Title:
Relevant Deduction.
From Solving Paradoxes Towards a General Theory

Author:
Gerhard Schurz

IPS-PREPRINTS
Annual 1990 No. 1

Edited by Gerhard Schurz und Alexander Hieke

Vorveröffentlichungsreihe am Institut für Philosophie der Universität Salzburg
Prepublication Series at the Department of Philosophy, University of Salzburg
RELEVANT DEDUCTION
From Solving Paradoxes Towards a General Theory

Gerhard Schurz

Abstract: This paper presents an outline of a new theory of relevant deduction which arose from the purpose of solving paradoxes in various fields of analytic philosophy. In distinction to relevance logics, this approach does not replace classical logic by a new one, but distinguishes between relevance and validity. It is argued that irrelevant arguments are, although formally valid, nonsensical and even harmful in practical applications. The basic idea is this: a valid deduction is relevant iff no subformula of the conclusion is replaceable on some of its occurrences by any other formula salva validitate of the deduction. The paper first motivates the approach by showing that four paradoxes seemingly very distant from each other have a common source. Then the exact definition of relevant deduction is given and its logical properties are investigated. An extension to relevance of premises is discussed. Finally the paper presents an overview of its applications in philosophy of science, ethics, cognitive psychology and artificial intelligence.

1. Four Paradoxes and Their Common Source

1.1. The General Problem of Paradoxes in Analytical Philosophy. Rudolf Carnap and Hans Reichenbach, to whom this volume is dedicated, belong to the most famous founders of modern Analytic Philosophy and Philosophy of Science. It was the program of this philosophical enterprise to base philosophy on systems of exact logical reasoning. This enterprise has developed in (at least) two branches, that of deductive and that of inductive (and/or statistical) reasoning. The following investigations focus on the first branch, deductive reasoning. However, there are good reasons to believe that the basic idea of this paper can also be applied to inductive reasoning. Although there has been considerable progress in realizing the program of Analytic Philosophy since the time of its founders, the enterprise of reconstructing philosophical concepts and principles within a system of exact symbolic logic has always been confronted with the obstinate and sometimes seemingly invincible phenomenon of paradoxes. These paradoxes typically emerge in the

---

1 This paper is based on the first part of my habilitation Schurz (1989). For various helps and comments I am indebted to Paul Weingartner, Andrzej Wróński, Georg Kreisel, David Miller, Kit Fine, Terence Parsons and Peter Woodruff.

2 This is work for the future. That it principally could be done was demonstrated in Schurz (1983b).
following succession of steps. One starts with a concept or principle put forward in natural language which is intuitively well-founded. Next, this concept or principle is formalized within the language of symbolic logic in a seemingly flawless way. From the formalized concept or principle logical consequences are then derived and retranslated into the natural language which are expected to give additional philosophical insights - until one suddenly recognizes that a conclusion is derivable which is intuitively nonsensical in a degree one never would have expected. Although the logical reconstruction has been undertaken in a seemingly perfect way, it has some completely unintuitive implications. This is the typical situation of paradoxes in Analytic Philosophy.

A clarification of the notion of "paradox". This notion is not used here in the 'strict' sense of a (prima facie well-founded) formal principle implying a contradiction, but in the 'loose' sense of a (prima facie well-founded) formal principle implying a consequence which contradicts an intuitive assumption. Of course, by formalizing this intuitive assumption the 'loose' paradox would be turned into a 'strict' paradox. So this distinction is not very deep. I don't think there exists a sharp or deep distinction between 'strict' and 'loose' paradoxes. It rather seems to me a matter of the degree of evidence of those intuitions which turn out to be inconsistent after formalization.

The paradoxes mentioned have been disappointing and sometimes even shocking for analytic philosophers. Some of them even have concluded that the entire program of basing philosophy on exact logical grounds must be given up. In the following we take a look at four typical paradoxes. Thereby we make use of the standard logical terminology: \( \neg \) (negation), \( \lor \) (disjunction), \( \land \) (conjunction), \( \rightarrow \) (material implication), \( \leftrightarrow \) (material equivalence), \( \forall \) (universal quantifier), \( \exists \) (existential quantifier); \( x, y \ldots \) (individual variables); \( a, b \ldots \) (individual constants); \( F, G, R \ldots \) (n-ary predicate letters); \( p, q, \ldots \) (propositional variables); \( \Box \) (alethic necessity), \( \Diamond \) (alethic possibility), \( O \) (deontic obligation), \( P \) (deontic permissibility). \( \mathcal{L} \) denotes the formal language, which is identified with the set of its well formed formulas. If \( \mathcal{L} \) is not explicitly specified, it always means the language of classical first-order logic. Capital Roman letters \( A, B, \ldots \) denote sentences and capital Greek letters denote sets of sentences in \( \mathcal{L} \). \( \mathcal{L} \subseteq \mathcal{L} \) denotes a logic. \( \vdash \in \text{Pow}(\mathcal{L}) \times \mathcal{L} \) denotes a deduction relation ["Pow" for "power set"], i.e. "\( \Gamma \vdash T, A \)" reads "sentence A is derivable from premise set \( \Gamma \)". \( \vdash_{L} \) denotes the deduction relation belonging to logic \( L \) defined by \( \Gamma \vdash_{L} A \) iff \( \Gamma \vdash T, A \) for a finite subset \( \Gamma \subseteq \Gamma \).\(^3\) We assume \( L \) is complete, so the syntactic concept \( \Delta \vdash_{L} A \) is equivalent to the semantical notion of a logically valid inference. If \( L \) is not explicitly specified it always means classical first-order predi-

\(^3\) The characteristic of \( \vdash_{L} \) is that it satisfies the deduction theorem and thus coincides with the semantic notion of validity as truth-conserving relation. \( \vdash_{L} \) may be defined also in different ways.
cate logic, thus $\vdash_T$ always means the deduction relation of classical logic. Further logical notations and concepts are explained later.

1.2 The Ross Paradox of Deontic Logic. Beginning with Ernst Mally, analytic ethicists started to develop the logic of norms. If one formalizes a norm as a sentence of the form $OA$, expressing "it is obligatory that $A$", then the expression "it is forbidden that $A$", $FA$, is definitionally equivalent to $O\neg A$, and the expression "it is permitted that $A$", $PA$, is definitionally equivalent to $\neg O\neg A$. What are the logical rules these normative sentences must obey? It seems obvious that every reasonable deontic logic must have the following rule:

(1) If $A$ is obligatory, and $B$ is a logical consequence of $A$, then $B$ is also obligatory.

For otherwise it would be possible that an ethical system permits something, namely $\neg B$, which logically implies something which the system forbids, namely $\neg A$. For example, it would be possible that a system of law forbids the prevention of free speech, but simultaneously permits pasting up the mouth of another person. This is completely unreasonable, and thus condition (1) is intuitively well-founded. The translation of this rule into formal language leads to the so-called law of monotonicity

(1*) $\vdash_TA \rightarrow B \Rightarrow \vdash_T OA \rightarrow OB$ -- or written as logical rule: $A \rightarrow B/OA \rightarrow OB$

There seems to be no objection to all that. But see what happens. A little logical reflection proves that the rule (1*) yields the following argument as a valid inference of 'every reasonable deontic logic':

(2) You should post the letter.
Therefore: You should post the letter or burn it.

Ross (1941) first recognized that the reasonable rule (1*) leads to this completely counterintuitive argument. He and other deontic logicans were so impressed, if not to say shocked, that they concluded there can't be something like a logic of norms.

1.3 The Hesse Paradox of Confirmation. According to a well-known idea of confirmation put forward by Popper⁴ and others, a scientific theory is confirmed by its true deductive consequences. This idea fits very well with real science, because the latter confirms its theories by successful predictions and explanations of empirical facts or laws which follow deductively from the theory plus initial and

---

⁴ Popper spoke of theory corroboration (cf. 1976, p. 212).
boundary conditions. In this way, Newton's theory was confirmed by its success in predicting the planetary orbits, Einstein's theory of special relativity was confirmed by its successful explanation of the Michelson-Morely experiment, etc.\(^5\)

So the concept of deductive theory confirmation seems perfectly intuitive. Representing a theory \(T\) as a set of sentences (in a given language \(\mathcal{L}\)), an obvious way to define the notion of deductive theory confirmation is the following (suggested e.g. by Hesse 1970, p. 50):

\[
(3) \text{ A sentence } S \text{ confirms a theory } T \text{ iff (i) } S \text{ is contentful (not } \mathsf{T}S), (ii) } T \text{ is consistent (not } T \mathsf{p} \land \neg \mathsf{p}), (iii) S \text{ is true [or 'rationally acceptable'], and (iv) } T \mathsf{TS}.
\]

Conditions (i) and (ii) are prerequisites for a non-trivial notion of theory confirmation. Depending on whether truth or rational acceptability of \(S\) is required in (iii), one gets the semantic versus the pragmatic version of (3).

Furthermore, the following condition of strengthening the confirmans is a reasonable principle of every 'logic of confirmation':

\[
(4) \text{ If } S \text{ confirms } T, \text{ and } S^* \text{ logically implies } S \text{ (} S^* \mathsf{TS} \text{), then } S^* \text{ confirms } T \text{ also.}
\]

For example, if certain astronomical data \(D\) confirm \(T\), then of course every more comprehensive observational report which includes (and hence logically implies) \(D\) confirms \(T\), too.

All this seems to be fine. But look what happens. Some easy steps of propositional logic show that (3) together with (4) imply the following result:

\[
(5) \text{ Every consistent theory } T \text{ is confirmed by every contentful and true [rationally acceptable] sentence } S, \text{ provided only that } \neg S \text{ is consistent with } \neg T.
\]

The proviso is very weak - e.g., it is satisfied if \(T\) and \(S\) have no predicate in common and contain no identity sign.\(^6\) For instance, "Peter is silly" confirms 'the theory of quarks'. This result is completely nonsensical. In (1970) Hesse showed in a similar manner that conditions (3) and (4) yield this paradoxical consequence, whence we

---

\(^{5}\) In view of well-known critiques I want to emphasize that these theoretical predictions are indeed mathematically-deductive. This is not changed by the often mentioned fact that the relation between the 'idealized' facts stated in the premises and the conclusion of the argument and the 'real' observed facts is mostly not that of identity but that of approximation.

\(^{6}\) This follows by virtue of the so-called property of Halldén completeness (Halldén 1951; Fitting 1983, p. 298), which says that \(\mathsf{T}(A \lor B)\) iff \(\mathsf{T}A\) or \(\mathsf{T}B\) provided \(A, B\) have no predicates in common and contain no identity sign. Classical predicate logic and all standard modal logics have this property. It does not hold for formulas with an identity sign; comp. e.g. \(\forall x \forall y(x=y) \mathsf{T} \neg (Fx \land \neg Fy)\).
call it the Hesse paradox. Several philosophers of science, e.g. Glymour 1980, have
drawn the conclusion that deductive logic is unable to reconstruct the process of
confirmation in science.

1.4 The Tichý-Miller Paradox of Verisimilitude. An important idea in philosophy of
science, deriving from Popper, is that of the verisimilitude of theories. Although the
ultimate aim of science is to give true theories, most of them are true only in an
'approximative' sense, but strictly speaking false. For example, Newton's theory,
although successful in explaining the physical laws in the 'middle' dimensions of space
and velocity, turned out to be false for dimensions of very high speed and of very
small space. In the same way, modern elementary particle physics has shown that a
lot of symmetry principles assumed in quantum mechanics can be broken. Since
most scientific theories are strictly speaking false, the idea of "truth" as a criterion
for progress in science does not seem to work. Nevertheless, as Popper argued,
there is progress in science, consisting of gaining theories which are better and
better approximations to the truth, coming closer and closer to the truth. So the
closeness to the truth, or verisimilitude, turns out to be the central concept in
philosophy of science in order to explicate the notion of scientific progress. Popper's
definition of verisimilitude is intuitively very convincing: a theory is the closer to the
truth, the more true and the less false consequences it has. Here, 'more' and 'less'
must not be understood in the sense of 'counting' - it would be unreasonable to
compare Einstein's relativity theory and Freud's psychoanalysis by counting
consequences. Rather, the verisimilitude of theories is only comparable if they speak
about the same domain of objects and properties, and the 'more' and 'less' has to be
understood in the sense of the set-theoretical inclusion relation. This leads to the
following definition of versimilitude, suggested by Popper (1963, p. 233):

(6) Notation: T denotes the set of all true and F the set of all false sentences. T, T'...
denote theories (sets of sentences). \( Cn(T) := \{ A | A \in T \land A \} \) is the set of T's
deductive consequences, \( (T)_t := Cn(T) \cap \mathbb{T} \) is the set of T's true consequences, and
\( (T)_f := Cn(T) \cap \mathbb{F} \) the set of T's false consequences. "\( T_1 >_t T_2 \)" expresses that T_1 is
closer to the truth, i.e. has more verisimilitude, than T_2. \( \subseteq \) stands for proper and \( \subset \) for
improper set inclusion. Then the definition runs:

"\( T_1 >_t T_2 \)" iff: [case 1:] \( (T_2)_t \subseteq (T_1)_t \) and \( (T_1)_f \supseteq (T_2)_f \), or

[case 2:] \( (T_2)_t \subseteq (T_1)_t \) and \( (T_1)_f \subset (T_2)_f \).

Definition (6) seems to be really intelligible. But look what happens. It follows by

7 Hesse has derived this paradox from the stronger assumption that the confirmation relation
is transitive - which is doubtful. Our assumption (4) is much weaker; so the paradox
becomes much more 'paradoxical'.
simple logical means that this definition has the following result:

(7) No false theory can be closer to the truth than any other theory.

The result was proved by Tichý (1974) and Miller (1974) independently of each other, whence we call it the Tichý-Miller paradox. Of course this means a breakdown of Popper's whole idea, since the entire purpose of the notion of versimilitude was to explicate the fact that scientific theories, although being false, can be closer to the truth that other ones. A lot of philosophers have drawn the radical conclusion that Popper's idea of versimilitude is unrealizable and must be explicated in a completely different manner

1.5 The Prior Paradox of Is-Ought Inferences. As is well-known, Hume (1739/40, p. 469) put forward a basic argument against the argumentative practice of many moralists of his (and our) time. He said that nothing about what ought (or ought not) to be the case can be deduced from what is (or is not). Hume's famous 'is-ought thesis' was very influential in modern Analytic Philosophy. Popper even claimed that it might be the most important point about ethics (1948, p. 154). Nevertheless, Hume's thesis was always controversial and still is in our time. Modern analytic ethicists have tried to demonstrate the truth of this thesis by more profound logical means, and the first step towards this end was of course to explicate Hume's thesis within a logical language. What is meant by a statement about 'is' and one about 'ought'? The prima facie most plausible explication of Hume's thesis was first suggested by Prior (1960). He introduced a dichotomic division of the set of all sentences into non-normative ones and normative ones. He characterized a normative sentence informally as a sentence which has a non-trivial normative content (gives non-trivial ethical information), whereas a non-normative one has not. But where exactly should this distinction be drawn? Three things are easy: (1.) So-called purely descriptive sentences - i.e. those which contain no occurrence of the obligation operator O - are non-normative. (2.) All sentences which are logically true have no content at all and hence are non-normative. (3.) So-called purely normative sentences - those which are built up from prime norms of the form OA with the help of logical symbols - are normative, provided they are not logically true. The problem arises with mixed sentences which have both descriptive and normative components. As Prior says, some mixed sentences, e.g. conditional

---

8 The breakdown of Popper's definition stimulated several alternative approaches to versimilitude. For a comprehensive collection see Kuipers (ed., 1987).
10 He called them 'ethical' and 'nonethical' ones (1969, p. 200).
obligations of the form □∀x(Fx → OGx), must be counted as normative, whereas others, e.g. trivial disjunctions like p ∨ Oq, should be counted as non-normative. To give an appropriate logical distinction between non-normative and normative sentences is certainly no easy task. However, one may assume that with the help of careful considerations such a - more or less appropriate - distinction can certainly be found. Thus Hume's thesis can be explicated as follows:

(8) Given an appropriate distinction between nonnormative and normative sentences, the is-ought thesis claims: No normative sentence is deducible from any consistent set of non-normative sentences.

Nothing seems to be objectionable about this explication of Hume's thesis. But now look at the following result, provable by propositional logic:

(9) Wherever the distinction between non-normative and normative sentences is drawn: (a) Hume's is-ought thesis is violated, i.e. there exist is-ought inferences; moreover (b) for every contingent non-normative sentence D the following holds: Either from D alone, or from ¬D together with another contingent non-normative sentence D', infinitely many normative sentences are deducible.

This result was first demonstrated by Prior (1960), whence we call it the Prior paradox. Several analytic ethicists, in particular Prior himself, have concluded that Hume's thesis must be abandoned. Other philosophers have concluded that mixed sentences must be excluded from the range of Hume's thesis. But recall that conditional obligations, which are the most important kind of ethical sentences, are mixed. So Prior's paradox has led analytic ethicists either to abandon or to greatly restrict Hume's thesis.

1.6 Three Kinds of Strategies Against the Paradoxes. There have been (at least) three kinds of strategies against paradoxes of this kind in Analytic Philosophy. The first strategy is that of refusing formal logic. It is sceptical and claims that the paradoxes are unsolvable. They show that formal logic is simply an inadequate means of philosophical reasoning. One should reason solely within natural language. This position, which is held e.g. by the so-called natural language philosophy, is not satisfying from the perspective of the original program of Analytic Philosophy. It would mean a relapse into the doubtful methods of traditional philosophy, which the founders of Analytic Philosophy have tried to overcome. Moreover, if Analytic Philo-

---

11 For example, Harrison (1972, p. 72), Kutschera (1982), pp. 29-31.
12 Cf. Stephan Körner (1979, pp. 377f), who distinguishes four strategies; the last three correspond to our strategies.
Sophy wants to do philosophy in a scientific manner, the position of the first strategy is not coherent, because scientific reasoning is in its most advanced parts based on mathematics, which itself is based on classical logic. So why should this logic, being so powerful in science, be avoided when one raises the issue of the foundations of science?

The other two strategies are optimistic. They believe that the paradoxes are solvable, but in different ways. According to the second strategy, the paradoxes show that classical logic is not correct. So one must change logics. This is the strategy of constructing a new (non-classical) logic. It is in particular the position of relevance logic initiated by Ackermann (1956), Anderson and Belnap (1975), which has become a well-established part of philosophical logic at the present time. Its strategy is to replace classical propositional logic by a weaker logic which avoids certain paradoxes of implication but nevertheless has the usual properties mathematicians expect from a logic. But apart from some simple paradoxical implications like ex falso quodlibet, the paradoxes mentioned above are not solved by these relevance logics. As will be shown later, it is just the fact that relevance logic keeps certain mathematical standard properties of a logic which prevents it from solving the above paradoxes.

The third strategy is that of restricting the given (classical) logic by additional relevance criteria. The theory of relevant deduction developed in what follows belongs to this strategy. It argues that one should distinguish between validity in the sense of mathematical logic and appropriateness with respect to applied arguments. The paradoxes rest on certain irrelevant deductions, which are, although mathematically valid, nonsensical and often enough harmful in applied arguments. Moreover the reason why these deductions are inappropriate as applied arguments is itself a logical one and can be defined within the framework of the underlying logic (which is usually - but not necessarily - the classical one). So the paradoxes have to be avoided by defining appropriate relevance criteria, singling out the relevant from the irrelevant deductions among all valid deductions. Such relevance criteria have often been considered to be ad hoc. However, what we will try to do here is to develop a concept of relevant deduction based on a single and well-defined criterion which covers all the paradoxes and moreover can be developed into a logical theory of relevant deduction which, in spite of its unusual mathematical properties, can be applied not only to the classical but to any kind of deductive system.

The deeper philosophical difference between the three strategies mentioned lies in their view about the relation between logic and reasoning. The first strategy claims an independence: reasoning is independent from logic (it even does not contain logic as a part). The second strategy, on the contrary, identifies both of them: logic = reasoning. The third strategy can be summarized in the slogan
"reasoning = logic + relevance". It is a 'refined' strategy in the sense that logic is viewed as a necessary but not sufficient part of reasoning - applied reasoning is based not only on criteria of logical correctness, but also on criteria of relevance.

We want to emphasize that if our view is the correct one, its consequences are far-reaching. As seen above, many philosophers have drawn radical conclusions from the paradoxes. In our time one very often hears the slogan that the original program of Analytic Philosophy which bases philosophy on systems of exact logic has failed because of these paradoxes. It has to be replaced by fundamentally different programs, like that of the structuralist philosophy of the Sneed-Stegmüller school, which replaces logical reconstructions by set theoretical ones, or by the programs of 'soft' philosophy which generally refute formal logics. If our view is the right one, then all these conclusions are ill-founded. The entire mistake of Vienna Circle philosophers was that they did not apply logics in the proper way. Imitating logicians like Tarski, Frege and Russell they tried to transfer classical logic in a direct way to natural language, not being aware of the need of restrictive relevance criteria. If the method of logical reconstruction is supplemented with a theory of relevant deduction, or of relevance in general, the paradoxes do not arise and the program works successfully.

2. Irrelevant Deductions - The Common Source of the Paradoxes

Having spoken so much about the general features of relevant deduction without having explained in what it really consists, the reader may have become impatient. Maybe this was intended. Let us now reveal the 'secret' by identifying the common source of the four paradoxes.

2.1 The Ross Paradox. With p for "the letter is posted" and q for "the letter is burned", the deductive inference underlying the Ross' paradox is

\[ (9) \ Op \ \mathcal{T}_L \ O(p \lor q) \quad [L = \text{any monotonic deontic logic}]^{13} \]

Intuitively, the particular oddity of this inference lies in the fact that the disjunctive component "q", i.e. "it is burned", has nothing to do with what the premises say, namely that the letter should be posted. The irrelevance of this disjunctive component manifests itself in the fact that instead of "it is burned" any sentence whatsoever may be disjunctively attached to the "the letter is posted". More precisely, the conclusion O(p \lor q) contains the component q which can be

---

13 The disjunction \( \lor \) is inclusive; however it goes through also if it is understood in the exclusive sense. In this case every occurrence of "p \lor q" has to be replaced by "(p \lor q) \land \neg(p \land q)\), and \( \neg(p \land q)\) has to be taken as an additional premise. On reasons mentioned in section 3.2, the defined symbol for exclusive disjunction must be replaced by its definiens in terms of \( \lor \), \( \neg \) and \( \land \) in order to detect "irrelevance".
replaced by any other formula A salva validitate of the deduction $O_{\top}T_{\top}(p \lor q)$ -- i.e. $O_{\top}T_{\top}(p \lor A)$ is valid for every $A \in \mathcal{L}$, e.g. A maybe the sentence "Suzy has a baby", in particular also $\neg p = "the letter is not posted"$. So this is our main idea: the conclusion of a given deduction is irrelevant iff the conclusion contains a component which may be replaced by any other formula, salva validitate of the deduction. Calling such a 'salva-validitate-replaceable component' an inessential one, we may say briefly: a conclusion is irrelevant iff it contains inessential components. In what follows we call a deduction with an irrelevant conclusion conclusion-irrelevant (as distinct from premise-irrelevant deductions introduced later).

Ross' inference is derived from the well-known inference of addition

$$10 \quad p \vdash_T p \lor q$$

by applying the principle ($1^*$). The inference of addition is obviously itself irrelevant, since the subformula $q$ of $O(p \lor q)$ is replaceable by any other formula salva validitate of the deduction, i.e. $O_{\top}T_{\top}(p \lor A)$ is valid for any $A$. As we will see, addition is the main culprit of irrelevant deductions, but not the only one.

2.2 Why Irrelevant Deductions are not only Useless but Harmful in Applied Contexts. Look at the following example. A satellite is going to crash onto the earth, and a journalist asks a scientist where the point of impact will be. The scientist answers "in the Atlantic Ocean", and the journalist, applying the inference of addition, writes next day in big letters in his newspaper "It follows from scientific calculations that the satellite will come down in London or in the Atlantic Ocean". The offence of this journalist does not consist in deriving a false conclusion, but in (intentionally) deriving an irrelevant one, which costs the frightened readers in London much pain and much money (since the sale rate of this journal increases enormously). The disastrous effect of the irrelevant deduction in this example is due to the fact that in practical speech situations the hearer assumes that if the speaker tells him a disjunction, say $A \lor B$, then the speakers knowledge $K_S$ about $A$ and $B$ is indeed incomplete, i.e. both $\neg A$ and $\neg B$ and thus also $A$ and $B$ are possible in $K_S$ [because, given $K_S \vdash_T A \lor B$, not $K_S \vdash_T A/B$ implies not $K_S \vdash_T \neg A$ and $\neg B$, respectively]. Equivalently expressed, the hearer assumes that $A \lor B$ is a relevant consequence of $K_S$ [because, given $K_S \vdash_T A \lor B$, it holds that $K_S \vdash_T A \lor B$ is conclusion-relevant iff not $K_S \vdash_T A$ and not $K_S \vdash_T B$]. The irrelevant conclusion together with this implicit assumption causes in the hearer an expectation which is not only irrelevant but wrong - namely that according to scientific calculations it is possible that the satellite will come down in London.

The devastating effect of the irrelevant conclusion in the Ross paradox has the same reason: if the speaker says that doing $A$ or doing $B$ is obligatory, the hearer implicitly assumes that the speakers normative principles $N_S$ are incomplete about $A$
and B, i.e. they neither require A nor B to be realized but leave a choice between them. Equivalently, the hearer assumes that \(O(A \lor B)\) is a relevant conclusion of \(N_S\) [because \(N_S \vdash T O(A \lor B)\) is conclusion-relevant iff not \(N_S \vdash L O A\) and not \(N_S \vdash L O B\)]. Thus, a solution of the Ross paradox is possible by requiring that in applied ethical contexts only relevant consequences of basic norms are acceptable.

We have recognized an important phenomenon: In applied contexts a conclusion-irrelevant argument has often harmful consequences because its evaluation is implicitly based on the wrong assumption that its conclusion is relevant and thus causes not only useless but wrong expectations in the hearer. This phenomenon, which we will meet also in the other examples, confirms the general assumption of our strategy that intuitive human thinking always combines logic with relevance. It will be further corroborated by the empirical test of section 5.3. It is for this reason that the theory of relevant deduction is not only a means of making arguments more useful or effective, but primarily a means of avoiding pathological and wrong results in logical formalizations of all sorts.

2.3 The Hesse Paradox. It is derived as follows. Let \(T\) be any contingent theory and \(S\) any contingent and true (acceptable) sentence. The following deductions

\[
\text{(11)} \quad T \vdash T \lor S \quad \text{(12)} \quad (ii) \quad S \vdash T \lor S
\]

are valid. \((T \lor S)\) is contentful because \(\neg T\) is consistent with \(\neg S\), and it is true (acceptable) because \(S\) is true (acceptable). So \((T \lor S)\) confirms \(T\) via deduction (11) according to definition (3) of deductive theory confirmation. Because of (12) and condition (4) of strengthening the confirmans, \(S\) also confirms \(T\).

But obviously both deductions have irrelevant conclusions according to our criterion. In (11), \(S\) is replaceable in the conclusion by any other sentence salva validitate; and \(T\) is replaceable in (12) in the same way. The irrelevance is harmless for the condition (4) of strengthening the confirmans, because - as one may reasonably argue - if the truth of \(S\) confirms \(T\), then any sentence \(S^*\) which logically implies \(S^*\) truth (whether relevantly or not) confirms \(T\), too. But the irrelevance is disastrous for the explication (3) of deductive theory confirmation. For it is justified to say that the truth of a deductive consequence \(C\) of \(T\) confirms \(T\) only if it is guaranteed that the facts on which the truth of \(C\) rests are relevantly connected with what \(T\) claims. But this is only true if \(C\) is a relevant consequence of \(T\). If \(C\) is an irrelevant consequence, as in the case of \(T \vdash C := T \lor S\), then it may happen that the fact which makes \(C\) true - here the fact-that-\(S\) - has nothing to do with what \(T\) says. To give an example: assume a descendant of Ptolemy would defend geocentric.

14 Discussions with Terence Parsons have shown me the importance of distinguishing between an irrelevant deduction which is just 'useless' and one which leads, together with other assumptions, to wrong results.
astronomy by arguing: "well, it does not fit to the facts of modern astronomy, but it
nevertheless has a lot of true confirmation instances; for instance it implies the
sentences "the sun rotates around the earth or grass is green", "the sun rotates
around the earth or snow is white", etc., which are all true". Needless to say, nobody
would take what our Ptolemy's descendent says seriously.

Again we recognize the phenomenon of section 2.2: an irrelevant argument leads
to a wrong result, because its evaluation relies on the wrong assumption that it would
be relevant. The wrong result here is the claim that $T \lor S$ would confirm $T$. Thus, for
an adequate notion of deductive theory confirmation it is essential that the
confirmans $C$ must be not only a valid but also a relevant consequence of $T$. With this
additional requirement the Hesse paradox no longer arises.

2.4 The Tichý-Miller Paradox. The proof that definition (6) leads to the
paradoxical result (7) consists of two subproofs, one proving that case 1 and the
other proving that case 2 of definition (6) can never be satisfied, provided the theory
is false. We start with the subproof concerning case 2, since it is this subproof which
rests on an irrelevant deduction. It runs as follows: Let $T$ be any false theory, $T^*$ any
other theory, and assume that according to the right conjunct of case 2, $(T)_f \subseteq (T^*_f)$. Thus there exists a false sentence $A \in (T^*_f)$ with $A \not\in (T)_f$. Because $T$ is false there
exists at least one false sentence $B \in (T)_f$. Now consider the implication $B \rightarrow A$.
Because of

(13) $A \models_p B \rightarrow A$

and $T^* \models_p A$ we have $T^* \models_p B \rightarrow A$, i.e. $B \rightarrow A \in Cn(T^*)$. Because $B \rightarrow A$ is true (since $B$
is false), $B \rightarrow A \in (T^*_f)$. On the other hand, because of $A \not\in Cn(T)$ and $B \in Cn(T)$ we
have $B \rightarrow A \not\in Cn(T)$ [otherwise $A \in Cn(T)$, contradicting the assumption]. Thus,$B \rightarrow A$ is a true consequence of $T^*$ which is not a consequence of $T$, whence the left
conjunct of case 2, $(T)_f \not\subseteq (T^*_f)$ is false.

But obviously $B \rightarrow A$ is an irrelevant conclusion of $A$ and hence of $T^*$ itself -- i.e.
$T^* \models_p C \rightarrow A$ holds for any $C$ (because of $\models_p (B \rightarrow A) \leftrightarrow (\neg A \lor B)$ this a variant of an
irrelevant addition). Such an irrelevant conclusion can not be counted as a real 'new
true consequence' of $T^*$, beyond $A$ itself. For example, if a physicist derives from his
theory $T$ the future arrival of a comet ($\rightarrow A$), then he will certainly not count a
sentence like "if the snow is green then $A$", or equivalently "the snow is not green or
$A$" as a further true consequence of $T^*$. Moreover, if $A$ turns out to be false our
physicist will not be proud of having derived at least the true consequence "the snow
is not green or $A$" from his theory $T^*$. Thus, the evaluation of the verisimulitude of
theories by their consequences presupposes that these consequences are relevant
ones. With this additional requirement, the inadequacy proof concerning case 2 is no
longer possible.

The subproof which shows that case 1 of definition (6) can't be satisfied is not (directly) connected with the irrelevance but with the *redundancy* of theory-consequences. It runs as follows: Assume T is a false theory, T* any other theory, and according to the left conjunct of case 1 of definition (6), \((T^*)_t \subseteq (T)_t\), i.e. there exists a true sentence \(A \in (T)_t\) with \(A \notin (T^*)_t\). Since T is false, there exists in addition a false sentence \(B \in (T)_f\). Now consider the conjunction \(B \land A\). \(B \land A\) is false, since B is false. Moreover, \(B \land A \in \text{Cn}(T)\) and hence \(B \land A \in (T)_f\), but \(B \land A \notin \text{Cn}(T^*)\) (because \(A \notin \text{Cn}(T^*)\)). Thus, \(B \land A\) is a false consequence of T which is not a consequence of \(T^*\), whence the right conjunct of case 1, \((T)_f \subseteq (T^*)_f\) can't be satisfied.

But clearly, \(B \land A\) is not a really *new* consequence of T beyond A and B itself, but simply a *redundant* repetition. An example: assume T* is Ptolemy's and T Copernicus' cosmology. Both theories have the false consequence A:="orbits are circular". In addition, T has the true consequence B="the earth moves around the sun" which T* has not. Then nobody would count the conjunction "orbits are circular and the earth moves around the sun" (=A\land B) as a further 'new' consequence of T; moreover if A becomes known to be false, neither the defender of T will be discouraged nor that of T* will be encouraged by the fact that T implies now the additional false consequence A\land B, which is not implied by T*.

To avoiding such misleading redundancies in the evaluation of verisimilitude with the help of relevant theory-consequences, it is necessary to decompose every relevant consequence into its *minimal* relevant conjunctive elements - the so-called *relevant consequence-elements*.. If the definition of verisimilitude is based on them, the inadequacy proof concerning case 1 is no longer possible, since A\land B is not a relevant consequence-element but a redundant conjunction.

### 2.5 Prior's Paradox

It rests on the following proof. Consider the mixed sentence D\lor OA, where D is any consistent descriptive sentence. We are not sure whether we should count it as non-normative or as normative. But certainly, p\lor Oq is *either* non-normative *or* normative, according to the dichotomic classification. Now consider the following two deductions, which are valid inferences in propositional logic:

\[
(14) \quad p \vdash \mathbf{T} p \lor Oq \quad (15) \quad \neg p, p \lor Oq \vdash \mathbf{T} Oq
\]

Since \(p\) is consistent, and Oq is not logically true, the following holds: If p\lor Oq counts as normative, then (14) is an example of an is-ought inference violating Hume's thesis in explication (8). If p\lor Oq counts as nonnormative, then (15) is an example of an is-ought inference. This proves that Hume's thesis is violated in every possible case, i.e. it proves (9a). Concerning the stronger claim (9b): Let D be any contingent non-normative sentence, and let \(\varphi\) denote the infinite set of propositional variables. If
D ∨ Op counts as normative, then D implies the infinitely many normative sentences D ∨ Op, with p ∈ \( \mathcal{P} \). If D ∨ Op counts as non-normative, then ¬D and D ∨ Op are contingent non-normative sentences implying the infinitely many sentences Op (p ∈ \( \mathcal{P} \)).

An obvious way to escaping this paradox, undertaken by many analytic ethicists (cf. Harrison 1972, Kutschera 1982), was to exclude mixed sentences like p ∨ Oq from the range of Hume's is-ought thesis. So instead of Prior's dichotomy, these authors introduced a trichotomous division among the set of all sentences into purely descriptive, purely normative and mixed ones and explicated Hume's thesis as follows: *No purely normative sentence which is not logically true is derivable from any consistent set of purely descriptive sentences.* We call this the *special Hume thesis SH.*

The exclusion of mixed sentences in SH is not a satisfying solution, because mixed sentences belong to the most important kinds of ethical sentences, as explained in section 1.5. A satisfying solution of Prior's paradox must show what Hume's general philosophical principle of the is-ought dichotomy implies for deductions containing mixed sentences. We need something what we will call the *general Hume thesis GH* - a natural generalization of Hume's thesis for deductions containing mixed sentences - in particular, mixed conclusions (cf. also Morscher 1984, cf. 426, 430).

We obtain the idea of this GH if we recognize that deduction(14) with purely descriptive premises and a mixed conclusion has an irrelevant conclusion in a special sense: the propositional variable q of the conclusion which lies within the scope of the obligation operator can be replaced salva validitate by any other formula. So what the conclusion says about *ought* is completely independent from what the premises say about *is.* If we look at further examples of deductions \( \Gamma \vdash A \) with purely descriptive premises \( \Gamma \) and a mixed conclusion A in standard deontic logics, we realize that they also have this salient feature: the propositional variables (atomic formulas) in A can be uniformly replaced by any other formulas on all those occurrences which lie within the scope of an obligation operator O, salva validitate of \( \Delta \vdash A \). For instance, consider deduction (16) - a new kind of irrelevant deduction (different from addition) which rests on introducing a superfluous logical truth as a conjunct (more on this in section 3.1). With (16) also every deduction of the form (16*) is valid:

\[
\begin{align*}
(16) & \quad p \vdash \Gamma, p \land (O(p \land q) \rightarrow Op) \\
\text{(16*)} & \quad p \vdash \Gamma, p \land (O(A \land B) \rightarrow OA)
\end{align*}
\]

(with \( L \) = the minimal deontic logic)

We speak here of a *completely ought-irrelevant* conclusion of a given descriptive premise set (this is a specialization of our more general irrelevance idea above,
which refers only to some occurrences). Based on this observation, the formulation of our general Hume thesis GH is now at hand: it claims that in every deduction \( \Gamma \vdash \text{A} \) with purely descriptive premises and a mixed conclusion, the conclusion is completely ought-irrelevant. Indeed if GH were true, this would fit perfectly with Hume's general principle of the is-ought dichotomy. The exciting question whether GH is indeed true and in which logics this is the case will be discussed in section 5.2.2. The ethical importance of GH is demonstrated by the following example. Assume an ethicist derives from a purely descriptive premise set the conclusion "if x is a murderer, then x should be punished by the death penalty" in a standard deontic logic L. If we knew that GH holds in L, then without even knowing which descriptive premises the ethicist has used we can conclude that the deduction is either incorrect or every sentence of the form "if x is a murderer, then A" follows from his premise set, in particular "if x is a murderer, x should not be punished" - which destroys the ethical relevance of our ethicist's argument.15

3. The Theory of Conclusion-Relevant Deduction

3.1 The Definition of Relevant Conclusion. In what follows the concept of relevant conclusion is developed in a logically systematic way. The main kind of irrelevant conclusion in the previous section was the addition \( p \vdash p \lor q \) and its 'derivatives' obtained by various logical transformations, like

\[
(17) \quad p \vdash q \rightarrow p; \quad p \vdash \neg p \rightarrow q; \quad p \wedge q \vdash q \lor (p \wedge q); \quad p \wedge q \vdash (p \lor q) \wedge (q \lor r); \quad \neg p \vdash \neg (p \wedge q)
\]

- the underlined subformulas are replaceable. It is characteristic of these 'derivatives' of addition that always single occurrences of subformulas are replaceable salva validitate. But the replacement of single subformula occurrences is not enough, as the second kind of irrelevant conclusions due to superfluous logical truths shows, which we met in example (16) of section 2.5. Derivatives of this kind are

\[
(18) \quad p \vdash q \wedge (q \lor \neg q); \quad p \vdash (p \lor q) \wedge (p \lor \neg p); \quad p \vdash (p \rightarrow q) \lor (q \rightarrow q), \quad p \vdash p \land (q \lor \neg q)
\]

etc. Here the conclusion has a subformula which is simultaneously replaceable on both of its (underlined) occurrences salva validitate, but not on its single occurrences alone - i.e. \( p \vdash p \land (A \lor \neg A) \), but not \( p \vdash p \land (A \lor \neg q) \), etc. Also superfluous logical truths have been the source of paradoxes. One example is the paradox in the

---

15 Let us remark that a modification of our general Hume thesis is also applicable to deduction (15) above with mixed premises and a purely normative conclusion, as a special case of the idea of relevant premises explained in section 4. For details cf. Schurz (1989), pp. 295ff and (1990b).
logic of 'epistemically neutral' perception sentences discussed by Barwise (1988, p. 24), which rests on an application of the intuitively plausible 'principle of logical equivalence' to the irrelevant equivalence $F_a \vdash (F_a \land G_a) \lor (F_a \land \lnot G_a)$.\(^{16}\)

Finally there exist also more complicated cases of irrelevant conclusions, the proof of which involves both addition and superfluous logical truths like

\[(19) \quad p \rightarrow q \land (p \land q) \land (q \land \lnot r); \quad p \land q \land (p \land q) \lor (q \land \lnot r) \].\]

Here, too, the underlined subformula is simultaneously replaceable on both occurrences, but not on the single one alone. By applying our replacement criterion to *multiple* occurrences we are able to cover all these cases by one simple relevance criterion.

We will give our relevance definition not only for propositional but also for first-order predicate logic, because this admits much more interesting applications. So far we have based our relevance criterion on the salva-validitate replacement of subformulas of any kind, i.e. possibly complex ones, by any other sentences. This condition has a relatively complicated formulation in predicate logic because of the necessary conditions prohibiting confusions of variables. To give some hints as to what is going on, consider the irrelevant deduction

\[(20) \quad \forall x(Fx \rightarrow Gx), \quad Fa \vdash (G_a \land \lnot \forall z Rzu) \lor (G_a \land Ryu) \]

obtained by introducing the superfluous logical truth ($\forall z Rzu \rightarrow Rxu$) and applying the law of distribution. If we view the underlined atomic formulas as different term substitution instances of the same atomic formula $Rx_1x_2$, namely $Rx_1x_2[z/x_1, u/x_2]$ and $Rx_1x_2[y/x_1, u/x_2]$, we see that they are simultaneously replaceable salva validitate by the corresponding term instances $B[z/x_1, u/x_2]$ and $B[y/x_1, u/x_2]$ for *any* other formula $B$ *provided* these term substitutions are correct and they do not contain a free variable different from $x_1$ and $x_2$ which gets bound by a quantifier after the replacement - here possibly the quantifier $\forall z$. Based on this observation, a general formulation of conclusion-relevance for predicate logic is possible based on the notion of the *occurrence-restricted substitution* of subformulas by any other formulas, which is obtained as a generalization of Kleene's notion of uniform substitution of formulas for predicates.\(^{17}\) We don't present this complicated definition. Fortunately an important simplification is possible. The definition of relevant

\(^{16}\) Roughly, the principle states that if a person sees a scene described by proposition A, and A is logically equivalent to B, then this person sees also the scene described by proposition B. Observe that these perception sentences are 'epistemically neutral' in the sense that they don't imply the person's belief that A (or B, respectively) is true.

\(^{17}\) Kleene 1971, 155-162; for details cf. Schurz 1989, pp. 52ff, def.s 1, 2 and 5.
conclusions based on the replacement of (possibly complex) subformulas by any others is provably equivalent to the simpler definition based on the replacement of predicate letters by any other predicate letters of the same arity. This will be the final version of our relevance definition, which can be stated in an informal version as follows:

(21) (Informal Definition) Assume $\Gamma \vdash A$. Then $A$ is a relevant conclusion of $\Gamma$ iff no predicate in $A$ is replaceable on some of its occurrences by any other predicate of the same arity, salva validitate of $\Gamma \vdash A$. Otherwise, $A$ is an irrelevant conclusion of $\Gamma$.

If $A$ is a relevant/irrelevant conclusion of $\Gamma$, we also say, the deduction $\Gamma \vdash A$ is conclusion-relevant/irrelevant, in symbols $\Gamma \vdash^F A / \Gamma \vdash^i A$ (resp.).

A further important simplification is due to the following theorem: Assume $\Gamma \vdash A$. Then: Some occurrences of a predicate are replaceable salva validitate by any other predicate of the same arity iff these occurrences are replaceable salva validitate by some new predicate of the same arity, i.e. one which does not occur in $\Gamma \cup \{A\}$. This makes it especially simple to test whether a given deduction $\Gamma \vdash A$ is conclusion-irrelevant: instead of investigating 'all possible' replacements of predicate occurrences, it is enough to find just one replacement by a new predicate in order to prove that $\Gamma \vdash A$ is conclusion-irrelevant.

Here are some examples of relevant and irrelevant conclusions.

Examples of relevant conclusions:
In propositional logic: $(p \land q) \models^F p, p \rightarrow q \vdash^F q; \neg p, p \rightarrow q \vdash^F \neg p; p \rightarrow q, q \rightarrow r \vdash^F p \rightarrow r; p \models^i p; p \models^F \neg p; p \models^F (q \rightarrow r); p \models^F (p \land q) \rightarrow (p \lor q); p \models^F (p \land q) \rightarrow (p \lor q); p \models^F (p \lor q) \rightarrow (p \land q); p \models^F (p \land q) \rightarrow (p \land q)$. In predicate logic: $\forall x Fx \models^F \exists x Fx; \forall x Fx \models^i Fa; \forall x Fx \models^F \exists x Fx; \forall x Fx \models^F \exists x Fx; \forall x Fx \models^F \exists x Fx$;

Examples of irrelevant conclusions (Underlined = replaceable; double underlined = replaceable independently from single underlined):
In propositional logic: $p \models^F p \lor q; p \rightarrow q \models^F (p \lor q) \rightarrow q; p \rightarrow q \models^F (p \land q) \rightarrow (q \land q); p \land q \models^F (p \land q) \rightarrow (q \land q); p \land q \models^F (p \lor q) \rightarrow (q \land q); p \land q \models^F (p \land q) \rightarrow (q \land q)$; In predicate logic: $\forall x Fx \models^F \forall x (F \land G) x; \forall x (F \rightarrow G) x \models^F \forall x (F \rightarrow G) x; \forall x (F \rightarrow G) x \models^F \forall x (F \rightarrow G) x; \forall x (F \rightarrow G) x \models^F \forall x (F \rightarrow G) x; \forall x (F \rightarrow G) x \models^F \forall x (F \rightarrow G) x; \forall x (F \rightarrow G) x \models^F \forall x (F \rightarrow G) x$;

\[18\] For a proof cf. Schurz (1989, pp. 57ff, prop.s 2, 3); Schurz/Weingartner (1987, prop. 1).
\[19\] This informal definition was first given in Schurz/Weingartner (1987).
\[20\] Its proof rests on simple uniform substitution; cf. Schurz (1989, pp. 57f, prop. 2.3).
We finally explain how the informal definition above can be stated in the standard mathematical form of an inductive definition as the basis of a general logical theory of conclusion-relevant deduction. The core concept is that of occurrence-restricted substitution of predicates. For $n \geq 0$, let $\mathbb{R}^n$ be the set of $\mathcal{L}$’s n-place predicates; and for every $A \in \mathcal{L}$, let $\mathbb{R}(A)$ denote the set of predicates occurring in $A$. A predicate-substitution function is a function with the domain $\bigcup_{n \geq 0} \mathbb{R}^n$ and $\pi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ for all $n \geq 0$.

For every formula $A \in \mathcal{L}$, $\pi A$ denotes the result of the uniform substitution of $\pi R$ for $R \in \mathbb{R}(A)$ in $A$ and is inductively defined as usual $[(\pi Rx_1 \ldots x_n) = (\pi R)x_1 \ldots x_n$, $\pi \neg B = \neg \pi B$, $\pi (B \lor C) = (\pi B \lor \pi C)$, $\pi \forall xB = \forall x \pi B$, $\pi oB = o \pi B$ for $o$ a modal operator].

Assuming that the occurrences of predicates in any formula $A \in \mathcal{L}$ are numbered from left to right, starting with 1, we can represent a set of predicate occurrences $\mu$ simply as a set of natural numbers $\mu \subseteq \mathbb{N}$. The result of applying the $\mu$-restricted $\pi$-substitution in $A \in \mathcal{L}$, denoted by $\pi \mu A$, is then simply defined as the result of replacing, for every $i \in \mu$, the $i$th predicate-occurrence $R$ in $A$ by $\pi R$, provided $A$ has at least $i$ predicate occurrences (otherwise nothing is replaced). The inductive definition is as follows: (1) $A = Rx_1 \ldots x_n$ $(n \geq 0)$: If $1 \notin \mu$, then $\pi \mu A = A$, and if $1 \in \mu$, then $\pi \mu A = A$. (2) $\pi \mu \neg B = \neg \pi \mu B$. (3) $\pi \mu (B \lor C) = (\pi \mu B \lor \pi \mu C)$, $\mu$ with $\mu^* = \{i - k_B \mid i \in \mu \text{ and } i > k_B\}$, where $k_B$ = the number of predicate occurrences in $B$. (4) $\pi \forall xB = \forall x \pi \mu B$. (5) $\pi \mu oB = o \pi \mu B$, for $o$ a modal operator.

An example: Assume $\pi F = H$, $\pi G = I$ and $\pi R = Q$, $\mu = \{1, 3\}$, $\mu^* = \{2, 4\}$, and $A = (\forall x(Fx \lor Gx) \rightarrow \exists y(\neg Fy \land Rx y))$. Then $\pi \mu A = (\forall x(Hx \lor Gx) \rightarrow \exists y(\neg Hy \land Rx y))$, $\pi \mu^* A = (\forall x(Fx \lor Ix) \rightarrow \exists y(\neg Fy \land Qxy))$, $\pi A = (\forall x(Hx \lor Ix) \rightarrow (\exists y(\neg Hy \land Qxy))$.

We obtain as the final definition:

(22) (Formal Definition): $\Gamma \vdash^\mathbb{F}_A \{ \Gamma \vdash^\mathbb{F}_A \} \text{ iff } \Gamma \vdash \Gamma \vdash \mu A$ and there exists no [some] non-empty $\mu \subseteq \mathbb{N}$ such that $\Gamma \vdash \pi \mu A$ holds for every $\pi$. 21

We call the irrelevance of a conclusion according to definition (22) also a partial irrelevance, since it is enough for being irrelevant that there exist some replaceable predicate occurrences, i.e. that $\mu$ is just non-empty. There are two successive strengthenings of this notion. The first is the notion of a completely O-irrelevant conclusion introduced in section 2.5 in order to solve the Prior paradox. Here every occurrence of a predicate in the conclusion lying within the scope of an obligation operator is replaceable salva validitate, i.e. $\mu$ coincides with the set of all predicate occurrences in O-scopes. Formally we define this notion as follows. For any formula $A \in \mathcal{L}$ and predicate substitution function $\pi$, $\pi^O A$ - the result of performing the O-restricted $\pi$-substitution in $A$ - denotes the result of replacing every predicate

21 A technical detail: "for every $\pi$" has to be understood here in a language-independent way, i.e. "for every $\pi : \mathbb{R}^* \rightarrow \mathbb{R}^*$ where $\mathbb{R}^*$ may be any extension of $\mathcal{L}$’s predicates."
R ∈ S(A) by πR on exactly those occurrences lying within the scope of some O. It is inductively defined as follows: (1.) \( π^0Rx_1 \ldots x_n = Rx_1 \ldots x_n \), (2.) \( π^0B = ¬π^0B \); (3.) \( π^0(B \lor C) = π^0B \lor π^0C \); (4.) \( π^0\lor xB = \lor xπ^0B \); (5.) \( π^0oB = oπ^0B \) for o a modal operator different from O, and [importantly] (6.) \( π^0OB = OπB \). We define now: The deduction \( Γ^tieq \pi^0A \) has a completely O-irrelevant conclusion iff \( Γ^tieqπ^0A \) holds for every π.

To give an example, the deduction (23)

\[
O(p \rightarrow q), Op \vdash L (Oq \lor Or) \land O(\bar{s} \lor \bar{s})
\]

(24) is only partially conclusion-irrelevant but not completely O-irrelevant, because only some (the underlined ones) but not all variable-occurrences in the conclusion lying within an O-scope are replaceable salva validitate. On the other hand, (24) is completely O-irrelevant.

A further strengthening is the notion of a completely irrelevant conclusion, in which all occurrences of predicates are replaceable salva validitate. We define this notion simply by applying π uniformly to the entire conclusion: \( Γ^tieq \pi^0A \) has a completely irrelevant conclusion iff \( Γ^tieqπ^0A \) for every π. The following theorem is provable by simple means (Schurz 1989, pp. 62f): \( Γ^tieq \pi^0A \) has a completely irrelevant conclusion iff either Γ is L-inconsistent or A is an L-theorem, where L may be any intuitionistic or classical modal logic which is Haldén-complete (cf. footnote 6). This means that our notion of a complete conclusion-irrelevance covers exactly the two "classical" paradoxes of implication, the elimination of which is the main target of relevance logics:

\[
(25) \text{falsum quodlibet: } p \land ¬p \vdash L q \quad (26) \text{verum ex quodlibet: } p \vdash L q ∨ ¬q
\]

In other words, the irrelevancies covered by relevance logics are very special case of those covered by our theory of relevant deduction - the special case of complete irrelevance.

Finally we mention that one may distinguish also between corresponding notions of relevant conclusions: If a conclusion is not partially irrelevant, i.e. relevant in the sense of def. (21), it can be called completely relevant; if it is not completely O-irrelevant, it is partially O-relevant; if it is not completely irrelevant, it is partially relevant.

3.2 Logical Properties and Comparison with Relevance Logic. As already mention in section 1.6, relevance logic identifies relevance with validity and constructs a 'new' logic, whereas our theory defines relevant deductions as a proper subclass of all valid deductions by restrictions within the given logic L. As a result of these different purposes, the general difference between the two approaches is this: Relevance logic has 'nice' logical properties; in return it solves hardly any paradoxes, except the classical falsum quodlibet and verum ex quodlibet. On the contrary, the conclusion-
relevant deduction does not have 'nice' logical properties; in return it solves a lot of paradoxes (indeed, I think all paradoxes which have to do with 'logical' irrelevance).

Before comparing our notion of conclusion-relevant deduction \( \mathcal{L} \) with relevance logics, we must pay attention to the fact that relevance logic is prima facie a system of relevant implication \( \rightarrow_r \). However, there exists also a notion of deduction in relevance logics which we denote by \( \vdash_r \) and which satisfies the usual deduction theorem \( A_1, \ldots, A_n \vdash_r B \) iff \( 'A_1 \land \ldots \land A_n \rightarrow_r B' \) is provable (cf. Anderson/Belnap 1975, p. 277f). So, a straightforward comparison of our \( \mathcal{L} \) with \( \vdash_r \) is possible. We will assume that the deduction relation \( \mathcal{L} \) to which our \( \mathcal{L} \) is relativized is that of classical logic (furthermore, we will simply say "relevant deduction" instead of "conclusion-relevant deduction").

(1.) The deductions \( \vdash_r \) of relevance logic are closed under substitution (like every standard logical deduction relation), whereas relevant deduction \( \mathcal{L} \) is not. On the other hand, the irrelevant deductions \( \mathcal{L} \) or our theory are closed under substitution whereas the corresponding classically valid inferences which are invalid in relevance logics are not. So here is the first crossroad: In relevance logics, relevance is closed under substitution, i.e. is a matter of form, whereas in our theory, irrelevance is closed under substitution, i.e. is a matter of form. Let us demonstrate by way of an example why we think that approach is the correct one. Consider the 'classical' ex falso quodlibet \( p \land \neg p \vdash_r q \). The irrelevance of this deduction is manifested in the fact that it is valid only because of the form of the premises, independent of what is stated in the conclusion. Therefore, not only \( p \land \neg p \vdash_r q \) but every deduction of the form \( A \land \neg A \vdash_r B \) must be counted as irrelevant, in particular also \( p \land \neg p \vdash_r p \). In the same way it can be argued for every irrelevant deduction that its instances must be irrelevant as well.

It holds in our approach that \( A \land \neg A \vdash_r B \) for every \( A, B \in \mathcal{L} \); i.e. not only \( P \land \neg p \vdash_r q \), but also \( p \land \neg p \vdash_r p \). Very different in relevant logic. Here \( p \land \neg p \vdash_r q \) is invalid, but \( p \land \neg p \vdash_r p \) is valid, and indeed must be valid there, because it is a substitution instance of the valid deduction \( p \land q \vdash_r p \), and \( \vdash_r \) is closed under substitution (Anderson/Belnap 1975, p. 158). On the other hand, the fact that \( p \land q \vdash_r p \) although \( p \land \neg p \vdash_r p \) is no problem for us because our \( \mathcal{L} \) is not closed under substitution.

(2.) The law of addition (Add), which is the 'paradigm' of an irrelevant deduction in our theory, is valid in relevant logics, i.e. \( p \vdash_r p \lor q \) holds (Anderson/Belnap 1975, p. 154). This, of course, implies that relevance logic can't solve all the main cases of paradoxes, because they all rest on the irrelevance of addition. But is has also another interesting consequence: disjunctive syllogism (DS) \( p \lor q, \neg p \vdash q \) is invalid in relevance logics, whereas it is relevant in our theory, \( p \lor q, \neg q \vdash \mathcal{L} p \). We can't see any reason for viewing (DS) as irrelevant. Anderson/Belnap exclude it on the ad hoc ground that (DS) together with (Add) would make ex falso quodlibet \( p \land \neg p \vdash q \) generally provable (1975, p.164f). In our view, what is irrelevant is not (DS) but (Add).
(3.) Two further important properties of 'standard' deduction relations are the cut rule: if $\Gamma \vdash A$ for all $A \in \Delta$ and $\Delta \vdash B$, then $\Gamma \vdash B$ (this implies transitivity), and monotonicity: if $\Gamma \vdash A$ and $\Gamma \subseteq \Delta$, then $\Delta \vdash A$ (cf. Rautenberg 1979, p.75f). Both properties are satisfied by $\vdash_r$ of relevance logics, but are not satisfied by our $\vdash^{cr}_L$. A counter-example to transitivity (cut rule) of $\vdash^{cr}_L$ is $(p \lor q) \land r \vdash^{cr}_L p \lor (q \land r)$, $(p \lor q) \land (p \lor r)$; a counterexample to monotonicity is $p \lor q \vdash^{cr}_L p \lor q$, but $\{p, p \lor q\} \vdash p \lor q$. The facts that $\vdash^{cr}_L$ is not closed under substitution, does not satisfy the cut rule and is non-monotonic are strong reasons for our strategy to distinguish between relevance and validity and not to view relevant deduction as a deduction relation in the standard sense. Nevertheless, the investigation of $\vdash^{cr}_L$ in the perspective of the general theory of deduction relations is highly interesting, for two reasons. First, there exist also 'non-standard' deduction relations and it is important to know whether $\vdash^{cr}_L$ - the 'relevant restriction' of the classical deduction relation $\vdash_L$ - cannot at least be seen as a 'deviating' deduction relation. Second, our concept of relevant deduction is generalizable to any kind of deduction relations, and it is very interesting to investigate it for deduction relations $\vdash$ which are weaker than the classical $\vdash_L$. Indeed we will see that in this case things may indeed change, and it may happen that for a weaker $\vdash$ its relevant restriction, $\vdash^{cr}_L$ satisfies standard properties of deduction relations. We will investigate this latter question in section 5.4, here we focus on the relevant restriction $\vdash^{cr}_L$ of the classical $\vdash_L$.

Can $\vdash^{cr}_L$ be viewed as a 'deviating' deduction relation? As mentioned above, irrelevant deductions may be useful in mathematical proofs. Indeed, it may happen that the proof of a relevant deduction contains an irrelevant intermediate proof step. For instance, the deduction $p, (p \lor q) \rightarrow r \vdash_L r$ is relevant, but its proof contains the irrelevant step (3); (1) $p$ [Prem], (2) $(p \lor q) \rightarrow r$ [Prem], (3) $p \lor q$ [Add, applied to 1], (4) $r$ [MP, applied to 2 and 3]. However, this proof may be replaced by another proof which contains only relevant proof steps, namely: (1) $p$ [Prem], (2) $(p \lor q) \rightarrow r$ [Prem], (3) $p \rightarrow r$ [from (2) by the relevant rule $(p \lor q) \rightarrow r \vdash^{cr}_L p \rightarrow r$], (4) $r$ [MP from (3) and (1)]. An important question is this: does every relevant deduction have a proof consisting only of relevant steps, or are there some relevant deductions which are only provable with the help of irrelevant steps? The former alternative would be an argument in favour of viewing $\vdash^{cr}_L$ as a 'deviating' deduction relation, the latter would be a strong argument against it.

Our question has a stronger and a weaker form. In the stronger form, it means: is there a correct and complete axiomatization of $\vdash^{cr}_L$, i.e. is there a decidable set of axioms and rules [e.g. in the form of a Gentzen-like system] from which all relevant deductions and only those are provable? The 'weakest' form of axiomatizability is recursive enumerability. For the latter, the question is answered by a theorem due to Kit Fine:
(27) (Theorem:) For every recursively enumerable (r.e.) logic $L$ closed under substitution and containing classical propositional logic: $\vdash_{\text{cr}}$ is r.e. only if $L$ is decidable.

Proof: It holds that $p \land \lnot A \vdash_{\text{cr}} p$ iff not $\vdash A$, provided $p \notin \mathcal{R}(A)$. Direction $\Leftarrow$: If $\vdash A$, then $p \land \lnot A$ is inconsistent whence $p \land \lnot A \vdash q$ for every $q \notin \mathcal{R}$, i.e. $p \land \lnot A \vdash_{\text{cr}} p$. Direction $\Rightarrow$: If $p \land \lnot A \vdash_{\text{cr}} p$, then $p \land \lnot A \vdash q$ for $q \notin \mathcal{R}(A,p)$, so $(p \lor \lnot p) \land \lnot A \vdash q \lor \lnot q$ [by uniform substitution and $p$, $q \notin \mathcal{R}(A)$]; so $\lnot A$ is $L$-inconsistent, i.e. $\vdash A$. From $p \land \lnot A \vdash_{\text{cr}} p$ iff not $\vdash A$ it follows that if $\vdash_{\text{cr}}$ is r.e., then also the set of $L$-non-theorems is r.e. Because $L$ is also r.e., $L$ is decidable. Q.E.D.

So for all logics $L$ containing classical first-order predicate logic, $\vdash_{\text{cr}}$ is not r.e. and hence has no correct and complete axiomatization.

The weaker form of the question is this: does there exist an axiomatization of $\vdash_{\text{cr}}$ such that every relevant deduction has a proof containing only relevant steps, even if there might also be some irrelevant deductions provable with it? In short, does there exist a complete and relevant axiomatization of $\vdash_{\text{cr}}$, even if it is not a correct one? This question is not answered yet. A further important question is whether there exists some interesting semantic representation of $\vdash_{\text{cr}}$.

Besides the question whether $\vdash_{\text{cr}}$ can be seen as a 'non-standard' deduction relation, there are further important questions about $\vdash_{\text{cr}}$. One example is the question to what degree $\vdash_{\text{cr}}$ depends on the choice of primitive propositional connectives. As shown in Schurz (1989, prop.5, pp. 73f), every set of primitive connectives among $\{\lnot, \lor, \land, \rightarrow\}$ is adequate for $\vdash_{\text{cr}}$, whereas $\leftrightarrow$ and exclusive disjunction are not adequate for $\vdash_{\text{cr}}$ because they may 'hide' certain irrelevancies. A second important question about $\vdash_{\text{cr}}$ is this: Is the set of relevant consequences of a sentence $A$ always logically equivalent to $A$ itself? If this were true then no information (in the sense of classical $L$) would get lost by replacing the classical deductive consequence class of a sentence by the set of its relevant consequences. This might be regarded as an important adequacy requirement for the application of $\vdash_{\text{cr}}$ to verisimilitude. It is easily proved that this requirement is satisfied iff the following holds:

(28) (Theorem/Conjecture:) For all $A \in \mathcal{L}$ there exists a $B \in \mathcal{L}$ such that $\vdash_{\mathcal{T}} A \iff B$ and $B \vdash_{\text{cr}} B$.

(28) has been proved for classical propositional logic, and it is easily proved for second-order predicate logic, but so far it is only a conjecture for first-order predicate logic.

\[\text{---}
\]

22 As a step in this direction, the notion of a "relevant model entailment" was introduced in Schurz (1987).
3.3 Comparison with other Relevance Criteria. There are some other criteria of relevant deduction which are located within our strategy of restricting classical logic. We show how our $\mathcal{I}^{cf}$ is related to them.

The Wright-Geach-Smiley criterion\textsuperscript{23} says that $A \rightarrow B$ is a relevant implication (or $A \mathcal{I} T B$ a relevant deduction, respectively) iff it is possible to prove $A \rightarrow B$ without proving $\neg A$ or $B$. Because of the formulation "possible", this criterion is equivalent to the following: $A \mathcal{I} T B$ is relevant iff neither $\mathcal{T} \neg A$ nor $\mathcal{T} B$. Thus, this criterion coincides with the exclusion of complete conclusion-irrelevance in the sense of section 3.1.

The Aristoteles-Parry-Weingartner criterion\textsuperscript{24} (called A-criterion in Wein-gartner/Schurz 1986) says that a deduction $\Gamma \mathcal{I} T A$ is relevant iff every propositional variable or predicate occurring in the conclusion occurs also in the premises, i.e. iff $\mathcal{R}(A) \subseteq \mathcal{R}(A)$. This criterion is properly contained in $\mathcal{I}^{cf}$; i.e. if $\Gamma \mathcal{I}^{cf} A$, then $\Gamma \mathcal{T} A$ satisfies the APW-criterion, but not vice versa. For instance, $p \mathcal{T} p \lor \neg p$, or $\forall x(Fx \rightarrow Gx)$, $Fa \mathcal{T} Ga \lor \exists x Fx$ are relevant according to the APW criterion but irrelevant in our theory. The deeper reason behind this is that APW-irrelevant deductions are not closed under substitution, which should be the case as argued in section 3.2.

The idea of 'salva validitate' replacements was invented by Körner, so our theory is much indebted to him. His idea was that a valid formula is 'essential' (or relevant, in our terminology) iff it contains no inessential components. He calls a component of a valid formula 'inessential' iff "it can salva validitate be replaced by its negation, i.e. if replacing every occurrence of the component by its negation leads from a valid to a valid logical implication".\textsuperscript{25} The formulation of Körner is mistaken because according to it, by uniform substitution every component of a valid formula would be inessential. Cleave(1973/74) corrected this mistake by applying Körner's salva validitate replacement only to single subformula-occurrences: A valid formula (in particular: implication, deduction) is 'essential' iff no single occurrence of a subformula is replaceable by its negations salva validitate. A generalization of this criterion to predicate logic was undertaken in Schurz (1983b), called K-criterion in Weingartner/Schurz(1986) and henceforth.

There is no simple logical relation between $\mathcal{I}^{cf}$ and the K-criterion. The most important differences, and I think improvements, of $\mathcal{I}^{cf}$ compared to the K-criterion are these. (1.) In predicate logic it may happen that a predicate of the conclusion is replaceable by its negation without being replaceable by any other predicate salva validitate. For instance, in the deduction $\neg \forall x Fx \land \forall x \neg Fx \mathcal{T} \exists x Fx \land \exists x \neg Fx$, $Fx$ is

\textsuperscript{23} This name was chosen by Anderson/Belnap (1975, pp. 215f) because these three authors suggested this criterion in different places.

\textsuperscript{24} I choose this name because this criterion was formulated by Parry (1933) and independently suggested by Weingartner (1985) who emphasized that it is satisfied by the Aristotelian syllogistics.

\textsuperscript{25} See Körner (1979, p. 378). For similar formulations cf. (1959, p. 63) and (1947), where he introduced his idea for the first time.
replaceable in the conclusion by $\neg Fx$, salva validitate, but it is not replaceable by any other atomic formula. So this deduction is conclusion-irrelevant according to the K-criterion, but it should be considered as conclusion-relevant, because here the premises are relevantly connected with every component of the conclusion. So the K-criterion leads to inadequate, namely too strong, results in predicate logic. An appropriate criterion must not be based on replacements-by-negation, but on replacements-by-any-other-formula. (2.) Whereas the K-criterion replaces only single occurrences, $\not{\mathsf{F}}x$ replace also several occurrences in order to eliminate the irrelevancies mentioned in section 3.1. For instance, the irrelevant deductions $p \not{\mathsf{T}} p \wedge (q \vee \neg q)$, $p \rightarrow q \not{\mathsf{T}} ((p \wedge r) \rightarrow (q \wedge r))$ or $p \wedge q \not{\mathsf{T}} (p \wedge r) \vee (q \wedge \neg r)$ are all relevant according to the K-criterion. (3) Finally, the K-criterion does not distinguish between premise-relevance and conclusion-relevance. For instance, simplification $p \wedge q \not{\mathsf{T}} p$ is also irrelevant according to it (moreover, it is formulated not only for valid implications or deductions but for any valid formulas). In the next section we will see that for good reasons, conclusion-relevance and premise-relevance have to be treated differently. I will introduce there a notion of relevant premises within our approach. However let me remark that my feelings about premise-relevance are much less 'certain' than about conclusion-relevance.

4. Extension of the Approach: Relevance of Premises

4.1 Why Premise-Relevance has Restricted Applications. The basic idea of relevant premises is simple: a premise $P \in \Gamma$ is relevant in the deduction $\Gamma \not{\mathsf{T}} A$ iff $P$ is necessary for deducing $A$, i.e. iff $\Gamma-\{P\} \not{\mathsf{T}} A$ is not valid also. According to this idea, $p,q \not{\mathsf{T}} p$ is the basic pattern of a premise-irrelevant deduction. The most important difference between premise-relevance and conclusion-relevance is that conclusion-relevance is required in every kind of applied argument, whereas premise-relevance is appropriate only in special cases of applications. The difference is also confirmed by the empirical test in section 5.3, according to which conclusion-irrelevant arguments are generally judged as 'invalid' by our intuitive thinking, whereas premise-irrelevant arguments are mostly judged as 'valid'. There are a lot of situations in which deductive arguments with irrelevant premises are perfectly reasonable. One example is the evaluation of a theory by deducing empirical consequences, as in verisimilitude. A deduction proving that, say, the law of plastic collision is a consequence of Newtonian physics is certainly relevant in an intuitive sense, although the gravitational law, an element of Newtonian physics, is unnecessary for deriving this special consequence. But this does not matter here because the premise set of Newtonian physics is allowed to contain, and in fact should contain, much more information than special laws derived from it. All that counts here is that the derived conclusion, the special law, is relevant. Another example is deductive
justification, where we justify the belief-that-C by deriving C from a premise set \( \Gamma \) which is already justified-to-believe: also here, the relevance of the premises in \( \Gamma \) is not a necessary requirement. The same holds for deductive predictions - deductive justifications of propositions which say something about the future. In all these and other examples, the requirement of relevant premises has at most the function of avoiding redundancies in the premises, but redundant premises don't make a deductive justification or predication unreasonable or senseless in applied arguments.

4.2 Paradoxes due to Irrelevant Premises. However there are also special contexts in which deductions must have relevant premises. The most important example is deductive theory confirmation: Recall explication (4), and add the following, intuitively very plausible, "general consequence-condition", proposed by Hempel (1965, p. 31):

(29) If sentences S confirms theory T, then S confirms also every logical consequence of T.

Then we immediately obtain the following paradoxical result:

(30) For every sentence S: If S confirms at least one theory, it confirms every consistent theory.

The proof is simple and comes in two variants. In the first variant, it is assumed that theories are sets of sentences. The proof of (30) then rests on:

(31) (i) if \( T \models S \), then \( T \cup T^* \models S \)

Assume S confirms T by (4). Then S is true [acceptable], not \( T \), not \( T \models \neg T \) and \( T \models S \). Thus \( T \cup T^* \models S \) by (31). So S confirms \( T \cup T^* \) by (4). Because \( T \cup T^* \models T^* \), S confirms also \( T^* \) by (29). But obviously the deduction (31) contains irrelevant premises, namely the elements of \( T^* \), which are superfluous for the deduction of S. The reason why irrelevant premises must be excluded in deductive theory confirmation is clear: if \( T \models S \) and S is true, then we are only justified in saying that S confirms T if indeed 'every part' of T was necessary to derive S. Otherwise S does not confirm T as a whole, but only that 'part' of T which was necessary to derive S.

The second variant of the proof of (30) assumes theories to be sentences. It rests then on:

(32) (i) if \( T \models S \), then \( T \land T^* \models S \)
Again, if $S$ confirms $T$ then $S$ confirms also $T \land T^*$ by (32), whence it confirms also $T^*$ because of $T \land T^* \mathcal{L} T^*$ and condition (29). Of course also the deduction (32) contains an irrelevancy in its premise set, but now it is not an entire premise but only a premise-conjunct which is superfluous for deducing $S$. Thus, we must apply the relevance criterion not only to the premises but also to their conjuncts.

Another kind of application where relevant premises are necessary are deductive-nomological explanations. From a deductive-nomological explanation "L,A/E" (L=law, A=antecedens, E=explanandum) it is usually required that A figures, at least potentially, as a cause of E. This is only possible if the information contained in A, and every 'part' of it, is indeed relevant for deducing E.

4.3 Logical Differences between Premise-Relevance and Conclusion-Relevance. There seem be formal parallels between irrelevant conclusions and irrelevant premises, via the law of contraposition. For instance, the irrelevant conclusion-disjunct in $p \mathcal{L} p \lor q$ turns into an irrelevant premise-conjunct $\neg p \land \neg q \mathcal{L} \neg p$ by contraposition. Moreover, if a premise of a premise-conjunct is superfluous in a deduction, then it is replaceable by any other formula salva validitate. Does this mean that we can obtain a satisfying definition of premise-relevance simply by applying our criterion for relevant conclusions to premises? - in the following way: $\Gamma \mathcal{L} A$ has relevant premises iff there exists no non-empty $\mu \subseteq \text{Nat}$ such that $\pi^\mu \Gamma \mathcal{L} A$ for every $\pi$. The answer is: not at all; this criterion would be much too strong. For, it would rule out modus ponens $p, p \rightarrow q \mathcal{L} q$ as premise-irrelevant, as well as every deduction $\Delta \mathcal{L} A$ with a premise set containing predicates which are not contained in the conclusion and are thus replaceable salva validitate (by uniform substitution). But the premises of relevant deductions are generally allowed to contain more information and in particular more predicates than the conclusion. It is the function of a deduction in an applied context to draw information from a set of accepted premises which is not immediately seen from the individual premises. This requires connecting the information contained in the single premises by means of a certain proof. Of course the premises may have contain information and more concepts than the conclusion. All that is required for their relevance is that each single premise is indeed necessary for deriving the conclusion.

So the basic idea of relevant premises is not that of essential predicates, explicated by salva validitate replacements applied to multiple occurrences, but the more simple idea of relevant premises. But as explained in section 4.2, this idea must be applied also to premise-conjuncts. So not only $p,q \mathcal{L} p$ but also $p \land q \mathcal{L} p$ has to be considered as irrelevant. Here the difficulty arises that conjuncts can also be hidden by various logically equivalent transformations. For instance if $p \land q \mathcal{L} p$, then also $\neg (\neg p \lor \neg q) \mathcal{L} p$ should be considered as premise-irrelevant, and if $p \land (q \lor r) \mathcal{L} p$, then
also \((p \land q) \lor (p \land r) \land p\). One suggestion for overcoming this problem would be to perform a logically equivalent 'splitting' of the premise set into its 'minimal conjuncts'. However this idea does not work because it may happen that a quantified sentence contains irrelevant conjuncts in its matrix without being itself splittable into conjuncts. For instance, since \(p \land q \land p\) is premise-irrelevant, \(\exists x(Fx \land Gx) \land p\) should also be counted as premise-irrelevant; but \(\exists x(Fx \land Gx)\) is not splittable into two conjuncts. A more complicated solution would be to apply the relevance idea to the conjuncts of the prenex conjunctive normal form \(\text{PKNF}(A)\) of a sentence \(A\) (this a a prenex NF with its matrix in conjunctive NF). But here a simplification is possible by the idea of salva validitate replacements, but now applied only to single occurrences, because the following theorem is provable by simple means: A premise contains a deductively superfluous conjunct in the matrix of its prenex conjunctive normal form iff it contains a single occurrence of a predicate which is replaceable salva validitate.

4.4 The Definition of Relevant Premises. We have seen that the idea of premise-irrelevance as superfluous premise-conjuncts is captured perfectly by the replacement criterion applied to single predicate occurrences, whereas the replacement criterion generally applied to multiple occurrences would be too strong.\(^{26}\) However there is one kind of irrelevance for which replacements applied to multiple occurrences are appropriate. This is the contrapositional counterpart of logically true conjuncts in the conclusion, namely redundant inconsistent disjuncts in the premises, and their 'variants' like:

\[(33)\] 

(i) \(p \lor (q \land \neg q), p \rightarrow r \models r\)  
(ii) \(p \lor q, p \lor \neg q, p \rightarrow r \models r\)

Here the premise set contains multiple predicates occurrences which are not only replaceable salva validitate, but also replaceable without changing its logical content. In this stronger case of replaceability without changing logical content, the application to multiple occurrences is adequate also for premise-relevance. To sum up, we obtain a definition of premise-relevance which combines two kinds of irrelevant premises: one based on replacements of single occurrences salva validitate, the other based on replacements of multiple occurrences without changing

---

\(^{26}\) In (1989, ch. 3.2) I suggested a different explication of premise-relevance based on a splitting of the premise set \(\Delta\) into 'minimal' conjuncts \(\text{Con}(\Delta)\) combined with the replacement criterion applied to multiple occurrences in these conjuncts. According to this criterion, a deduction \(\Delta \models A\) has relevant premises if no element \(P \in \text{Con}(\Delta)\) contains a predicate which is replaceable on some occurrences in \(P\) salva validitate. I think now that this explication is too strong. For instance, it makes the deductions \(\exists x(Fx \land (Fx \rightarrow Gx)) \models \exists xGx\) or \(\exists x(Fax \land \forall y(Fyx \rightarrow Gyx)) \models \exists xGax\) premise-irrelevant, because its premise is not conjunctively splittable and \(F\) is simultaneously replaceable salva validitate on both occurrences. But these deductions are just quantified variations of modus ponens and thus intuitively premise-relevant.
logical content. As in the case of conclusion-relevance, we treat predicate occurrences as sets $\mu \subseteq \text{Nat}$; single occurrences correspond to singletons \{n\} with $n \in \text{Nat}$. For applying the notion of occurrence-restricted predicate substitution $\pi^\mu$ to sets of sentences we assume the sentences in $\Gamma$ ordered from left to right in a fixed (say alphabetical) ordering; the predicate occurrences in $\Gamma$ are then ordered from left to right within this sequence by natural numbers.

(34) (Definition:) $\Gamma \vdash \pi A$ has relevant premises, in short $\Gamma \vdash^R \pi A$, iff

(34.1) there exists no $n \in \text{Nat}$ such that $\pi^{(n)} \Gamma \vdash \pi A$ holds for all $\pi$.

(34.2) there exists no non-empty $\mu \subseteq \text{Nat}$ such that $\Gamma \vdash^R_\pi \pi^\mu \Gamma$ holds for all $\pi$.

Examples of premise-relevant deductions in propositional logic are: $p \vdash \pi p$; $\neg p. (p \lor q) \vdash \pi q$; $p.p \rightarrow q \vdash \pi q$; $p.(p \rightarrow q) \vdash q$; $\neg (\neg p \lor (p \rightarrow q)) \vdash q$; $p \rightarrow q, \neg q \vdash \pi p$; $\neg (p \lor (p \rightarrow q)) \vdash q$; $p \rightarrow (p \lor q) \vdash \pi q$; $p \lor (q \lor r) \vdash (p \lor q) \lor r$; $p \lor (q \lor r) \vdash (p \lor q) \lor r$; $p \rightarrow q, q \rightarrow r \vdash \pi r$; $p \rightarrow q \land (q \rightarrow r) \vdash r$; $p \lor q \lor p \land q$; etc. Premise-relevant deductions in predicate logic are: $\forall x F(x, a \wedge b) \vdash_a x Fx$; $\forall x (F(x \rightarrow G(x)) \wedge F(x \rightarrow a \vdash \pi Ga); \forall x (F(x \rightarrow G(x)) \wedge \neg G(a) \vdash \pi \neg Fa); \forall x (G(x \rightarrow H(x)) \vdash \forall x (F(x \rightarrow H(x)); \forall x (F(x \rightarrow G(x)), \exists x F(x \rightarrow x \land \neg \exists x (F(x \rightarrow G(x)) \land \exists x Fx$; etc. Premise-irrelevant deductions in propositional logic are (underlined =replaceable): $p \rightarrow \neg p \vdash \pi p$; $p \land \neg p \vdash \pi q$; $p \lor \neg p \vdash \pi q$; $p \vdash \pi p$; $p \land q \vdash \pi p$; $p \lor q \rightarrow r \vdash \pi p \lor q \rightarrow r$; $p \lor q \rightarrow p \lor q$; $p \lor q \rightarrow p \lor q$; $p \lor q \rightarrow p \lor q$; $p \lor q \rightarrow p \lor q$; etc. Premise-irrelevant deductions in predicate logic are: $\forall x (F(x \wedge G(x)) \vdash \forall x Fx ; \exists x (F(x \wedge G(x)) \rightarrow Hx), Fa \vdash \pi Ga; \forall x (F(x \rightarrow \exists y (G(x \wedge H(x)))) \vdash \pi \exists x (F(x \wedge H(x)) \vdash \forall x Fx$; etc.

5. Applications of the Approach - An Overview

5.1. Applications in Philosophy of Science

5.1.1 Deductive Theory Confirmation and Deductive Explanation. These are the only cases where premise-relevance is also needed. All other applications listed in the following require only conclusion-relevance.

In accordance with sections 2.3 and 4.2, a proper notion of deductive theory confirmation is obtained by adding to explication (3) the requirement that the deduction $\Gamma \vdash \pi S$ must be premise-relevant as well as conclusion-relevant.

The application of deductive relevance conditions to explanations is more subtle because of a lot of other requirements for explanations which have to be combined with relevance in the right way. For details of the application of premise-relevance and conclusion-relevance to explanations, cf. Schurz (1983a, 1988 and 1989).

5.1.2. Verisimilitude. As explained in section 2.4, Popper's idea of verisimilitude can be explicated in a paradox-free and functioning way, if it is based on the set of
relevant consequence-elements of a theory. With the help of $\mathfrak{D}^{\mathcal{L}}_T$ they are defined as follows:

(35) (Definition:) $A$ is a relevant consequence element of $T$ [where $T \subseteq \mathcal{L}$, $A \in \mathcal{L}$] iff $T \mathfrak{D}^{\mathcal{L}}_T A$ and there exists no $B, C \in \mathcal{L}$ such that $T \vdash A \leftrightarrow (B \land C)$, $T \mathfrak{D}^{\mathcal{L}}_T (B \land C)$ and both $B$ and $C$ are shorter than $A$.

It is provable that every theory $T$ has a non-empty set of relevant consequence-elements, denoted by $(T)_r$, which is logically equivalent to $T$ provided the conjecture (28) of section 3.2 holds, and that the elements of $(T)_r$ have further nice properties, e.g. they don't contain redundancies of the form $A \lor A$ or $\neg \neg A$. The new definition of verisimilitude is now obtained in analogy to Popper's definition (6) with the help of the true and the false part of $(T)_r$, which we denote as $(T)_r^t := (T)_r \cap T$ and $(T)_r^f := (T)_r \cap \mathcal{F}$ ($T$ for the set of all true and $\mathcal{F}$ for the set of all false sentences). The only difference is that we must now replace the inclusion relation $\subseteq$ by the deducibility relation $\mathfrak{D}^{\mathcal{L}}_T$, and $\models$ by $\models^{\mathcal{L}}$ because the set $(T)_r$ is of course not deductively closed, whence $\models^{\mathcal{L}}$ does not coincide with $\subseteq$ as in Popper's definition based on the classical consequence class $\text{Cn}(T)$.

(36) (Definition:) $T_1 >_T T_2$ iff

either [case 1]: $(T_1)_r^t \models^{\mathcal{L}} (T_2)_r^t$ and $(T_2)_r^f \models^{\mathcal{L}} (T_1)_r^f$,

or [case 2]: $(T_1)_r^t \models^{\mathcal{L}} (T_2)_r^t$ and $(T_2)_r^f \models^{\mathcal{L}} (T_1)_r^f$.

A further important simplification is possible because it is provable that the sets $(T)_r^t$ and $(T)_r^f$ may be replaced by their so-called 'maximal reductions' without changing the verisimilitude. These 'maximal reductions' are any subsets $X \subseteq (T)_r^t$ and $Y \subseteq (T)_r^f$ which are a basis of $(T)_r^t$ and $(T)_r^f$, respectively, in Tarski's sense [i.e. $X$ but no proper subset of $X$ is logically equivalent with $(T)_r^t$, the same with $Y$ and $(T)_r^f$]. In most cases of theories, these maximal reductions have only a small number of elements which can be found and listed straightforwardly. Therefore definition (36) of verisimilitude is indeed practically feasible and applicable to non-trivial cases.

27 See Schurz/Weingartner (1987) and Schurz (1989, p. 145, prop. 14). For the purpose of an easy notion of verisimilitude we assume here that $\mathcal{L}$ contains the verum $\top$ and the falsum $\bot$ as undefined constants. By doing this, a logical truth contains $\top$ and an inconsistent theory contains $\bot$ as the only relevant consequence element.

28 Note that although $(T)_r$ is logically equivalent to $T$ provided (28) holds, the sets $(T)_r^t$ and $(T)_r^f$ are of course not logically equivalent to $(T)_t$ and $(T)_f$, respectively (otherwise our definition could not solve the Tichý-Miller paradox).

of scientific theories, as demonstrated in Schurz/Weingartner (1987) by means of several examples. In the case of quantitative theories, definition (36) has to be combined with a numerical approximation measure. By replacing the set of all true sentences by a given body of evidence, the notion of theory-progress can be defined in a similar way (Schurz 1987).

5.1.3 Knowledge Unification and Scientific Understanding. The understanding of the world provided by scientific theories consists in unifying the manifolds of empirical phenomena by a few basic laws and principles. This view, put forward by several scientists and philosophers of science, seems to be one of the most promising approaches to scientific understanding and explanation at present. Among other advantages, it gives an account in what explanation and understanding go beyond mere prediction or confirmation, without involving a relapse into unscientific traditional notions like explanation as "deduction from 'ultimate' principles", etc. However, the attempts to explicate the notion of unification more precisely have lead to serious difficulties. Let us call the set of sentences describing phenomena and principles a knowledge system $K$. Intuitively, $K$ is the more unified, the more sentences in it are inferable from the less ones. In this way, the knowledge system $K_N$ of Newtonian physics was very unified because a lot of empirical phenomena, say $P_1, ..., P_n$, were derivable from the Newtonian theory $T_N$. But the obstacle is an explanation paradox which was already raised by Hempel and Oppenheim (cf. Hempel 1965, p. 275, fn. 33) and had beset the approaches of Friedman (1974) and Kitcher (1981) towards unification - the paradox of conjunction: There exists also a trivial sentence, from which all the phenomena $P_1, ..., P_n$ are inferable, namely the conjunction $P_1 \land ... \land P_n$. This is of course not a 'real' unification. How can the 'real' unification of the phenomena $P_1, ..., P_n$ by $T_N$ be logically distinguished from the 'spurious' unification by $P_1 \land ... \land P_n$? Given our method of relevant consequence elements, the answer is at hand: we have to represent the knowledge system $K$ by the set of its relevant consequence elements $(K)_r$, which does not contain superfluous conjunctions like $P_1 \land ... \land P_n$ or any other irrelevant or redundant elements. If the above idea of unification is applied to $(K)_r$, a paradox free notion of the degree of unification of a knowledge system can be defined and fruitfully applied, as Schurz (1988) and Schurz/Lambert (1990) try to show.

5.1.4 The Empirical Significance of Theoretical Terms. It was an epoch-making discovery by Rudolf Carnap in (1936/37) that some fundamental scientific terms, like mass or force in physics, are not definable by empirical terms. Carnap called them theoretical terms. But how distinguish then, within the class of empirically undefinable terms, between these scientifically useful theoretical terms and

---

Friedman's definition in (1974) was refuted by Kitcher (1976) on various grounds, some of them connected with the conjunction paradox. For Kitcher's approach in (1981) the paradox of conjunction is a difficult problem.
unscientific 'metaphysical' terms? Given that the distinctive feature of scientific theories is that they have empirical consequences, Carnap (1956) suggested the following obvious idea: theoretical terms of science are empirically significant, i.e. they play an essential role in the derivation of empirical consequences, whereas metaphysical terms don't play such a role. Again, the problem was the correct explication. Carnap's suggestion was, in a simplified version, the following. Let $T \subseteq \mathcal{L}$ be a theory and $t$ a term in the language of $T$. Then: $t$ is empirically significant within $T$ iff $t$ occurs in a sentence $A \in T$ such that some empirical (i.e., observational) sentence $E$ follows from $T$ but not from $T \cup \{A\}$ alone. As is well known, Carnap's explication was beset by a lot of paradoxical consequences due to irrelevance and redundancy. A solution is possible by making use of the idea of unification based on relevant consequence elements, as I have tried to show in my (1989, pp.158ff). The basic idea is this: It is the main achievement of a scientific theory to unify its empirical consequences, and a theoretical term $t$ is empirically significant within $T$ iff the empirical unification provided by $T$ were not possible in a 'subtheory' of $T$ which does not contain $t$. To formalize this idea, let $T \subseteq \mathcal{L}$ be a scientific theory, $\mathcal{L}_e \subseteq \mathcal{L}$ be the purely empirical sublanguage of $\mathcal{L}$ (i.e. all sentences containing theoretical terms are in $\mathcal{L}$-$\mathcal{L}_e$), let $E(T)=\{A \in \mathcal{L}_e | \forall T,A\}$ be $T$'s empirical content. $(T)_r$ is the set of $T$'s relevant consequence elements. We call a minimal empirically equivalent axiomatization of $(T)_r$ any subset $\Delta \subseteq (T)_r$ such that $E(\Delta)=E(T)$ and there exists no $\Delta^* \subseteq (T)_r$ of smaller cardinality than $\Delta$ satisfying $E(\Delta^*)=E(T)$. We define: $t \in \mathcal{L}$ is empirically significant in $T$ iff there exists no minimal empirically equivalent axiomatization of $(T)_r$ which does not contain $t$.

5.1.5 Knowledge Representation. The three applications above suggest that the method of relevant consequence elements can be suggested as a general method of knowledge representation. To find an adequate way of knowledge representation plays a crucial role in a lot of areas, e.g., revisions of belief systems and counterfactual reasoning, which is also important for artificial intelligence. So there is a wide range of possible future applications.

31 Carnap's definition (1956, p. 51) is more complicated, but the basic problem can be explained away means of this simplified version (cf. also Stegmüller 1970, p. 347).

32 Cf. the discussion in Stegmüller (1970). To give an example: Assume $T$ is the conjunction $A \land B$ with $A$="God exists" and $B$="Grass is green". The term "God" would be empirically significant within $T$ according to the above definition. To avoid this inadequacy, we have to split $T$ into its conjuncts (Stegmüller 1970, p. 347). But which conjuncts? $A \land B$ is logically equivalent with the split-set $T_1=\{A,B\}$ as well as with $T_2=\{A, A \rightarrow B\}$. The term "God" would be empirically significant in the latter but not in the former (Stegmüller 1970, p. 348f). But clearly, $A \rightarrow B$ is an irrelevant consequence. A solution with the help of relevant consequence-elements lies at hand. Unfortunately, Stegmüller (1970, p. 361) concluded that Carnap's idea of empirical significance would have irreparably collapsed', and he went a very different way in his later structuralist approach to T-theoreticity. For a critique of the latter cf. Schurz (1990a).
5.2 Applications in Ethics.

5.2.1 Applications in Deontic Logic. Besides the Ross paradox, a lot of further deontic paradoxes rely on irrelevant conclusions. One example is the 'paradox of free choice', which relies on

(37) \( Pp \rightarrow P(p \lor q) \)

If it is permitted that John laughs, then it is permitted that he laughs or murders.

(cf. Stranzinger 1977, p. 151). Prior's paradox of 'derived obligation' (1954, p. 64) is reconstructed by Æqvist (1984, p. 640) in four versions:

(38) \( \neg p \rightarrow (p \rightarrow Oq) \)

If John is not elected, then if he is elected, he should murder.

(39) \( Op \rightarrow (q \rightarrow Op) \)

If John should be elected, then if he murders he should be elected.

(40) \( O \neg p \rightarrow O(p \rightarrow q) \)

If John should not be elected, then it should be the case that if he is elected, he murders.

(41) \( Op \rightarrow O(q \rightarrow p) \)

If John should be elected, then it should be the case that if he murders he is elected.

All the deductions (37)-(41) are conclusion-irrelevant (q is replaceable). So these paradoxes are avoided by the principle motivated in section 2.2 that in applied contexts only relevant consequences of norms are acceptable.

5.2.2 Applications to the Is-Ought Problem. The general Hume thesis GH, which we introduced in sections 2.5 and 3.1, turns out to be very important for the logical investigation of the is-ought problem. GH is provable for a broad range of logics - in fact a much broader class than that class in which SH, the special Hume thesis of section 2.5, holds. Various results about GH and SH are achieved in my (1989) and (1990b). The field of logics in which I have performed the investigation is that of bimodal predicate logics, containing the alethic necessity operator \( \Box \) and the deontic obligation operator O. The inclusion of \( \Box \) is necessary for a satisfactory philosophical investigation of the is-ought problem, because most naturalist theories of ethics, claiming 'ought' to be derivable from 'is', include in their descriptive premises necessity statements, for instance laws of nature, necessary properties of human life or social collectives, etc. The two most important results about GH, one syntactical and one semantical, are as follows.

Let \( \mathcal{L}(\Box, O) \) be the language of bimodal first-order predicate logic without identity symbol. Let \( \Pi(\Box, O) \) be the class of normal logics in \( \mathcal{L}(\Box, O) \) for constant domain, defined as follows. \( L \in \Pi(\Box, O) \) iff \( L \) contains the usual axiom schemes of classical predicate logic and that of the minimal normal modal logic for \( \Box \) and O, the Barcan
formulae for □ and O, and any additional set of propositional modal axiom schemes, denoted by θ_L, and is closed under substitution in Λ(□, O), modus ponens, ∀-generalization rules and □- and O-necessitation rule. Semantically, the logics in Π(□, O) are characterized by Kripke frames <W,R,S>, with W ≠ ∅ a non-empty set of possible worlds, R the alethic accessibility relation and S the deontic ideality relation. The models <W,R,S,D,v> based on them have D as a constant domain and v as a valuation function, with truth clauses etc. defined as usual. The additional propositional set θ_L expresses the special properties of □ and O, semantically characterized by the special properties of R and S. If a given logic L ∈ Π(□, O) is defined with the help of a given certain θ_L as above, we say that L representable by θ_L. L is called axiomatizable by θ_L iff θ_L is a decidable. Note that one and the same L may be representable by different additional sets θ_L.

Π(□, O) is the underlying class of the logical investigation. Observe that the set θ_L may contain any kind of axiom schemes, in particular also bridge principles between is and ought like □p → Op or even p → Op. Of course if is-ought bridge principles were included among the axiom schemes, Hume's thesis would trivially fail. However, to claim is-ought bridge principles as logical axioms would be a 'petitio principii' because their logical validity is exactly what is philosophically doubtful; in fact most philosophers would not regard them as analytical. So the question of primary interest is whether GH holds in all logics which don't contain such is-ought bridge principles among their axiom schemes. That this is indeed the case is the content of our first theorem. We call an axiom scheme X in Λ(□, O) an is-ought bridge principle iff it contains some scheme letter which has in X some occurrence within the scope of an O and some occurrence outside of the scope of any O. E.g., □A → OA, OA → □A, OA → A are is-ought bridge principles. On the other hand, □A, OB, □A → OB are not is-ought bridge principles - they express no relevant connection between is and ought, because they contain no variable occurring both in a descriptive and in a normative 'scope'. We say that L is representable without is-ought bridge principles iff L is representable with a set of additional axiom schemes θ_L which contains no is-ought bridge principle. The theorem then states:

(42) (Theorem:) For all L ∈ Π(□, O): GH holds in L iff L is axiomatizable without is-ought bridge principles. (Proof: Schurz 1989, ch. II.5.3; 1990b).

The second theorem gives a semantic characterization of the logics L ∈ Π(□, O) in which GH holds. Provided they are frame-characterizable at all, these logics are characterized by so-called completely is-ought separated frames. This is the semantic counterpart of the notion of a completely ought-irrelevant deduction and is defined as follows. <W,R,S> is completely is-ought separated iff there exists a partition of W into W_1 and W_2 [i.e. W_1 ≠ ∅, W_2 ≠ ∅, W_1 ∩ W_2 = ∅ , W_1 ∪ W_2 = W] such that:
all ideal worlds, i.e. those worlds $\alpha \in W$ deontically reached by some worlds $[\exists \beta : S \alpha ]$, lie in $W_2$, and (2.) $R$ does not connect $W_1$ and $W_2$, i.e. $R \subseteq (W_1 \times W_1) \cup (W_2 \times W_2)$. The important fact about such a frame is that in any model based on it, the truth of purely alethic-descriptive propositions is determined by $\forall W_1$ alone and the truth of purely normative propositions is determined by $\forall W_2$ alone, so both can be varied independently from each other. Our theorem reads as follows:

(43) (Theorem:) For all $L \in \Pi(\Box, O)$ which are frame-characterizable: $GH$ holds in $L$ iff $L$ is characterized by the class of all is-ought separated frames for $L$.

Two astonishing results are these. First, $SH$ turns out to fail in many logics in $\Pi(\Box, O)$, in particular in all logics with an 'alethic' fragment which is not Halldén complete (cf. fn. 6). Second, as soon as identity is included, even $GH$ fails, due to the relevant is-ought inference $a=b \vdash O(a=b)$.

5.3 Application to Cognitive Psychology. Relevant deduction can be investigated not only from the logical but also from the psychological point of view, as an empirically measurable cognitive behaviour of persons. In the motivation of our theory it was pointed out that in all cases of applied arguments we intuitively draw only relevant conclusions; in other words, our intuitive logical thinking is a combination of validity plus relevance. This cognitive hypothesis implies the following empirical prediction: the intuitive concept of a valid logical inference of adults entails the relevance of the conclusion. More precisely, we expect that logically untrained adults, confronted with logical arguments which they have to judge as valid or not, will only consider valid and conclusion-relevant arguments as valid, but conclusion-irrelevant ones as invalid. In an empirical test which I performed with a sample of 30 logically untrained students this prediction was confirmed in a very high degree. Also our hypothesis of section 4, that as distinct from conclusion-irrelevance, premise-irrelevant argument are intuitively not considered as 'invalid', was strongly confirmed.

I confronted the test persons with four kinds of arguments: paradigm cases of valid relevant arguments - in short RVA's; paradigm cases of logically invalid 'arguments' - abbreviated IA's; paradigm cases of valid arguments with irrelevant conclusions - abbreviated ICA's; and paradigm cases of arguments with irrelevant premises - abbreviated IPA's. The RVA's were judged correctly (i.e. as 'valid') by most of the students: projected on an ordinal scale from 0 to 4 - 0 for a completely incorrect answer, 4 for a completely correct answer - the mean score of the students for RVA's was 3,34. The IA's were also judged correctly by the majority of the students, although they made more mistakes here, in accordance with well-known psychological results: the mean score here was 2,78. But the ICA's were judged by
most of the students incorrectly, i.e. as 'invalid' (although they are logically valid) - the mean score here was 0.7. A t-test comparing the results for the PVA's with the ICA's gave a difference between RVA's and IA's with a significance level of $p \leq 0.0005$. Finally, the IPA's were judged correctly by the majority of the students - the mean score was 2.87. A t-test comparison gave a difference between ICA's and IPA's with a significance level of $p \leq 0.0005$, whereas there was no significant difference between PVA's and IPA's at a significance level of $p \leq 0.05$. For a detailed presentation cf. Schurz (1989). Interesting questions which are subject to further empirical studies concern different kinds of relevance, interconnections with other 'failures', dependencies on other parameters like the interpretation of concepts, etc.

5.4. Applications to Artificial Intelligence (AI). As mentioned in section 3.2, the fact that $\mathfrak{L}^c_r$ is not a deduction relation in the standard sense does not mean that $\mathfrak{L}^c_r$ could not have standard properties if the underlying deduction relation is weaker than the classical $\mathfrak{L}_r$. We first introduce the necessary concepts for generalizing our theory to deduction relations of any kind.

To save space, let $\mu$ always range over nonempty sets of natural numbers (and, as above, $\pi$ over predicate-substitution functions). We call a deduction relation $\vdash$ internally relevant (w.r.t. the conclusion) iff $\Delta \vdash A$ implies $\exists \mu \forall \pi: \Delta \vdash \pi A$. As a further notion, we call a deductive system $\vdash$ (c $\mathfrak{L}_r$) classically relevant iff $\Delta \vdash A$ implies $\exists \mu \forall \pi: \Delta \vdash \pi A$. The more important notion is that of internal relevance. For every deduction relation $\vdash$ we define its internally relevant restriction $\vdash^c_r$ by $\Delta \vdash^c_r A$ iff $\Delta \vdash A$ and $\exists \mu \forall \pi: \Delta \vdash \pi \mu A$. That a deductive system $\vdash$ is internally relevant means now exactly that $\vdash = \vdash^c_r$. As we know from section 3.2, $\mathfrak{L}_r$ is neither classically nor internally relevant. Of course $\mathfrak{L}^c_r$ is classically relevant and thus also internally relevant. Our question can now be reformulated as follows: Do there exist internally relevant deduction relations $\vdash \subset \mathfrak{L}_r$ which are axiomatizable and satisfy standard properties? Weaker deductive systems play an important role in AI as techniques of theorem proving. These techniques should be as efficient in search as possible. So it is not unplausible to expect that they satisfy relevance requirements. In what follows we show that the deduction relation underlying the simple but nevertheless important inference mechanism of PROLOG is internally relevant. Similar consideration can certainly be extended to the resolution method in general.

PROLOG has a very simple language, denoted by $\mathcal{L}(\text{PRO})$. Its well-formed formulas are so-called Horn-clauses. They can be written into the form $\Delta A_1 \land \ldots \land A_n \rightarrow A_{n+1}$ Here $A_i$ denote atomic formulas of the language of predicate logic (including function symbols, individual variables and constants), and $n \geq 0$. For $n>0$ such a formula is called a 'rule', for $n=0$ it reduces to an atomic formula and is
called a 'fact'. All PROLOG formulas are to be read as universally generalized. The
deductive inference mechanism of PROLOG, denoted by \( \models_{\text{PRO}} \), can be formalized by
the following axiomatic system (A,B... denote formulas \( \in \mathcal{L}(\text{PRO}) \), i.e. Horn-clauses,
and \( \Gamma, \Delta \) sets of them; \( A_t \) denote atomic formulas; \( x_i \) variables, \( t_i \) terms).

**Axioms:** Every instance of:
(Substitution) \( A \models_{\text{PRO}} A[t_1/x_1, \ldots, t_n/x_n] \)
(Resolution) \( (A_1 \land \ldots \land A_n \rightarrow A_{n+1}), A_1, \ldots, A_n \models_{\text{PRO}} A_{n+1} \)
(Repetition) \( \Delta \models_{\text{PRO}} A_t \), for all \( A_t \in \Delta \) provided \( \Delta \) is finite.

**Rules:**
(Cut) If \( \Delta \models_{\text{PRO}} A_t \) and \( \{A_t\} \models_{\text{PRO}} B \), then \( \Delta \cup \Gamma \models_{\text{PRO}} B \)

Important properties of \( \models_{\text{PRO}} \), following from this axiomatization, are: If \( \Delta \models_{\text{PRO}} A \) then \( \Delta \) is finite and \( A \) is atomic (induction on the length of the proof); \( \models_{\text{PRO}} \) is monotonic
(from repetition and cut), \( \models_{\text{PRO}} \) is closed under \( \pi \)-substitution (but of course not closed
under substitution in general, because of the restriction of Horn-clauses). \( \models_{\text{PRO}} \) is
decidable. Without proof we claim that our axiomatization is correct, i.e. it holds that
\( \Delta \models_{\text{PRO}} A \) according to the above axiomatization iff a PROLOG program containing \( \Delta \)
in its knowledge base answers the question "?-A" with "yes" after a finite time.\(^{34}\)

Before proving that \( \models_{\text{PRO}} \) is internally relevant we note an important connection
between our definition of conclusion-relevance and the APW-criterion of section 3.3.
As we saw there, they do not coincide if applied to single deductions. But as general
properties of deduction relations closed under \( \pi \)-substitutions, these two criteria
indeed coincide.

\(^{(48)} \) (Theorem:) Let \( \models \) be a deduction relation which is closed under \( \pi \)-substitutions.
Then: \( \models \) is internally relevant iff \( \models \) satisfies the APW-criterion, i.e. \( \Delta \models A \) implies
\( \Re(A) \subseteq \Re(A) \).

**Proof:** \( \bar{U} \): Assume \( \models \) does not satisfy the APW-criterion, i.e. there exists \( \Delta \models A \) and
\( R \in \Re(A) \) such that \( R \notin \Re(\Delta) \). Let \( \pi \) be a substitution function replacing only \( R \) by a new
predicate of the same arity. Because \( \models \) is closed under \( \pi \)-substitutions we have
\( \Delta \models \pi A \) for every \( \pi \) (where \( \mu \) coincides here with all \( R \)-occurrences and thus can be

\(^{33} \) The PROLOG-notation deviates from that of predicate logic: all implication are written in the
inverse form \( \models \); "\( \models \)" is replaced by "\( \models \)"; \( \land \) by \( \models \); individual variables are written with
capital letters (or words starting with such), whereas individual constants, function symbols
and predicate letters are written with small letters (or words starting with such). For
example, the PROLOG-notation of \( Rxy \land Fx \rightarrow Gxy \) would be \( g(X,Y) \models r(X,Y), f(X) \). For a

\(^{34} \) Provided its clauses and rules are put in an appropriate ordering which avoids infinite loops.
More on the theoretical foundations of PROLOG programs, also called simply logic
programs, see Lloyd (1984).
omitted); so $\vdash$ is not internally relevant. 1: Assume $\vdash$ is not internally relevant, so there exists a $\Delta$ and A with $\exists \mu \forall \pi: \Delta \vdash \pi^\mu A$. Let $\pi$ be a substitution function replacing the predicates in A by ones which are not contained in $\Delta$.\(^{35}\) We obtain a deduction violating the APW-criterion. Q.E.D.

$\vdash_{\text{PRO}}$ is closed under $\pi$-substitution. So for proving its internal relevance it suffices to prove that:

(49) (Theorem:) $\Delta \vdash_{\text{PRO}} A$ implies $\mathcal{R}(A) \subseteq \mathcal{R}(\Delta)$.

Proof: Induction on the length of the proof. For $\Delta \vdash_{\text{PRO}} A$ an axiom, (Substitution), (Resolution-Rule) or (Repetition), the theorem holds obviously. Assume $\Delta \cup \Gamma \vdash_{\text{PRO}} A$ is proved from $\Delta \vdash_{\text{PRO}} B$, $\{B\} \cup \Gamma \vdash_{\text{PRO}} A$ by cut rule. By induction hypothesis, $\mathcal{R}(A) \subseteq \mathcal{R}(\{B\} \cup \Gamma)$ and $\mathcal{R}(B) \subseteq \mathcal{R}(\Delta)$, so $\mathcal{R}(A) \subseteq \mathcal{R}(\Delta \cup \Gamma)$. Q.E.D.

(Corollary:) $\vdash_{\text{PRO}}$ is internally relevant. Proof: From theorems (48) and (49).

$\vdash_{\text{PRO}}$ is not classically relevant because it might be that $\Delta \vdash_{\text{PRO}} A$, where $\Delta$ is inconsistent. But the subset of $\vdash_{\text{PRO}}$-deductions with classically consistent premises is also classically relevant, as is easy to show.

Generalizations of these investigations for other proof-techniques in AI, like the general resolution method, are possible. Besides investigating existing proof mechanisms in AI on their internal (and classical) relevance, there are further interesting applications of the theory of relevant deduction to AI. For instance, by introducing defined operators for classical negation and disjunction in PROLOG it is then possible to implement a program which finds out whether a classically valid deduction is relevant or not.\(^{36}\)

References


\(^{35}\) This $\pi$ always exists, even for infinite $\Delta$, because of what was said in fn. 21.

\(^{36}\) Such a program can't be complete for first-order predicate logic. For propositional logic it was implemented by Don Fallis at the University of California at Irvine.


