), no. 3, pp. 491–512.

[43] D. Hume, ***An Enquiry Concerning Human Understanding***, Harvard Classics, vol. 37, Collier & Son, 1910, URL: 18th.eserver.org/hume-enquiry.html.

[44] P.N. Johnson-Laird, ***Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness***, Harvard University Press, Cambridge, MA, 1983.

[45] P.N. Johnson-Laird and R.M.J. Byrne, ***Deduction***, Lawrence Erlbaum, Hove, UK, 1991.

[46] J.R. Josephson and S.G. Josephson (editors), ***Abductive Inference: Computation, Philosophy, Technology***, Cambridge University Press, New York, 1994.

[47] C. Karp, ***Languages with Expressions of Infinite Length***, North Holland, Amsterdam, 1964.

[48] K.T. Kelly and C. Glymour, *Why probability does not capture the logic of scientific justification*, ***Contemporary Debates in Philosophy of Science*** (C. Hitchcock, editor), Blackwell, Malden, MA, 2004, pp. 94–114.

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

[50] I. Lakatos, *Changes in the problem of inductive logic*, ***Mathematics, Science and Epistemology*** (J. Worrall and G. Currie, editors), Philosophical Papers Volume 2, Cambridge University Press, Cambridge, UK, 1978, pp. 128–200.

[51] D. Lehmann and M. Magidor, *What does a conditional knowledge base entail?*, ***Artificial Intelligence***, vol. 55 (1992), pp. 1–60.

[52] L. Magnani, *Model-based and manipulative abduction in science*, ***Foundations of Science***, vol. 9 (2004), pp. 219–247.

[53] P. Maher, *Probability captures the logic of scientific confirmation*, ***Contemporary Debates in Philosophy of Science*** (C. Hitchcock, editor), Blackwell, Malden, MA, 2004, pp. 69–93.

[54] J. McCarthy, *Epistemological problems of artificial intelligence*, ***Proceedings of the Fifth International Joint Conference on Artificial Intelligence*** (R. Reddy, editor), vol. 2, MIT Press, Cambridge, MA, 1977, URL: ijcai.org, pp. 1038–1044.

[55] J.L. McClelland, D.E. Rumelhart, and the PDP Research Group (editors), ***Parallel Distributed Processing: Explorations in the Microstructure of Cognition***, vol. 2: *Psychological and Biological Models*, MIT Press, Cambridge, MA, 1986.

[56] T.A. Meyer, J. Heidema, W.A. Labuschagne, and L. Leenen, *Systematic withdrawal*, ***Journal of Philosophical Logic***, vol. 31 (2002), no. 5, pp. 415–443.

[57] T.A. Meyer, W.A. Labuschagne, and J. Heidema, *Refined epistemic entrenchment*, ***Journal of Logic, Language, and Information***, vol. 9 (2000), pp. 237–259.

[58] A.C. Michalos, ***The Popper-Carnap Controversy***, Martinus Nijhoff, The Hague, 1971.

[59] Y. Moinard and R. Rolland, *Characterizations of preferential entailments*, ***Logic Journal of the IGPL***, vol. 10 (2002), no. 3, pp. 245–272.

[60] A. Mostowski, *On a generalization of quantifiers*, ***Fundamenta Mathematicae***, vol. 44 (1957), pp. 12–36.

[61] G.L. Murphy, ***The Big Book of Concepts***, MIT Press, Cambridge, MA, 2002.

[62] G.L. Murphy and B.H. Ross, *The two faces of typicality in category-based induction*, ***Cognition***, vol. 95 (2005), pp. 175–200.

[63] A. Musgrave, *How Popper [might have] solved the problem of induction*, ***Philosophy***, vol. 79 (2004), pp. 19–31.

[64] D.N. Osherson, E.E. Smith, O. Wilkie, A. López, and E. Shafir, *Category-based induction*, ***Psychological Review***, vol. 97 (1990), pp. 185–200.

[65] C.S. Peirce, ***Collected Papers of Charles Sanders Peirce*** (C. Hartshorne and P. Weiss, editors), vol. I-VI, Harvard University Press, Cambridge, MA, 1931-1935, Quotations according to volume and paragraph.

[66] ———, ***Philosophical Writings of Peirce*** (J. Buchler, editor), Dover, New York, 1955.

[67] ———, ***Collected Papers of Charles Sanders Peirce*** (A.W. Burks, editor), vol. VII-VIII, Harvard University Press, Cambridge, MA, 1958, Quotations according to volume and paragraph.

[68] J.L. Pollock and J. Cruz, ***Contemporary Theories of Knowledge***, 2nd ed., Rowman & Littlefield, Lanham, 1999.

[69] K. Popper, ***The Logic of Scientific Discovery***, Hutchinson, London, 1959.

[70] ———, ***Conjectures and Refutations***, Routledge and Kegan Paul, London, 1963.

[71] ———, *The problem of induction*, ***A Pocket Popper*** (D. Miller, editor), Fontana, Oxford, 1983, pp. 101–117.

[72] K. Popper and D. Miller, *A proof of the impossibility of inductive probability*, ***Nature***, vol. 302 (1983), pp. 687–688.

[73] L.J. Rips, *Inductive judgments about natural categories*, ***Journal of Verbal Learning and Verbal Behavior***, vol. 14 (1975), pp. 665–681.

[74] R.L. Rivest and R. Sloane, *A new model for inductive inference*, ***Theoretical Aspects of Rationality and Knowledge: Proceedings of the 2nd Conference*** (M.Y. Vardi, editor), Morgan Kaufmann, San Mateo, CA, 1988, URL: www.tark.org, pp. 13–27.

[75] E. Rosch, *Cognitive representations of semantic categories*, ***Journal of Experimental Psychology: General***, vol. 104 (1975), no. 3, pp. 192–233.

[76] ———, *Principles of categorization*, ***Cognition and Categorization*** (E. Rosch and B.B. Lloyd, editors), Erlbaum, Hillsdale, NJ, 1978, pp. 27–48.

[77] D.E. Rumelhart, J.L. McClelland, and the PDP Research Group (editors), ***Parallel Distributed Processing: Explorations in the Microstructure of Cognition***, vol. 1: *Foundations*, MIT Press, Cambridge, MA, 1986.

[78] D.E. Rumelhart, P. Smolensky, J.L. McClelland, and G.E. Hinton, *Schemata and sequential thought processes in parallel distributed processing models*, ***Parallel Distributed***

Processing: Explorations in the Microstructure of Cognition (J.L. McClelland, D.E. Rumelhart, and the PDP Research Group, editors), vol. 2: *Psychological and Biological Models*, MIT Press, Cambridge, MA, 1986.

[79] B. Russell, ***Introduction to Mathematical Philosophy***, George Allen and Unwin, London, 1919.

[80] K. Segerberg, ***An Essay in Classical Modal Logic***, Filosofiska Studier, vol. 13, Uppsala Universitet, Uppsala, 1971.

[81] Y. Shoham, ***Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence***, The MIT Press, Cambridge, MA, 1988.

[82] S.A. Sloman and D.A. Lagnado, *The problem of induction*, ***The Cambridge Handbook of Thinking and Reasoning*** (K.J. Holyoak and R.G. Morrison, editors), Cambridge University Press, Cambridge, UK, 2005, pp. 95–116.

[83] P. Smolensky, *Connectionist AI, symbolic AI, and the brain*, ***Artificial Intelligence Review***, vol. 1 (1987), no. 2, pp. 95–109.

[84] R.J. Solomonoff, *A formal theory of inductive inference – part i*, ***Information and Control***, vol. 7 (1964), pp. 1–22.

[85] ———, *Complexity-based induction systems*, ***IEEE Transactions on Information Theory***, vol. 24 (1978), no. 5, pp. 422–432.

[86] G. Sundholm, *Systems of deduction*, ***Handbook of Philosophical Logic*** (D. Gabbay and F. Guenthner, editors), vol. 2, Kluwer, Dordrecht, 2nd ed., 2001, pp. 1–52.

[87] ———, *Varieties of consequence*, ***A Companion to Philosophical Logic*** (D. Jacquette, editor), Blackwell, Oxford, 2002, pp. 241–255.

[88] A. Tarski, ***Logic, Semantics, Metamathematics: Papers from 1923 to 1938*** (J. Corcoran, editor), Hackett, Indianapolis, 2nd ed., 1983, Translated by J.H. Woodger.

[89] P. Thagard, *The best explanation: Criteria for theory choice*, ***Journal of Philosophy***, vol. 75 (1978), pp. 76–92.

[90] P. Thagard and C. Shelley, *Abductive reasoning: Logic, visual thinking, and coherence*, ***Logic and Scientific Methods*** (M.L. Dalla Chiara, K. Doets, D. Mundici, and J. van Benthem, editors), Kluwer, Dordrecht, 1997, pp. 413–427.

[91] B. Walliser, D. Zwirn, and H Zwirn, *Abductive logics in a belief revision framework*, ***Journal of Logic, Language, and Information***, vol. 14 (2005), pp. 87–117.

[92] D. Walton, ***Abductive Reasoning***, University of Alabama Press, Tuscaloosa, 2004.

[93] Z. Zhu, *Similarity between preferential models*, ***Theoretical Computer Science***, vol. 353 (2006), pp. 26–52.

SCHOOL OF COMPUTING
UNIVERSITY OF SOUTH AFRICA
PRETORIA, SOUTH AFRICA
E-mail: britzk@unisa.ac.za

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTH AFRICA
PRETORIA, SOUTH AFRICA
E-mail: heidej@unisa.ac.za

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF OTAGO
DUNEDIN, NEW ZEALAND
E-mail: willem@cs.otago.ac.nz