PART I

BACKGROUNDS

In the three following chapters work in philosophy, artificial intelligence, and logic that is relevant for the present investigations is reviewed. In the first chapter, The philosophy of induction, *I discuss the philosophical backgrounds of this study. I concentrate on philosophers who have studied induction from a logical perspective, most notably Peirce and Hempel. The main conclusion drawn from this chapter is that the dichotomy between explanatory and confirmatory induction proposed and defended in this thesis is already present, albeit implicit, in the work of Peirce and Hempel. The next chapter is called* Approaches to computational induction. *It draws upon work in machine learning (a subfield of artificial intelligence) on inductively learning concepts, logic programs, and logical theories from examples. I indicate how the latter two problems can be reformulated as problems of explanatory and confirmatory induction, respectively. The third chapter,* The analysis of non-deductive reasoning, *is mainly devoted to one article by Kraus, Lehmann and Magidor that provided the main inspiration for my work. In that article the authors set out to "study general patterns of nonmonotonic reasoning"; in this thesis I have set out to do the same for induction.*

CHAPTER 2

THE PHILOSOPHY OF INDUCTION

— in which the philosophical backgrounds of this study are discussed, leading to the conclusion that the dichotomy between explanatory and confirmatory induction proposed and defended in this thesis is already present, albeit implicitly, in the work of Peirce and Hempel —

THE TIME THAT philosophy was considered the Mother of Science is long gone, and disciplines, such as logic, that were once among the central concerns of philosophers, have gained a separate identity. However, foundational studies like the present one cannot afford to ignore the reflections of philosophers of all time. I therefore start my investigations by considering the philosophy of induction.

I do not claim to give an overview of this subject that can be called complete in any sense. While hard in general, this task is even more formidable in the case of induction, which has always been a philosophical battlefield. Instead, I will concentrate on those philosophers whose work provides some of the foundations upon which this thesis will be built: Charles Sanders Peirce and Carl G. Hempel. Before considering their work in more detail, however, I will give a brief historical overview of parts of the battlefield. The second half of the chapter is devoted to a discussion of various issues that are especially significant in the context of this thesis.

§4. THE 'PROBLEM OF INDUCTION'

Induction has been studied by philosophers of all times, but there appears to be little agreement about a general theory of inductive reasoning, or even about what the questions are which such a theory is designed to answer. As a result, there exists a lot of confusion and disagreement about what the so-called 'Problem of Induction' actually is (which is why I will surround the term by quotes). In this section, I will give a — necessarily brief and subjective — historical overview of the aspects most relevant to the present discussion.

Induction was recognised already by Aristotle as a distinguished mode of reasoning in his *Analytica Posteriora*. He gives a syllogistic account of what is nowadays called complete or mathematical induction, a proof strategy in mathematics used to prove theorems involving the infinite set of all natural numbers. However, complete induction is actually a deductively valid form of reasoning, and is therefore not relevant for the present discussion.

Scientific method, as we know it today, started to take shape during the Renaissance. Francis Bacon was one of the first philosophers to stress the importance of induction as a method of scientific theory formation, in his *Novum Organum* (1620). An inductive scientific methodology was further worked out by John Stuart Mill in his *System of Logic* (1843). Mill identifies four ‘Methods of Experimental Enquiry’, and he formulates five ‘Canons’ or principles underlying and justifying those methods. After the methods and canons have been formulated, Mill gives a particularly pregnant formulation of what remains to be done:

> ‘Why is a single instance, in some cases, sufficient for a complete induction, while in others, myriads of concurring instances, without a single exception known or presumed, go such a very little way towards establishing an universal proposition? Whoever can answer this question knows more of the philosophy of logic than the wisest of the ancients, and has solved the problem of induction.’
> (Mill, 1843, Book III, Chapter VIII, p.314)

A century earlier, ‘the greatest of modern philosophers’[11] had shone his light upon this problem. It was generally conceived that the scientific method was concerned with explicating the necessary causal relations that exist between physical objects in the world. Likewise, induction was seen as reasoning from observations to logically necessary causal laws. However, in his *Treatise on Human Nature* (1739), the Scottish philosopher David Hume concluded that the idea of a causal law is something that exists solely in the mind, a mental image provoked by observing a number of occasions in which events of a certain kind are consistently followed by events of another kind. In other words, there is no logical necessity inherent to causal laws, and any attempt to formulate a firm logical basis for induction is doomed to fail.

Hume’s scepsis, directed at metaphysical notions such as substance and causality, did not lead him to a radical solipsism. He was keenly aware of the practicality of induction as a systematic method to derive new knowledge, and his argument should be seen as refuting the logical status of necessity commonly ascribed, in his times, to inductive conclusions, and not, as it is sometimes perceived, as denouncing the inductive method itself. Hume was the first to formulate the logical status of inductive conclusions in terms of *probability* rather than necessity. However, this inductive probability does not refer to states of affairs in the world in terms of relative frequencies, as stochastic probability does, but it expresses the degree to which we are willing to accept the hypothesis on the basis of available evidence — in other words, it is *subjective*. The concept of inductive probability, which has been worked out in considerable detail by twentieth-century philosophers such as Keynes and Carnap, will not be our major concern in this thesis.[12]

The American philosopher Charles Sanders Peirce noted that the view of induction as assessing inductive probabilities presupposes the inductive hypothesis as given, and leaves the question as to where these hypotheses come from unanswered. Peirce introduced

---

[11] According to (Goodman, 1954).

[12] For a discussion of the philosophical aspects of inductive probabilities, see the chapter with the same name as the present one in (Keuzenkamp, 1994).

the term *abduction* to denote the process of forming explanatory hypotheses, and claimed that it is a process of logical inference — not because humans actually form inductive hypotheses by means of some inductive proof procedure (on the contrary, the hypothesis 'comes to us like a flash', and is 'an act of insight'; Peirce compares it to knowledge obtained through direct perception), but rather because it has 'a perfectly definite logical form'. We will find ample opportunity to discuss this logical form in this thesis.

The ideas of Peirce seem to have gone unnoticed for a long time, especially because at the time of his writings (the turn of the nineteenth century) the formalisation of logic was still insufficient for his needs. The term 'abduction' for the process of forming explanatory hypotheses has not been widely accepted in philosophy (it has been introduced recently in computer science, although there it usually receives a much more restricted interpretation). On the other hand, the inclination towards taking logic as the basis of scientific reasoning was yet to reach its peak, in the writings of the so-called Wiener Kreis (around 1930).

The Wiener Kreis, which included the philosophers and logicians Rudolf Carnap and Carl G. Hempel, advocated a doctrine known as logical positivism, a radical form of empiricism and anti-metaphysicism. Being heavily influenced by the early Wittgenstein and his *Tractatus Logico-Philosophicus*, they believed that most of the traditional philosophical problems are in fact pseudo-problems, arising from imprecise or incorrect use of language. The hallmark of real knowledge, they argued, is *verifiability*, and the scientific method is the only way to achieve such knowledge. A different position was taken by the Austrian philosopher Karl R. Popper, who proposed another criterion to distinguish between scientific and non-scientific knowledge[13]: the criterion of *falsifiability*. Popper drew attention, like Hume had done before him, to the fact that scientific hypotheses can never be conclusively verified — however, they **can** be conclusively falsified.

Although many of its views are now considered too radical, the Wiener Kreis has had an enormous influence on philosophy of science, in that many of the main questions have been reformulated into logical terms. Clearly, Popper has been much more influential than any member of the Wiener Kreis when it comes to methodological issues in science; but it must be added that his famous slogan 'conjectures and refutations' is worked out rather sketchy as far as the first conjunct is concerned[14]. Members of the Wiener Kreis, most notably Hempel and Carnap, have investigated, in considerable detail, the question what it means to say that certain conjectures are suggested or *confirmed* by the evidence, while others are not. Hempel's analysis is qualitative, while Carnap's is quantitative. Although Carnap seems to think that a qualitative analysis is a rather thin extract of the real thing, it is Hempel's analysis that provides a foundation upon which much of my subsequent investigations will be built.

---

[13]Unlike the Wiener Kreis, Popper did not consider non-scientific knowledge to be nonsensical.

[14]'Induction, i.e. inference based on many observations, is a myth. It is neither a psychological fact, nor a fact of ordinary life, nor one of scientific procedure. The actual procedure of science is to operate with conjectures: to jump to conclusions' (Popper, 1963, p.53). However, the term 'conjecture' is not described or defined any further — it does not even occur in the index of *Conjectures and Refutations*.

After having identified the intellectual treasurers to whom I can be held tributary, I will now proceed to present their work in somewhat more detail. My intention is to be as faithful as possible to their original writings, leaving my own subjective interpretations and criticisms for the subsequent discussion (§7).

§5. THE LOGIC OF ABDUCTION

In a series of lectures on Pragmatism delivered in 1903, Peirce distinguishes three types of reasoning: deduction, induction, and abduction. Induction 'consists in starting from a theory, deducing from it predictions of phenomena, and observing those phenomena in order to see *how nearly* they agree with the theory'. Furthermore,

> 'The justification for believing that an experiental theory which has been subjected to a number of experimental tests will be in the near future sustained about as well by further such tests as it has hitherto been, is that by steadily pursuing that method we must in the long run find out how the matter really stands.' (CP 5.170)[15]

Note that Peirce claims that 'induction consists in starting from a theory' — that is, it aims at assessing the plausibility of a **given** theory, rather than **constructing** that theory from observations.

However, inductive hypotheses do not come out of the blue, and this is where abduction comes into play:

> 'Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea; for induction does nothing but determine a value, and deduction merely evolves the necessary consequences of a pure hypothesis.
>
> Deduction proves that something *must* be; Induction shows that something *actually is* operative; Abduction merely suggests that something *may be*.
>
> Its only justification is that from its suggestion deduction can draw a prediction which can be tested by induction, and that, if we are ever to learn anything or to understand phenomena at all, it must be by abduction that this is to be brought about.
>
> No reason whatsoever can be given for it, as far as I can discover; and it needs no reason, since it merely offers suggestions.' (CP 5.171)

In other words, *abduction is the process of conjecturing inductive hypotheses*, constrained by the requirement that they should comply with the available observations. Abduction represents the purely logical part of inductive reasoning.

Peirce proceeds by defining the logical form of abduction. 'It must be remembered', he writes, 'that abduction, although it is very little hampered by logical rules, nevertheless is

---

[15]References with the prefix 'CP' refer to Hartshorne, Weiss & Burks (eds.), *Collected Papers of Charles Sanders Peirce* (1931-58).

logical inference, asserting its conclusion only problematically or conjecturally, it is true, but nevertheless having a perfectly definite logical form.’ Peirce then defines this logical form, as follows.

> ‘Long before I first classed abduction as an inference it was recognized by logicians that the operation of adopting an explanatory hypothesis — which is just what abduction is — was subject to certain conditions. Namely, the hypothesis cannot be admitted, even as a hypothesis, unless it be supposed that it would account for the facts or some of them. The form of inference, therefore, is this:
>
> The surprising fact, C, is observed;
>
> But if A were true, C would be a matter of course,
>
> Hence, there is reason to suspect that A is true.
>
> Thus, A cannot be abductively inferred, or if you prefer the expression, cannot be abductively conjectured until its entire content is already present in the premiss, “If A were true, C would be a matter of course.” ’
> (CP 5.188)

In short, the view of induction that Peirce offers in these marvellously lucid citations is this. Inductive reasoning[16] consists of two steps: (*i*) formulating a conjecture, and (*ii*) evaluating the conjecture. Both steps take the available evidence into account, but in quite different ways and with different goals. The first step, abduction, requires that the conjectured hypothesis *explains* the observations; having a definite logical form, it represents a form of inference. The second step, induction, evaluates how well predictions offered by the hypothesis agree with reality; it is not inference, but assigns a numerical value to a hypothesis.

We can paraphrase this model, to which I will refer as *Peirce’s model*, as follows: ‘the goal of inductive reasoning is to find an explanation for some observations, the predictions of which comply with other observations’. Although the notion of an explanation is not explicitly defined, we can abduce from the last citation above that *A* explains *C* whenever ‘if A were true, C would be a matter of course’. I will argue in §7 that, although there is room for various interpretations here, this should be interpreted as ‘*A* deductively entails *C*’.

Given this paraphrase, we can summarise the main claims defended in this thesis as follows:

- (*i*) the notion of explanation is actually a **parameter** in Peirce’s model, that can be instantiated to deductive entailment but also to, for instance, plausible entailment;
- (*ii*) at a higher level of abstraction, the **goal** of inductive reasoning is also a parameter, that can be instantiated to ‘finding a hypothesis that explains some observations’, but also to ‘finding a hypothesis that is confirmed by some observations’.

Although not the main motivation, the way the second claim is substantiated in this

[16] I will use the term ‘inductive reasoning’ for the combined process of abductive or conjectural reasoning and inductive validation.

thesis has been influenced by work of Hempel on the concept of confirmation, which I will discuss now.

§6. THE LOGIC OF CONFIRMATION

How shall we decide whether certain observational evidence confirms a scientific theory? Several options come to mind, one being to define a binary relation *Cf* between two logical formulas, such that *Cf*(*E*,*H*) formalises the proposition '*E* confirms *H*'. Another option is to define instead a function assigning a number between 0 and 1 to *H* given *E*. The first option, leading to a qualitative concept of confirmation, has been worked out by Hempel in a paper published in 1943; a less technical account, which concentrates on the philosophical issues involved, can be found in (Hempel, 1945). The second option, leading to a quantitative concept of degree of confirmation, has been elaborated by Carnap. I will now concentrate on the qualitative concept, and discuss its relation with the quantitative concept in §7.

Adequacy conditions

Before considering possible definitions of this relation of confirmation, Hempel presents a number of logical conditions of adequacy, to be satisfied by any such definition. These logical conditions, which are represented below, are not independent :

> 'The logical requirements are stated in three groups. In each group, the fulfillment of the first condition entails that of all others. Those other conditions are mentioned for two reasons; first, because most of them represent important characteristics which would generally be sought in an adequately defined concept of confirmation; and secondly, because some apparently reasonable alternative definitions which we shall examine, turn out to satisfy some of those weaker conditions, but not the strongest of each group. Confrontation with the different requirements explicitly stated will thus provide a yardstick for the appraisal of what might be termed the degree of adequacy of a proposed definition of confirmation.'
> (Hempel, 1943, p.127)

There are some differences in presentation of the adequacy conditions in the two papers — the exposition below can be found in (1945, pp.103–106)[17]. Note that an observation report is a formula without variables. Furthermore, a formula is compatible with a class[18] of formulas if their conjunction is consistent.

(H1) *Entailment condition*: any sentence which is entailed by an observation report is confirmed by it.

(H1.1) Any observation report is confirmed by itself[19].

[17] Condition (H4) is added on p.110, note 1.

[18] Hempel uses the word 'class'; I will consider it synonymous with 'set'.

[19] This condition was not explicitly mentioned by Hempel, but is added to facilitate later discussions.

(H2) *Consequence condition*: if an observation report confirms every one of a class *K* of sentences, then it also confirms any sentence which is a logical consequence of *K*.

(H2.1) *Special consequence condition*: if an observation report confirms a hypothesis *H*, then it also confirms every consequence of *H*.

(H2.2) *Equivalence condition*: if an observation report confirms a hypothesis *H*, then it also confirms every hypothesis which is logically equivalent with *H*.

(H2.3) *Conjunction condition*: if an observation report confirms each of two hypotheses, then it also confirms their conjunction[20].

(H3) *Consistency condition*: every logically consistent observation report is logically compatible with the class of all the hypotheses which it confirms.

(H3.1) Unless an observation report is self-contradictory[21], it does not confirm any hypothesis with which it is not logically compatible.

(H3.2) Unless an observation report is self-contradictory, it does not confirm any hypotheses which contradict each other.

(H4) *Equivalence condition for observations*: if an observation report *B* confirms a hypothesis *H*, then any observation report logically equivalent with *B* also confirms *H*.

These adequacy conditions are intended to formalise certain intuitions about the concept of confirmation. Whether we find them acceptable or not depends on whether our intuitions about that concept are the same as Hempel's. Let us, therefore, investigate the motivation offered by Hempel for the main adequacy conditions.

The entailment condition is motivated by noting that 'entailment is a special case of confirmation' (1945, p.102). Elsewhere, Hempel remarks that entailment 'might be referred to as the special case of *conclusive* confirmation' (p.107). These are important remarks, because they offer some insight in what Hempel has in mind when he talks about confirmation. Furthermore, they will be a starting point for the logical formalisation of the concept of confirmation in chapters 6 and 7.

The consequence conditions (H2) and (H2.1) state that the relation of confirmation is closed under weakening of the hypothesis or set of hypotheses (H_1 is weaker than H_2 iff it is logically entailed by the latter). Hempel justifies this condition as follows (1945, p.103): 'an observation report which confirms certain hypotheses would invariably be qualified as confirming any consequence of those hypotheses. Indeed: any such

[20]This condition is considered on p.106.

[21]By the entailment condition, a contradictory observation report confirms every hypothesis. As Hempel notes, 'it is possible to exclude the possibility of contradictory observation reports altogether (…). There is, however, no important reason to do so.' (p.103, note 1). In chapters 6 and 7 I will choose to exclude them.

consequence is but an assertion of all or part of the combined content of the original hypotheses and has therefore to be regarded as confirmed by any evidence which confirms the original hypotheses.'

The consistency condition (H3) seems to be more problematic. Hempel does not really give a motivation for this condition, but immediately remarks that it 'will perhaps be felt to embody a too severe restriction'. He considers to possibility of dropping (H3) while retaining (H3.1) and (H3.2) (p.106). However, combination of (H3.1) with the conjunction condition (H2.3) implies (H3); consequently, if we drop (H3) while keeping (H3.1), we should drop (H2.3), and *a fortiori* (H2), as well. Hempel decides to keep them all — without, it must be added, a very clear justification.

The confirmation paradox

The adequacy conditions (H1–4) are mutually compatible, in the sense that their combination does not lead to counterintuitive results. This might be taken to indicate that each of them indeed formalises some aspect of a single intuitive notion. There is, however, a fifth rule, with considerable intuitive appeal, which is incompatible with (H1–4). This problem is known as the *confirmation paradox*. While there are various formulations of this paradox, we will follow here Hempel's original treatment; the reader is referred to (Hesse, 1974) for a more elaborate discussion.

Let H_1 and H_2 be two theories such that the latter includes the former, in the sense that everything entailed by H_1 is also entailed by H_2. Suppose E is confirming evidence for H_1; shouldn't we conclude that it confirms H_2 as well? To borrow an example of Hempel: 'Is it not true, for example, that those experimental findings which confirm Galileo's law, or Kepler's laws, are considered also as confirming Newton's law of gravitation?' (1945, p.104). This intuition is formalised by the following condition:

(H5) *Converse consequence condition*: if an observation report confirms a hypothesis H, then it also confirms every formula logically entailing H.

The problem is, however, that this rule is incompatible with consequence conditions (H2) and (H2.1). This can be seen as follows: in order to demonstrate that E confirms H for arbitrary E and H, we note that E confirms E by (H1.1), so by the converse consequence condition E confirms $E \wedge H$; but then E confirms H by (H2.1). Thus, we see that the combination of two intuitively acceptable conditions leads to a collapse of the system into triviality, a clearly paradoxical situation.

Hempel solves the problem on the formal level by dropping the converse consequence condition in favour of the consequence conditions. However, on the intuitive level the paradox remains, since Hempel does not provide a clear justification of his choice. I will argue at the end of the present chapter that the confirmation paradox can be fully dissolved by making a clear distinction between a confirmed hypothesis and an explanatory hypothesis. For instance, the converse consequence condition does make sense for hypotheses that are required to be explanatory (in the sense of Peirce) rather than confirmed. The duality between explanatory and confirmatory reasoning we encounter here for the first time will be a leading theme throughout the thesis.

The satisfaction criterion of confirmation

I will now proceed to explain Hempel's proposal for a definition of confirmation satisfying the above adequacy conditions, following the less technical exposition in (Hempel, 1945). The basic idea is simple and elegant:

> 'Consider the simple case of the hypothesis H: '$\forall x$: Raven(x)$\rightarrow$Black(x)', where 'Raven' and 'Black' are supposed to be terms of our observational vocabulary. Let B be an observation report to the effect that Raven(a)$\wedge$Black(a) $\wedge$ $\neg$Raven(c)$\wedge$Black(c) $\wedge$ $\neg$Raven(d)$\wedge\neg$Black(d). Then B may be said to confirm H in the following sense: There are three objects altogether mentioned in B, namely a, c, and d; and as far as these are concerned, B informs us that all those which are ravens (*i.e.* just the object a) are also black. In other words, from the information contained in B we can infer that the hypothesis H does hold true within the finite class of those objects which are mentioned in B.' (Hempel, 1945, p.108; with slight modifications regarding the logical symbols)

To formalise the notion of 'a hypothesis being true within a class of objects', Hempel introduces the concept of the *development of a hypothesis H for a finite set of individuals C*, which 'states what H would assert if there existed exclusively those objects which are elements of C' (p.109). The formal definition can be found in (Hempel, 1943, p.131); I will introduce the concept by means of a few examples. The development of the hypothesis

$$H_1 = \forall x\colon P(x) \vee Q(x)$$

for the set $\{a, b\}$ is $(P(a)\vee Q(a))\wedge(P(b)\vee Q(b))$; the development of

$$H_2 = \exists x\colon P(x)$$

for the same set is $P(a)\vee P(b)$; and the development of a formula without variables, such as

$$H_3 = P(c)\vee Q(c)$$

is that formula itself, regardless of the set of individuals. In perhaps more familiar terminology we might say that the development of H for a set C is the set of ground instances of H over the Herbrand universe C[22]. It should be noted that Hempel does not include function symbols in his language.

For reasons which will be exhibited shortly, the relation of confirmation is defined in terms of a narrower relation of direct confirmation:

> An observation report E *directly confirms* a hypothesis H if E entails the development of H for the class of those objects which are mentioned in E.

Here, an observation report is a formula without variables, such as $E = P(a)\wedge P(b)$. Thus, E directly confirms H_1 and H_2, but not H_3. However, the relation of direct confirmation

[22] Assuming that H is a set of formulas in clausal form (without existential quantifiers).

does not satisfy the equivalence condition (H2.2), since E does not directly confirm

$$H_4 = H_1 \wedge H_3 = (\forall x\colon P(x) \vee Q(x)) \wedge (P(c) \vee Q(c))$$

although this formula is logically equivalent with H_1. This problem is overcome in the following definition:

> An observation report E *confirms* a hypothesis H if H is entailed by a class of formulas each of which is directly confirmed by E.

So, $P(a) \wedge P(b)$ confirms each of H_1–H_4.

Hempel calls this definition the *satisfaction criterion of confirmation* 'because its basic idea consists in construing a hypothesis as confirmed by a given observation report if the hypothesis is satisfied in the finite class of those individuals which are mentioned in the report' (1945, pp.109–110). This idea suggests an alternative formalisation of this criterion, namely in terms of satisfaction by a specially constructed model. This will be considered in the next section.

Hempel rounds off his 1945 paper with the following considerations:

> 'A general definition of confirmation, couched in purely logical terms, was developed for scientific languages of a specified and relatively simple logical character. The logical model thus obtained appeared to be satisfactory in the sense of the formal and material standards of adequacy that had been set up previously.
> (…) Among the open questions which seem to deserve careful consideration, I should like to mention the exploration of concepts of confirmation which fail to satisfy the general consistency condition; the extension of the definition of confirmation to the case where even observation sentences containing quantifiers are permitted; and finally the development of a definition of confirmation for languages of a more complex logical structure than that incorporated in our model.
> (Hempel, 1945, pp.120–121)

I will return to these issues below as well.

§7. DISCUSSION

I will now relate the issues raised above to each other and to the rest of this thesis. Specifically, I will investigate the significance of Hempel's adequacy conditions for characterising the logical form of abduction as put forward by Peirce. I will then briefly discuss the confirmation paradox and what I consider as the cause of it. Furthermore, I will reformulate Hempel's satisfaction criterion of confirmation in the terminology of clausal logic. Finally, I will discuss the relation between a qualitative notion of confirmation as put forward by Hempel, and the quantitative concept of degree of confirmation studied by Carnap.

The logical form of abduction

Recall that Peirce defined the logical form of an abductive argument as follows:

> The surprising fact, *C*, is observed;
> But if *A* were true, *C* would be a matter of course,
> Hence, there is reason to suspect that *A* is true.

The problem with this definition is that it contains a few nonlogical phrases, *viz*. *C* being a 'surprising fact' and 'there is reason to suspect' that *A* is true. The first phrase can be made precise by introducing a background theory *T*, and stipulating that *T* should not logically entail *C*. Thus, if *C* is known to be true by virtue of *T*, it will not allow for any abductive explanation. On the other hand, the reader will probably agree that this represents a nonessential borderline case, not unlike the treatment of inconsistent theories in deductive logic: such a theory entails anything, yet one could make a case for not inferring anything from such a theory. For the moment, I will choose the option of allowing any abductive explanation if *C* is already known.

The phrase 'there is reason to suspect that *A* is true' requires some careful consideration. Clearly, Peirce couldn't have used the phrase '*A* is true' instead, since some of the possible abductive explanations of *C* will probably be false. However, there seems to be a mixture of syntactic and semantic issues here. Consider, as an illustration, the deductively valid scheme of *modus ponens*[23]:

> A
>
> $A \rightarrow B$
>
> $\therefore B$

This scheme can be paraphrased in several ways. One possibility is

> If *A* is true,
> and 'if *A* then *B*' is true,
> then *B* is true.

However, the validity of *modus ponens* does not depend on the truth of any of the propositions involved — the following argument is still deductively valid:

> If the moon is made of green cheese,
> and, if the moon is made of green cheese then 2+2=4
> then 2+2=4.

although the first premise is false.

This illustrates that a scheme like *modus ponens* should be construed so as to describe the beliefs held by a particular deductive reasoner *X*:

> If *X* holds *A*,
> and *X* holds $A \rightarrow B$,
> then *X* holds *B*.

[23] The conclusion sign ∴ separates the premises from the conclusion.

Whether the beliefs held by *X* are actually true is a completely different issue. Tarskian semantics proves that if we start from true propositions, *modus ponens* will never derive a false proposition, but it is clear that this cannot hold for any abductive scheme. I thus propose to paraphrase Peirce’s scheme in terms of an abductive reasoner *Y*:

> If *Y* observes *C*,
> and ‘if *A* wer[redacted]rue, *C* would be a matter of course’,
> then *Y* conje[redacted]res *A*.

I write ‘*Y* [redacted]*jectur*[redacted] *A*’ rather than ‘*Y* holds *A*’ in order to indicate that *Y* will not be inclined to[redacted]dopt e[redacted]ductive explanations she conceives as possible.

It remai[redacted] to dec[redacted]e upon the interpretation of the second premise of this scheme. Let us try ‘*Y* h[redacted] *C*, it seems likely that *Y* will believe that *C* is tr[redacted] — bu[redacted] by definition of material implication, *Y* necessarily believes that $A \rightarrow$[redacted] true f[redacted] any *A*, implying that *Y* is ready to conjecture any proposition whatsoev[redacted]tation is that *A* should *logically* entail *C* rather than mate[redacted]y, i.e. [redacted]$\rightarrow C$[24]. We thus arrive at the following paraphrase:

> If [redacted]serves *C*,
> an[redacted]
> the[redacted] conjectures *A*.

For conv[redacted]f *Y* observes *C*, then *Y* conjectures *A*’ to ‘*Y* holds that *C* is [redacted]ductively explained by *A*’; usually the reference to the abductive reasoner *Y* will be o[redacted]ted, leading to the final scheme:

> If [redacted]$\rightarrow C$,
> the[redacted]*C* is abductively explained by *A*.

Note that Peirce presented his scheme as a **definition** of the logical form of abduction, i.e. *C* is abductively explained by *A* if, and only if, *A* logically entails *C*. In this thesis I will also consider more liberal forms of abduction, such that the above scheme represents an adequacy condition which is to be satisfied (in one direction only) by any form of abduction, rather than a logical equivalence.

Adequacy conditions for abduction

Having thus clarified Peirce’s proposed definition of the relation ‘is abductively explained by’, we can investigate which of the adequacy conditions considered by Hempel it satisfies. These conditions are reproduced below, with phrases like ‘confirms’ replaced by ‘is explained by’. The conditions are renumbered (P1–5); (~~P1~~) denotes that this condition is invalid for Peirce’s definition.

[24] I.e. *A* materially implies *C* in every possible interpretation; this condition may be considered too strong, but note that it will be treated in the sequel as a sufficient rather than a necessary condition.

(~~P1~~) *Entailment condition*: any sentence which is entailed by an observation report explains it.

  (P1.1) Any observation report explains itself.

(~~P2~~) *Consequence condition*: if an observation report is explained by every one of a class *K* of sentences, then it is also explained by any sentence which is a logical consequence of *K*.

  (~~P2.1~~) *Special consequence condition*: if an observation report is explained by a hypothesis *H*, then it is also explained by every consequence of *H*.

  (P2.2) *Equivalence condition*: if an observation report is explained by a hypothesis *H*, then it is also explained by every hypothesis which is logically equivalent with *H*.

  (P2.3) *Conjunction condition*: if an observation report is explained by each of two hypotheses, then it is also explained by their conjunction.

(~~P3~~) *Consistency condition*: every logically consistent observation report is logically compatible with the class of all the hypotheses by which it is explained.

  (~~P3.1~~) Unless an observation report is self-contradictory, it is not explained by any hypothesis with which it is not logically compatible.

  (~~P3.2~~) Unless an observation report is self-contradictory, it is not explained by any hypotheses which contradict each other.

(P4) *Equivalence condition for observations*: if an observation report *B* is explained by a hypothesis *H*, then any observation report logically equivalent with *B* is also explained by *H*.

(P5) *Converse consequence condition*: if an observation report is explained by a hypothesis *H*, then it is also explained by every formula logically entailing *H*.

The entailment condition (P1) is clearly invalid — entailment is not a special case of 'is explained by'. However, (P1.1) is valid, and we will see shortly that it can be used to derive a replacement for the entailment condition. Clearly, the consequence conditions (P2) and (P2.1) are invalid as well: explanations cannot be arbitrarily weakened. Equivalence condition (P2.2) remains valid. Conjunction condition (P2.3) is valid only by virtue of the fact that inconsistent explanations are allowed — which is also the reason why (P3.1) is invalid. The other consistency conditions (P3) and (P3.2) are clearly invalid. The equivalence condition for observations (P4) is trivially valid.

Finally, and most interestingly, the converse consequence condition (P5) is valid: a given explanation can be arbitrarily strengthened. Furthermore, since every observation report is explained by itself by (P1.1), we can use the converse consequence condition to derive the following:

(P1) *Converse entailment condition*: an observation report is explained by every formula logically entailing it.

In other words: converse entailment is a special case of 'is explained by'; (P1) represents one half of Peirce's definition of abduction. As already indicated, I will consider several alternative definitions, but each of them satisfies the converse entailment condition (P1) — with a minor modification, since I will require that any observation report and any explanation be compatible. Note that this also means that the conjunction condition (P2.3) becomes invalid, and (P3.1) becomes valid. The exact list of adequacy conditions for abductive or, as I will call it, explanatory reasoning is given at the end of the chapter.

The confirmation paradox

A question we did not yet address is whether these adequacy conditions for 'is explained by' are mutually compatible. For instance, doesn't the converse consequence condition (P5) reintroduce the confirmation paradox? This latter question is easily answered: it does not, since the consequence conditions (P2) have been disabled. The more general question is left for chapter 7, where a formal analysis demonstrates that they are indeed compatible.

This, then, is the solution I propose to get rid of the confirmation paradox: to make a clear distinction between the statements 'these observations confirm this hypothesis' and 'these observations are explained by this hypothesis'. The fundamental difference between the property of being a confirmed hypothesis, and the property of being an explanatory hypothesis, is that the former property is passed on to logical consequences, as expressed by the consequence condition, and the latter is passed on to logically stronger explanations, as expressed by the converse consequence condition. If the consequence condition seems intuitively valid, it is because we think of confirmed hypotheses; if the converse consequence conditions seems intuitively valid, it is because we think of explanatory hypotheses. The 'confirmation paradox' is not a paradox at all, but demonstrates rather convincingly that 'confirms' and 'is explained by' are quite different relations.

This being said, it may seem natural to study the combination of these two relations — that is, to investigate a formalisation of the statement 'these observations confirm **and** are explained by this hypothesis'. Since this statement is a simple conjunction of the two statements previously mentioned, the resulting binary relation is the intersection of the relations 'confirms' and 'is explained by'. This approach would avoid paradoxes such as the confirmation paradox, since the two concepts of confirmation and explanation are not confused, but clearly separated before they are combined. However, I believe that the compound concept can only be successfully studied after a sufficient understanding of its parts has been obtained. It is this understanding that is primarily pursued in this thesis.

Furthermore, and more importantly, inductive reasoning can also be successfully applied to infer hypotheses that do not have explanatory power, but exhibit certain regularities implicit in the observations. That is, *the goal for which inductive hypotheses are sought influences the logical relation between evidence and hypothesis*. Therefore, I think that it makes sense to speak about two genuinely distinct forms of inductive reasoning: the 'Peircean' form of induction, which I will call *explanatory induction*, and which is aimed at finding explanations of the observations; and the 'Hempelian' form, which I will call *confirmatory induction*, the primary aim of which is to find other generalisations, with no or limited explanatory power. In order to stress the fact that I am mostly concerned with the first step of inductive hypothesis formation, I will usually speak of explanatory and confirmatory *reasoning* rather than induction. I will use the term *conjectural reasoning* whenever the question whether it is explanatory or confirmatory is immaterial.

The satisfaction criterion of confirmation

I will now demonstrate that Hempel's ideas, which led him to the formulation of the satisfaction criterion of confirmation, can also be used to derive a slightly different but very similar criterion, which seems to be more faithful both to his original ideas and to the notion of a *satisfaction* criterion.

Consider Hempel's original example, cited above: the observations

$$\text{Raven}(a)\wedge\text{Black}(a) \wedge \neg\text{Raven}(c)\wedge\text{Black}(c) \wedge \neg\text{Raven}(d)\wedge\neg\text{Black}(d)$$

confirm the hypothesis $\forall x$: Raven(x)→Black(x), since the latter is true as far as the objects a, c and d are concerned. Hempel formalises this by reconstructing what the hypothesis has to say about each of these objects. Alternatively, we can reconstruct what the observations have to say if a, c and d are the **only** (relevant) objects. This can be effected by switching from a formula expressing the observations to a designated **interpretation**. Basically, there is only one interpretation over the domain $\{a, c, d\}$ satisfying the observations. We can then say that any formula that is satisfied by this designated interpretation is confirmed by the observations.

This raises a number of issues. What if the knowledge that d is not a raven is missing from the observations? Clearly, in that case there are two interpretations over the domain $\{a, c, d\}$ satisfying the observations, one in which d is a raven, and one in which d is not. Should we require a confirmed hypothesis to be satisfied by both interpretations? If so, the above hypothesis is not confirmed. Alternatively, we could require that a confirmed hypothesis be satisfied by at least one of these interpretations, or we could select one of them as the designated interpretation (for instance by stipulating that if d is not known to be a raven, it is expected not to be).

Another problem arises if the hypothesis talks about objects not present in the observations. Recall that, according to Hempel's criterion, the observations $P(a)\wedge P(b)$ confirm $\forall x$: $P(x)\vee Q(x)$, but also ($\forall x$: $P(x)\vee Q(x)$) $\wedge$ ($P(c)\vee Q(c)$) and hence $P(c)\vee Q(c)$. Should we map c to one of the observed objects a and b? Should we say that, since it is not known whether c is a P or a Q, it is expected to be neither? Or should we restrict attention to hypotheses that talk only about objects present in the observations?

Clearly, the current proposal falls short of being a precise definition, but an improvement will have to wait until chapter 7. What I want to demonstrate here is that there is, at least intuitively, quite a close connection between Hempel's criterion of confirmation and so-called designated or *preferred model semantics*. In fact, I think that defining confirmation in terms of a designated model or models reflects more closely the intuitions about confirmation, including the intuitions given by Hempel.

I should add that the idea of applying preferred model semantics to formalise inductive hypothesis formation is not new. As will be described in chapter 3, similar ideas have been proposed by (Helft, 1989) and (De Raedt, 1993). These authors did not formulate their work in terms of the concept of a confirmed hypothesis, however.

Degree of confirmation and its relation to the qualitative concept

There is one remaining issue we need to discuss here, which is the supposed quantitative nature of confirmation. Indeed, it seems very natural to ask, when a body of evidence is said to confirm a hypothesis, to which degree this is so. In this section I intend to address this issue by spending a few words on the work of Rudolf Carnap, who has proposed a rather elaborated system of what he calls 'inductive logic'.

Before investigating his proposal, I should like to remark that Carnap's writings are classic cases of the scientific precision advocated by the Wiener Kreis. Consider, as a case in point, his notion of explication:

> 'The task of ***explication*** consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second. We call the given concept (or the term used for it) the ***explicandum***, and the exact concept proposed to take the place of the first (or the term proposed for it) the ***explicatum***. The explicandum may belong to everyday language or to a previous stage in the development of scientific language. The explicatum must be given by explicit rules for its use, for example, by a definition which incorporates it into a well-constructed system of scientific either logicomathematical or empirical concepts.' (Carnap, 1950, p.3)

Carnap then proceeds by stating the requirements for an explicatum: similarity to the explicandum, exactness, fruitfulness, and simplicity. The reason that I chose this citation is that it nicely characterises what I have set out to do in this thesis: explicating the logical aspects of inductive reasoning.

In his *Logical Foundations of Probability* (1950) Carnap develops a system of *inductive logic* based on a quantitative notion of *degree* of confirmation or *c-function*, which is a function $c(H,E)$ assigning a number between 0 and 1 to a hypothesis H on the basis of evidence E. Such c-functions should be *regular*, in the sense that they obey some general adequacy conditions for probability measures. As has been remarked in §4, such a degree of confirmation is not a statistical fact about the world in terms of relative frequency — in fact, according to Carnap it should be interpreted as an 'estimate of the relative frequency' (§41D, p.168). Carnap formulates his conception of an inductive logic as follows:

> 'What we call inductive logic is often called the theory of nondemonstrative or nondeductive inference. Since we use the term 'inductive' in the wide sense of 'nondeductive', we might call it the theory of inductive inference... However, it should be noticed that the term 'inference' must here, in inductive logic, not be understood in the same sense as in deductive logic. Deductive and inductive logic are analogous in one respect: both investigate logical relations between sentences; the first studies the relation of [entailment], the second that of degree of confirmation which may be regarded as a numerical measure for a partial [entailment]... The term 'inference' in its customary use implies a transition from given sentences

> to new sentences or an acquisition of a new sentence on the basis of sentences already possessed. However, only deductive inference is inference in this sense.' (§44B, pp.205–6)

Thus, in Carnap's view an inductive logic is based on a notion of 'partial' entailment, expressing how well a conclusion is confirmed by the premisses: the higher the degree of confirmation, the more plausible the conclusion becomes.

I should like to stress that I disagree with Carnap in two, quite crucial respects. First of all, in this thesis I do **not** interpret inductive as 'nondeductive' — on the contrary, the main motivation for my research has always been that induction should (and can) be defined in its own right, and not negatively in terms of something else. Secondly, I disagree with Carnap on the conception of a logic. As will be elaborated in chapter 5, what Carnap calls 'inductive logic' is in fact a semantics for estimating the plausibility of a formula given the truth of others. Although certainly useful, such a semantics does not induce a notion of proof (as noted by Carnap). However, there exists an alternative notion of semantics which does have a natural proof-theoretical counterpart, and which I call *preservation semantics*. It is the combination of preservation semantics, proof procedure, and metatheory which I will call a logical system.

What should interest us, however, is what Carnap has to say about the qualitative concept of confirmation and the work of Hempel. Carnap notes in §86 that a qualitative relation of confirmation can be defined by $c(H,E)>c(H,\mathbf{true})$[25]. It should be noted that until this point no particular *c*-function has been fixed[26]; this translation merely serves to transform the adequacy conditions for regular *c*-functions into adequacy conditions for qualitative confirmation and *vice versa*. Carnap critically examines Hempel's adequacy conditions, and finds that quite a few of them don't correspond to the regularity of *c*-functions. While Carnap accepts the equivalence conditions (H2.2) and (H4), as well as (a qualified form of) the entailment condition (H1), the remaining conditions are invalidated by certain regular *c*-functions. Specifically, Carnap rejects both the consequence condition and the converse consequence condition, and remarks about the consistency condition (H3) that 'it seems to me not even plausible' (§87, p.476), which seems to be mainly caused by the perceived implausibility of the derived condition (H3.2). On the other hand, (H3.1), which basically states that evidence and hypothesis are compatible, is accepted.

Finally, Carnap considers Hempel's definition of confirmation. It is clear from the foregoing discussion of the adequacy conditions, some of which are rejected by Carnap, that this definition cannot be unconditionally accepted by Carnap either. In an interesting analysis, he shows that Hempel's definition is overly restrictive in the following sense: it 'holds only in the special case where the evidence ascribes to *all* individuals essentially occurring in it the property in question' (§88, p.481). The reason that this is considered overly restrictive is that, in Carnap's view, the main function of a qualitative notion of confirmation is that it reflects the qualitative aspects of a quantitative measure:

---

[25] In fact, Carnap also takes background knowledge B into account by requiring that $c(H,B\wedge E)>c(H,B)$; however, in order to retain the parallel with Hempel's work B is set to $\mathbf{true}$.

[26] The degree of confirmation suggested by Carnap is, in modern terminology, the fraction of Herbrand models of E in which H is also true (§110).

> 'Some years ago those who worked on these problems expected that, if and when a definition of degree of confirmation were to be constructed, it would be based on a definition of a nonquantitative concept of confirming evidence. However, today it is seen that this is not the case either for *dc* [a quantitative concept put forward by Hempel] nor for my definition of *c**, and it is not regarded as probable that it will be the case for other definitions which will be proposed. It appears at present more promising to proceed in the opposite direction, that is, to define a quantitative form of the concept of confirming evidence on the basis of an explicatum for degree of confirmation...' (§88, p.481)

In summary, the observed discrepancies between Carnap's desiderata for a quantitative measure of confirmation and Hempel's desiderata of a qualitative relation of confirmation lead Carnap to the conclusion that the latter notion, as perceived by Hempel, is inadequate to model the qualitative aspects of the former notion, as perceived by Carnap. This conclusion seems correct to me — but then, why should we want to develop an independent qualitative notion if it can be derived from the quantitative measure? This point is noted by Carnap as well:

> 'The task of finding an adequate explicatum for the classificatory concept of confirmation defined in purely classificatory, that is, nonquantitative terms is certainly an interesting problem; but it is chiefly of importance for those who do not believe that an adequate explicatum for the quantitative concept of confirmation can be found.' (§86, p.467)

Thus, Carnap considers the qualitative concept of confirmation as a rather poor, and in fact superfluous, surrogate for the real thing: a quantitative concept.

However, in my view the two concepts are conceived for quite different reasons, and the relation between them is less transparent than Carnap seems to think. I will not have space to defend this here — in a way, this whole thesis serves as an illustration of this point. To reiterate some remarks I made earlier: it will be shown in chapter 7 that the qualitative concept of confirmation does give rise to a full-fledged logical system, including a proof procedure, while the quantitative concept does not do so, even if we would only derive conclusions that are maximally confirmed by the evidence. In the words of Popper:

> 'Those who identify confirmation with probability must believe that a high degree of probability is desirable. They implicitly accept the rule: 'Always choose the most probable hypothesis!' Now it can be easily shown that this rule is equivalent to the following rule: 'Always choose the hypothesis which goes as little beyond the evidence as possible!'
> (Popper, 1963, pp.289–90)

In fact, Carnap has recognised this problem, but dissolves it by stating that inductive logic has a different goal:

> 'Inductive logic alone does not and cannot determine the best hypothesis on a given evidence... This preference is determined by factors of many different kinds...' (§46, p.221)

§7. *Discussion*

What I want to argue here is that Carnap's conception of an 'inductive logic' is actually a procedure which, *given some logical system*, can be used to estimate the plausibility of the formulas derived by that logical system[27]. That is, such a truth-estimating procedure is subordinate to, and only useful in conjunction with, some other logical system (see chapter 5 for a further discussion of this subject). In this thesis my primary interest lies with such a logical system for induction, and not with a truth-estimating procedure which takes the hypothesis as an input.

§8. SUMMARY AND CONCLUSIONS

In this chapter I have outlined the main philosophical aspects of induction relevant for the subject of this thesis. I have put special emphasis on the work of Peirce and Hempel, which I will take as a starting point for my subsequent investigations. From Peirce I borrow the idea of taking *explanation* as the central notion of inductive hypothesis formation. From Hempel I borrow the idea of taking *confirmation* as an alternative notion. Furthermore, I will use and formalise Hempel's tool of *adequacy conditions*.

However, unlike Hempel and Carnap, I consider it necessary to develop several, alternative sets of adequacy conditions or systems, each of them designed to formalise a distinguished kind of induction. I will distinguish two main families of systems, one formalising explanatory induction, the other formalising confirmatory induction. This distinction is justified by the idea that any formal theory of induction should take into account the purpose which the inductive conclusion is intended to fulfil. In explanatory induction this purpose is explaining the observations, while in confirmatory induction the purpose is to find non-explanatory but confirmed theories expressing regularities implicit in the observations.

This distinction also provides a solution to the confirmation paradox, which results from combining the consequence condition and its converse. I have argued that, while the consequence condition formalises an intuition about confirmatory reasoning, the converse consequence condition belongs to the realm of explanatory reasoning. We thus obtain two sets of adequacy conditions: (C1–4) for confirmatory reasoning, and (E1–4) for explanatory reasoning.

(C1) *Entailment condition*: any sentence which is entailed by a consistent observation report is confirmed by it.

(C1.1) Any consistent observation report is confirmed by itself.

(C2) *Consequence condition*: if an observation report confirms every one of a set *K* of sentences, then it also confirms any sentence which is a logical consequence of *K*.

(C2.1) *Special consequence condition*: if an observation report confirms a hypothesis *H*, then it also confirms every consequence of *H*.

---

[27]Carnap's work is closely related with so-called Bayesian methods for dealing with uncertainty in a numerical way (Pearl, 1987).

(C2.2) *Equivalence condition*: if an observation report confirms a hypothesis *H*, then it also confirms every hypothesis which is logically equivalent with *H*.

(C2.3) *Conjunction condition*: if an observation report confirms each of two hypotheses, then it also confirms their conjunction.

(C3) *Consistency condition*: every consistent observation report is compatible with the set of all the hypotheses which it confirms.

(C3.1) *Special consistency condition*: an observation report is compatible with any hypothesis which it confirms.

(C3.2) An observation report does not confirm any hypotheses which contradict each other.

(C4) *Equivalence condition for observations*: if an observation report *B* confirms a hypothesis *H*, then any observation report logically equivalent with *B* also confirms *H*.

Insofar there are differences with Hempel's conditions, these are caused by the fact that Hempel allows contradictory observation reports (which confirm any hypothesis), while I don't.

The adequacy conditions for abductive or, as I will call it in this thesis, explanatory reasoning are as follows.

(E1) *Converse entailment condition*: an observation report is explained by every consistent hypothesis entailing it.

(E1.1) Any consistent observation report explains itself.

(E2) *Converse consequence condition*: if an observation report is explained by a hypothesis *H*, then it is also explained by every consistent formula entailing *H*.

(E2.1) *Equivalence condition*: if an observation report is explained by a hypothesis *H*, then it is also explained by every hypothesis which is logically equivalent with *H*.

(E3) *Special consistency condition*: an observation report is compatible with every hypothesis by which it is explained.

(E4) *Equivalence condition for observations*: if an observation report *B* is explained by a hypothesis *H*, then any observation report logically equivalent with *B* is also explained by *H*.

Notice that the converse consequence condition (E2) is not valid for a set *K* of hypotheses, in the manner of the consequence condition (C2): most likely, alternative explanatory hypotheses will be incompatible.

I have also raised some points that merely serve as a preview of claims made further on in this thesis. I have suggested that an alternative formalisation of Hempel's satisfaction criterion of confirmation can be based on the notion of a designated or preferred model semantics, as will be elaborated in chapter 7. Furthermore, I have argued that a relation of confirmation is more than just a qualitative plaster cast of a function assigning degrees of confirmation, since the former, unlike the latter, gives way to a full-fledged logical system. This raises the question as to what constitutes a logical system, a question that will be considered in chapter 4.

* *