Acting against one's best judgement
Peijnenburg, Adriana Johanna Maria

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1996

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
CHAPTER III
THE LOGICAL CONNECTION ARGUMENT

0. Introduction

As we have seen in Chapter II, Hempel holds that action explanations look like causal explanations, although reasons are not causes. Philosophers claiming a fundamental difference between the two kinds of explanation have therefore criticised Hempel in one of two ways. Some have argued that, since reasons are not causes, action explanations cannot possibly be causal explanations. Others maintained that action explanations do differ from causal ones, although reasons are causes. The first course is followed by adherents of the later Wittgenstein; of the second Donald Davidson is the leading man. The present chapter deals with the first kind of criticism; Davidson’s causal theory on action explanation will be discussed in the next chapter.

1. Intentions, reasons, causes

The view that reasons differ from causes, and that therefore explanations by reasons are not causal explanations, gained influence through Wittgenstein’s preliminary studies for the Philosophical Investigations, viz. The Blue and Brown Books. It is not easy to gather from these notes a clear idea of what Wittgenstein II has in mind, but it is clear enough that he occasionally warns against confusing reasons with causes (Wittgenstein 1958, 14ff; cf. 110ff; 143). In doing so, Wittgenstein appears to regard the cause of an action as something from which a series of reasons can originate:

"Thus when the chain of reasons has come to an end and still the question "why?" is asked, one is inclined to give
Furthermore, Wittgenstein considers the statement that your action has such-and-such a cause to be a hypothesis, on the ground that you cannot know, but merely conjecture what the cause of your action is. On the other hand, the assertion that your action has such-and-such a reason cannot be called hypothetical, for it is presumed that one knows his own motives (Wittgenstein 1958, 15 - notice that Wittgenstein here uses the terms 'reason' and 'motive' interchangeably). This is not to say, Wittgenstein tells us, that reasons are a subclass of causes, viz. the causes that we know and recognise; Wittgenstein denies that "a motive is a cause of which we are immediately aware, a cause 'seen from the inside’, or a cause experienced" (Wittgenstein 1958, 15). The point is rather that 'cause' and 'reason' stem from different language games. If taken in conjunction with the appropriate empirical laws, causes enable us to predict actions; reasons, on the other hand, escape from a language game in which the notions of prediction and empirical laws feature. As far as this point is concerned, The Blue and Brown Books matches the work which it prepares: the claim that the difference between reasons and causes hinges on a difference between language games, is borne out by the Philosophical Investigations I, notably by Sections 630-633.

These brief remarks from The Blue and Brown Books were developed further by philosophers such as Von Wright, Melden, and Anscombe. In elaborating Wittgenstein’s ideas on reasons and causes, they frequently referred to the concept intention and to related notions: intentional, intentionality, intending and so on. Now Wittgenstein himself,

26 Cf. the Philosophical Investigations:

"211. How can he know how he is to continue a pattern by himself - whatever instruction you give him? - Well, how do I know? - if that means 'Have I reasons?' the answer is: my reasons will soon give out. And then I shall act, without reasons.

212. When someone whom I am afraid of orders me to continue the series, I act quickly, with perfect certainty, and the lack of reasons does not trouble me. ....

326. We expect this, and are surprised at that. But the chain of reasons has an end."
in *The Blue and Brown Books*, refers to the concept of intending only twice (Wittgenstein 1958, 32 and 147). Unfortunately these references are not only fleeting, but even give off a contradictory glow. For the one says that intending "is neither a particular state of mind nor a particular mental process" (p. 32), whereas the other has it that intending does "refer to certain ... states of mind" (p. 147). Perhaps, as Wittgenstein seems to suggest on page 32, an intention consists in a *combination* of many states of mind.

In general, ‘intentionality’ is said to have two senses, a broad sense and a restricted one. The broad sense is connected to Brentano’s epistemology and to Husserl’s phenomenology. Here ‘intention’ and ‘intentionality’ refer to a supposed essential feature of all mental phenomena, viz. that those phenomena are about or are directed towards something. The Brentano-Husserl notion of the directedness of the mental springs from the mediaeval notions *intentio* and *intendere* which, in their turn, are rooted in Aristotle’s idea that when we know or want an object, our mind takes on the shape of that object. On the other hand, the restricted meaning of ‘intention’ and ‘intentionality’ does not have roots that stretch out that far in the history of philosophy. In its restricted meaning ‘intention’ refers to one particular way of being directed towards something; here intending plainly boils down to the everyday notion of *planning to do something*.

It is not altogether clear which of the two meanings Wittgenstein II and his interpreters such as Winch and Anscombe have in mind. Judging from their interest in ordinary language, it is probably the second, restricted meaning. But even if we assume that all of them use the restricted meaning and accordingly take ‘intending’ to mean ‘planning to do something’,

---

27 As has been lucidly explained by Føllesdal, there are interesting differences between Brentano and Husserl insofar as their conceptions of ‘intentionality’ as ‘directedness of the mental’ are concerned. For Brentano, the directedness of a mental phenomenon towards an object implies the existence of that object. Husserl, on the other hand, maintains that a mental phenomenon can also be directed towards something that does not exist (Husserl mentions dreams and hallucinations). Thus, whereas Brentano can describe intentionality by referring to the object towards which the mental phenomenon is directed, Husserl must follow a different route. In Føllesdal’s explanation, Husserl "[o]kused on what the directedness consists of: what are the features of the mental thanks to which it is always as if it has an object?” (Føllesdal 1986, 109-110). Husserl called the collection of these features the *noema*, which, according to Føllesdal’s illuminating explication, Husserl regarded as “a generalized theory of meaning, thereby tying together intension with a *t* and intension with an *s*.” (Føllesdal 1986, 110).
differences remain. For some interpreters regard planning to do something as a state or an episode, others look upon it as a property, and still others consider it to be a (mental) event, or even an act.\textsuperscript{28}

\textsuperscript{28} Thus Peter Winch considers an intention as an act, and "a statement of intention" as an announcement of an act (Winch 1958, 81). Winch’s standpoint might be confirmed by \textit{Philosophical Investigations I}, Section 644: “I am not ashamed of what I did then, but of the intention which I had’ - And didn’t the intention reside also in what I did?” (emphasis by Wittgenstein).

A.I. Melden, on the other hand, regards an intention as a motive (Melden 1961, 83). Melden describes a motive as something that explains an action (Melden 1961, 88-89). However, he stresses that a motive is not an event distinct from the action and therefore cannot be seen as the cause of the action (Melden 1961, 89). According to Melden, a "statement of intention" is a way "to explain oneself" or "a declaration of a man’s reasons for doing what he does" (Melden 1961, 101 and 176-177). Following Wittgenstein (Wittgenstein 1958, 15), Melden thus regards ‘reason’ and ‘motive’ as synonyms. Yet Winch appears to take an opposite view: "The terms ‘reason’ and ‘motive’ are not synonymous" (Winch 1958, 82).

Stuart Hampshire seems to regard an intention as a property of an act, in the sense that it is the quality which makes an act someone’s act. An intention gives the agent "his sense of being in the world", and it prevents him "from thinking of himself as a neutral point" (Hampshire 1959, 69); it renders continuity in the discontinuity between what you are doing now and what you did in the past (Hampshire 1959, 72). Hampshire’s view might be prompted by what Wittgenstein writes in Section 659 of the \textit{Philosophical Investigations I}: "Why do I want to tell him about an intention too, as well as telling him what I did? - Not because the intention was also something which was going on at that time. But because I want to tell him something about myself, which goes beyond what happened at that time. I reveal to him something of myself when I tell him what I was going to do.” (emphasis by Wittgenstein).

Von Wright considers an intention to be an inner aspect of an action. In Von Wright’s account, an action has two "aspects", an inner and an outer one. The outer aspect is twofold. It encompasses the "immediate outer aspect" as well as the "remote outer aspect". The former is the "muscular activity - e.g. a turning of the hand or raising of an arm". The latter is "some event for which this muscular activity is causally responsible, - e.g. the turning of a handle or the opening of a window" (Von Wright 1971, 86-87). The inner aspect, on the other hand, is "the intentionality of the action, the intention or will ‘behind’ its outer manifestations" (Von Wright 1971, 86). That the term ‘intention’ in Wittgenstein’s writings denotes a by definition inner aspect is perhaps confirmed by passages like that in the \textit{Philosophical Investigations I}, Section 641: "... the most explicit expression of intention is by itself insufficient evidence of intention.”

G.E.M. Anscombe wrote an entire book on the subject of intention. She distinguishes three major uses of the term ‘intention’ (see below, Chapter VI, Section
However, I do not wish to get entangled in an abortive attempt to find out what ‘intention’ (in the restricted sense) really means. Like the terms ‘reason’, ‘belief’ and ‘desire’, the term ‘intention’ is probably best regarded as a node in a whole gamut of mental terms, the meanings of which have become bogged down in the morass of our linguistic intuitions. Either we leave them where they are and give up analysing them, or we try to pull them out from the swamp. If we choose the second option, there is some work for us to do. For dragging them out of the mud and exposing them to public view requires that we clean them up, that is, polish, stylise and somewhat formalise them. As we have seen, Hempel started some cleaning with respect to the terms ‘belief’, ‘desire’ and ‘reason’; he swept away vagueness from the realm of terminology by stating that beliefs and desires are dispositions, and by considering ‘reason’ as a term that denotes a belief/desire pair. Moreover, Hempel made those dispositions quantifiable by referring to mathematical decision theory. Like Hempel, I will eschew vagueness by making a decision: I stipulate that an intention is synonymous with a reason in Hempel’s sense. This stipulation has of course a footing in re. For desiring $x$ and believing that $y$ is the means for $x$ (which is an example of a reason in Hempel’s sense) very much resembles planning to do $y$ (which is an example of an intention in the strict sense). Hence I shall, from now on, use the term ‘intention’ interchangeably with the Hempelian ‘reason’. In doing so, I act in the spirit of Anscombe, for whom intentions and reasons are intimately connected (cf. the previous footnote; but see Bratman 1987 for a dissident view). Apparently the same spirit entered Donald Davidson when he wrote:

2.4.8). One of them occurs when we speak of intentional actions. Anscombe defines intentional actions as actions to which the question ‘Why’ is given application. A why-question in Anscombe’s sense is a question to which the answer may “(a) simply mention past history, (b) give an interpretation of the action, or (c) mention something future” (Anscombe 1968 (1957), 24). Anscombe adds that in cases (b) and (c) the answer is characterised as a reason for acting.

In Edward’s Encyclopedia of Philosophy Bruce Aune classifies the different meanings of ‘intention’ by bringing them all under two headings: intention as a disposition and intention as an event. Some experts on Wittgenstein will claim that this classification leaves Wittgenstein out, because in his usage ‘intention’ would neither denote a disposition nor an event. The difference is roughly that one can determine the presence of a disposition or an event whereas it is impossible to determine the presence of an intention (Michel ter Hark, private communication).
"An intention to act (or to refrain from acting) requires both a belief and a desire or pro-attitude: a desire or pro-attitude toward outcomes or situations with certain properties, and a belief that acting in a certain way will promote such an outcome or situation." (Davidson 1987, 40).

"Someone who acts with a certain intention acts for a reason; he has something in mind that he wants to promote or accomplish. A man who nails boards together with the intention of building a squirrel house must want to build a squirrel house, or think that he ought to ... and he must believe that by nailing the boards together he will advance his project." (Davidson 1978, 83).

I will say more on Davidson’s notion of intending in Chapter VI, Section 2.4.8. For the moment only two points are of importance: first, that I treat intentions and reasons for acting as synonyms, and second, that many philosophers following Wittgenstein II argued for a difference between reasons or intentions on the one hand and causes on the other. I shall discuss the consequences of the second point below.

2. The logical connection argument

When Winch, Von Wright, Melden, Hampshire, and a host of other Wittgensteinians state that reasons differ from causes, they mean that the relation between reasons and actions differs from the relation between causes and effects. More particularly, they claim that the relata are logically independent in the latter relation, but not in the former one. This claim has become widely known as the logical connection argument (LCA) (Stoutland 1970).

In spite of the frequency with which it was promoted, the LCA was never wholly convincing. For all of the suggestive writings on the subject, it never became clear what exactly is meant by the statement that reasons and actions are logically dependent upon each other. I therefore fully agree with Simon Evnine when he declares that "the logical connection thesis ... has never been stated as clearly as one would like, and many of its
In my view, a large part of the uncertainty about the LCA arises from the fact that it has been described in at least three different ways. In the description by Melden and Stoutland the LCA becomes a *semantical argument* (Melden 1961, 89ff; cf. Stoutland 1970). As a semantical argument the LCA says that descriptions of intentions are conceptually or analytically linked to descriptions of actions, whereas the links between causes and effects are said to be empirical or synthetical. Thus, if $D-1$ is ‘wanting $A-2$’ and $B-1$ is ‘believing that $A-1$ is a means for achieving $A-2$’, then $D-1$ and $B-1$ analytically imply ‘doing $A-1$’. Consequently, the statement in which $D-1$ and $B-1$ are ascribed to a person $P$ logically or analytically implies the statement that $P$ will do $A-1$. Kathleen Lennon, on the other hand, describes the LCA as an *ontological argument* (Lennon 1990, 73ff). In Lennon’s description the LCA states that intentions are necessarily linked to actions, whereas the links between causes and effects are said to be contingent or empirical. Hence, if $P$ has $D-1$ and $B-1$, then $P$ will necessarily do $A-1$.

Finally, Von Wright strongly suggests that the LCA is an *epistemological argument* (Von Wright 1971, 107ff). As such the LCA implies that if one knows $P$ to have $D-1$ and $B-1$, one can know a priori that $P$ does $A-1$. To use Von Wright’s word: one can only "verify" that $P$ does $A-1$ if one knows that $P$ has $D-1$ and $B-1$, and vice versa. "In this mutual dependence of the verification of premises and the verification of conclusions in practical arguments consists, as I see it, the truth of the Logical Connection Argument." (Von Wright 1971, 117).

For my purposes, the distinction between semantics, ontology and epistemology is irrelevant. Therefore I once more make a clean sweep: any distinction between the semantical pair analytic/synthetic, the ontological pair necessary/contingent, and the epistemological pair a priori/a posteriori will be ignored in this chapter (but not in all the subsequent ones!). I deem the LCA to be at the same time a semantical, an ontological and an epistemological argument.

There is still another difference that I shall ignore, namely that between logical truth and analytical truth. A sentence like

$$(p) \text{ Socrates is white or Socrates is not white}$$

is often called logically true, since its truth can be ascertained solely by
taking the logical words ‘or’ and ‘not’ into account; the meanings of the non-logical or descriptive terms (‘Socrates’, ‘white’) are irrelevant here. On the other hand, a sentence like

(q) If John is a bachelor, then John is not married

is (merely) analytically true, for to ascertain the truth of (q), we do have to know the meanings of non-logical words (‘John’, ‘bachelor’, ‘married’). However, I shall treat (p) and (q) as if they are on a par, and consequently disregard any distinction between ‘logical’ on the one hand and ‘analytical’ (‘tautological’, ‘conceptual’, ‘deductive’, or whatever) on the other.

Obviously, if the LCA is correct, then Hempel’s schema of action explanation is false. As we have seen in Chapter II, the latter schema can be represented by (AE hs ), which was our simplification of (AE b):

\[
\begin{align*}
  (\text{II.17}): \quad & \forall(x) \left\{ \left( R(x) \land D-1(x) \land B-1(x) \right) \rightarrow A-1(x) \right\} \\
  (\text{II.18}): \quad & R(a) \\
  (\text{AE hs}): \quad & \text{II.19): } D-I(a) \\
  \quad \quad & \text{II.20): } B-I(a) \\
  \quad \quad \quad & \text{II.21): } A-I(a)
\end{align*}
\]

In (AE hs) (II.17) is supposed to be an empirical law. However, if the LCA is correct, then the schema for action explanation becomes (AE lca):

\[
\begin{align*}
  (\text{III.1}): \quad & \forall(x) \left\{ \left( D-1(x) \land B-1(x) \right) \rightarrow A-1(x) \right\} \\
  (\text{II.19}): \quad & D-I(a) \\
  (\text{II.20}): \quad & B-I(a) \\
  \quad \quad \quad & \text{II.21): } A-I(a)
\end{align*}
\]

from which the premise (II.18) has disappeared. Moreover, since according to the LCA (III.1) is analytical, that premise can be left out too. This means that (AE hs) can be reduced to (AE lca):

\[
\begin{align*}
  (\text{AE lca}): \quad & \forall(x) \left\{ \left( D-1(x) \land B-1(x) \right) \rightarrow A-1(x) \right\} \\
  (\text{II.19}): \quad & D-I(a) \\
  (\text{II.20}): \quad & B-I(a) \\
  \quad \quad \quad & \text{II.21): } A-I(a)
\end{align*}
\]
III: The Logical Connection Argument

\[ (\text{AE}^{\text{lca}}) \]

\[ (\text{II.19}): D-1(a) \]

\[ (\text{II.20}): B-1(a) \]

\[ (\text{II.21}): A-1(a) \]

(the ‘a’ added to ‘lca’ symbolises the absence of (III.1)). The schema \((\text{AE}^{\text{lca}})\) bears a close resemblance to the schema for action explanation that was put forward by the most influential defender of the LCA, Georg Henrik von Wright. According to Von Wright the explanation of an action takes the form of \((\text{AE}^{\text{wright}})\):

\[ (\text{III.2}): P \text{ intends to bring about } A-2 \]

\[ (\text{III.3}): P \text{ considers that he cannot bring about } A-2 \text{ unless he does } A-1 \]

\[ (\text{III.4}): \text{Therefore, } P \text{ sets himself to do } A-1 \]

(Von Wright 1971, 96). Von Wright calls \((\text{AE}^{\text{wright}})\) a practical inference, but also speaks of a practical argument, a practical reasoning, and even a practical syllogism. He thus underlines the far-reaching similarity between \((\text{AE}^{\text{wright}})\) on the one hand and Aristotle’s practical syllogism on the other. In Chapter I we have seen what the latter looks like, and it is easy to perceive its resemblance to Von Wright’s schema. I return to this point at the end of the present chapter, but meanwhile I regard \((\text{AE}^{\text{lca}})\) as a

---

29 Von Wright has \(A, p\) and \(a\) for \(P, A-2\) and \(A-1\) respectively. On pages 96-106 Von Wright imposes a number of modifications on \((\text{AE}^{\text{wright}})\). His “final formulation of the inference schema” is on page 107:

"From now on \(P\) intends to bring about \(A-2\) at time \(t\)
From now on \(P\) considers that, unless he does \(A-1\) no later than at time \(t'\), he cannot bring about \(A-2\) at time \(t\)

Therefore, no later than when he thinks time \(t'\) has arrived, \(P\) sets himself to do \(A-1\), unless he forgets about the time or is prevented."

Notice that in Von Wright’s analysis two different times occur, and that \(t = t' + \Delta\), where \(\Delta > 0\). In the analysis by Hempel, which is based on Carnap’s study of dispositions, only one single time-variable is used (see the text).
formalised version of \((AE^{\text{wright}})\), and the latter as a simplified practical syllogism à la Aristotle.

The rest of the present chapter is devoted to a comparison of Hempel’s schema \((AE^{\text{hs}})\) with \((AE^{\text{lcaa}})\). My final claim will be that the two schemata do not differ. This may sound fairly surprising, for when we contrast \((AE^{\text{lcaa}})\) with \((AE^{\text{hs}})\) two evident differences catch the eye. I call them DIF-1 and DIF-2:

DIF-1: premise (II.17) - the empirical law - is present in \((AE^{\text{hs}})\) and absent from \((AE^{\text{lcaa}})\)

DIF-2: premise (II.18) - the rationality assumption - is present in \((AE^{\text{hs}})\) and absent from \((AE^{\text{lcaa}})\).

Stringent though they may look, I shall nevertheless argue that DIF-1 and DIF-2 are in fact pseudo-differences. As will be seen below, it follows from Hempel’s assumptions that the property of being a rational agent lacks empirical import, so that (II.18) becomes necessarily true and is hence not testable by experiment. This means that we may eliminate (II.18) from \((AE^{\text{hs}})\), which makes a pseudo-difference out of DIF-2. Also, Hempel’s reasoning indicates that (III.1) is conceptually true. Given the analyticity of (II.18), this entails that (II.17) cannot be empirical in character but must be analytical too. That turns DIF-1 into a pseudo-difference. Finally, I shall demonstrate that the analyticity of (II.17) and that of (II.18) presuppose one another. In other words: DIF-1 and DIF-2 are two sides of the same coin.

Of course, if DIF-1 and DIF-2 are pseudo-differences in this sense, then Hempel’s ideas on the function and characteristics of scientific explanations are called into question. For then Hempel’s idea that all scientific explanations fall under a common denominator becomes badly bruised. As we have seen in Chapter II, Section 1, this common denominator consists of four conditions of adequacy, (CA1)-(CA4). However, if DIF-1 and DIF-2 are fake, then two conditions are defied. Condition (CA3) - that all premises have to be testable - is defied by the untestability of (II.18), whereas condition (CA2) - that the explanans contains at least one empirical
III: The Logical Connection Argument

law - is broken by the analyticity of (II.17). Moreover, if DIF-1 and DIF-2 are pseudo-differences, then Hempel’s schema for action explanation (AE^hs) is deficient, and the LCA is sound. This is not to say that the LCA provides us with a correct model for action explanation; as we will see, it does not, since it confronts us with the *akrasia* problem. When I call the LCA a sound argument, I mean that it is adequate as a critique of Hempel’s theory. It is sound *to the extent that from Hempel’s assumptions it follows that* (AE^lcaas) *underlies (AE^hs)*. It is not sound in the sense that (AE^lcaas) is unsatisfactory as a schema for action explanation.

It should be kept in mind that my treatment of the LCA deviates from the true adherents attitude. Most LCA adherents do hold that (AE^lcaas) represents a satisfactory model for action explanation. By treating the LCA solely as a destructive but sound criticism of Hempel’s view, thereby disavowing any of its constructive claims, I abstain from the usual LCA tenets.

I begin by showing that DIF-2 is a pseudo-difference (Section 3). I argue how according to Hempel’s own assumptions (II.18) transforms into an empty and untestable phrase, thus explaining how Hempel violates his own condition (CA3). Then, in Section 4, I show that the violation of (CA3) by the untestability of (II.18) is one side of a coin which has the violation of (CA2) by the analyticity of (II.17) as its other side. Thus I demonstrate that DIF-1 is a pseudo-difference too. Section 5 is devoted to three objections that might be brought against my argument. Finally, in Section 6, I shall explain how the discussion between Hempelians and LCA adherents on the explanation of action has brought us back to the age-old problem of *akrasia*.

---

30 Of course, the analytical character of (II.17) also violates (CA3). This violation, however, is not very important for my argument, since I wish to focus on the connection between the infringement of (CA3) by (II.18) and that of (CA2) by (II.17). In particular, I will argue that both violations branch from the same trunk, or, as I shall express this, are two sides of the same coin.
When do we call someone a rational agent? What does it mean to say that a person $P$ acts rationally at time $t$? As we have seen in the previous chapter, Hempel deems it to mean that $P$'s action at $t$ "suits" his reasons at $t$. This means that $P$ has the disposition to follow a rule of which decision theory says that it applies to $P$’s wants and beliefs at $t$. I have already stated that here, in Part Two, I take decision theory for granted. Here I accede to Hempel’s belief that decision theory might be an effective mathematical representation of rational acting. For my objections are directed toward another part of Hempel’s analysis.

According to Hempel, the ability to act rationally, i.e. to follow decision theoretical rules, is not an ordinary disposition. It is a disposition of a special kind, namely a broad higher-order-disposition. As I shall argue below, it is precisely this idea that causes the problems for Hempel’s model: regarding rationality as a broad higher-order-disposition turns the rationality assumption (II.18) into a tautology. That the rationality assumption might be empty has been suggested by Dennett in several writings (Dennett 1969; Dennett 1971; Dennett 1975). Also, Donald Davidson hinted at the point:

"Hempel says rationality is a kind of character trait: some people have it and some don’t, and it may come and go in the same individual. No doubt some people are more rational than others, and all of us have our bad moments. ... But reference to such a trait does not seem to me to provide the generality for reason explanations that Hempel wants. For in the sense in which rationality is a trait that comes and goes, it can’t be an assumption needed for every reason explanation. People who don’t have the trait are still agents, have reasons and motives, and act on them ... So being in a position to call a person rational, irrational, or nonrational in this sense presupposes that we have already found it possible to give reason explanations of his actions ... At this point the assumption of rationality seems in danger of losing content." (Davidson 1976, 266ff).
In other words, Davidson suggests that reason explanations such as Hempel’s (AE$^{xy}$) tacitly presuppose the necessary truth of a rationality assumption such as (II.18). The mere possibility that this rationality assumption may be false, Davidson argues, renders these reason explanations paradoxical and causes the whole enterprise of explaining actions out of reasons to collapse like a house of cards. In later writings Davidson has specified and expanded this intuition, tracing it back to Quine’s indeterminacy of translation and from there indirectly to Brentano’s irreducibility of the mental (Davidson 1982, 1985, 1987; Quine 1960, 221).

Stimulating though these studies may be, they nowhere explain how precisely Hempel’s analysis renders (II.18) untestable. To see how it does, I shall scrutinise Hempel’s opinion that ‘being a rational agent’ expresses a broad higher-order-disposition. Since Hempel’s analysis is based on Carnap’s study of disposition concepts, I start by recollecting the relevant issues of Carnap (1936-1937) (3.1). Subsequently, I shall examine how Hempel uses these issues in his analysis of dispositions (3.2), broad dispositions (3.3) and higher-order dispositions (3.4). Section 3.5 consists in an excursion into the phenomenon that is called epistemic interdependence. In Section 3.6 some conclusions are drawn concerning the higher-order disposition of being rational.

### 3.1 Carnap’s reduction sentences

During the early stages of his concern with cognitive significance, Carnap held the view that a language $L$ is meaningful only if every non-logical or descriptive term of $L$ can be explicitly defined in terms referring to either immediate phenomenal experience (Carnap 1928) or physical objects (Carnap 1932). By an explicit definition of a descriptive predicate $Q$ with one argument Carnap understands a sentence of the form (D):

$$(D): \{(x) \ Q(x) \leftrightarrow \ldots x\ldots\}$$

where ‘$\ldots x\ldots$’ represents a sentential function with $x$ as its only free variable. The descriptive symbols occurring in ‘$\ldots x\ldots$’ are predicates which refer either to perceptions or to observable objects, substances, or events. Moreover, the explicit definition of a predicate with more than one argument has a form which is similar to (D).
The analyticity of the rationality assumption

In ‘Testability and Meaning’, however, Carnap distances himself from his previous position (Carnap 1936-1937). He now no longer believes that the cognitive significance of $L$ requires the explicit definability of every descriptive $L$-term. For $L$ may contain descriptive terms which are clearly meaningful but nevertheless escape definition as in (D): the notorious disposition predicates like ‘smellable’, ‘visible’, ‘brittle’. Carnap’s argument to the effect that these predicates resist explicit definition is tailored to one-place predicates, but it can easily be extended to predicates with more than one argument, or even to functions representing quantitative characteristics. It goes as follows.

Suppose we want to define the disposition predicate ‘$Q_3$’ meaning ‘is soluble in water’. Suppose further that we already have predicates ‘$Q_1$’ and ‘$Q_2$’ such that ‘$Q_1(x,t)$’ means ‘the substance $x$ is placed in water at time $t$’ and ‘$Q_2(x,t)$’ means ‘$x$ dissolves at $t$’. Then (D) indicates that ‘$Q_3$’ can be defined as in (D):

\[
(D'): (x) \{Q_3(x) \leftrightarrow (t)(Q_1(x,t) \rightarrow Q_2(x,t))\},
\]

in words: ‘$x$ is soluble in water’ means ‘if $x$ is put into water, then $x$ dissolves’ (Carnap 1936-1937, 440). But on (D’) ‘$Q_3(x)$’ is already true if ‘$Q_1(x,t)$’ is false for all $t$; ergo, (D’) says that objects which have never been placed in water for this reason are soluble. Since that is not what we mean when we attribute solubility to something, Carnap gives up the idea that every non-logical term of $L$ can be defined in explicit definitions like (D).

It has often been argued that the problem just described stems from the way in which the conditional in (D’) is specified. Explicit definitions like (D’) are phrased in extensional, truth-functional logic where conditionals are specified as material implications: they are true when their antecedent is false or when their consequent is true. A manner of avoiding these infamous paradoxes of material implication would be to use not only truth-functional connectives, but also modalities, especially conditionals in which causal connections can be expressed. So far, however, this line has proved to be exceedingly resistant. Moreover, Carnap’s evasion of the paradoxes is different. He tries to insert disposition terms in extensional logic, not by

---

31 As I explained in the footnote belonging to Section 2, Carnap uses one single time-variable, $t$. 

79
explicit definitions, but by *reduction pairs*.

A reduction pair (RP) introducing a one-place predicate like ‘\(Q_3\)’ is a conjunction of two so-called reduction sentences having the form \((r_1)\) and \((r_2)\):

\[
(r_1): \, (x) \, (t) \, [Q_1(x,t) \rightarrow (Q_2(x,t) \rightarrow Q_3(x))] \\
(RP) \\
(r_2): \, (x) \, (t) \, [Q_4(x,t) \rightarrow (Q_5(x,t) \rightarrow \neg Q_3(x))].
\]

Here ‘\(Q_1\)’ and ‘\(Q_4\)’ describe experimental conditions which must be realised in order to discover whether or not \(Q_3\) applies to a certain substance; ‘\(Q_2\)’ and ‘\(Q_5\)’ describe the results of these experiments.

(RP) is called a reduction pair because by sentences \((r_1)\) and \((r_2)\) the meaning of ‘\(Q_3\)’ is reduced to the meaning of the four predicates ‘\(Q_1\)’, ‘\(Q_2\)’, ‘\(Q_4\)’ and ‘\(Q_5\)’. It is assumed that those predicates either belong to the vocabulary \(V\) of the language \(L\) in which we want to introduce ‘\(Q_3\)’ or are previously introduced into \(L\) by other reduction pairs which ultimately use only the vocabulary \(V\). A further assumption is that there exists at least one object to which either the properties \(Q_1\) and \(Q_2\) or the properties \(Q_4\) and \(Q_5\) are rightly attributed, in other words, that

\[
(\exists x) \, (\exists t) \, [(Q_1(x,t) \land Q_2(x,t)) \lor (Q_4(x,t) \land Q_5(x,t))]
\]

holds. For if the latter sentence were false, (RP) would specify neither ‘\(Q_3\)’ nor ‘\(\neg Q_3\)’ for any object, hence would not specify the meaning of ‘\(Q_3\)’ at all, and therefore could not be said to introduce or determine the meaning of ‘\(Q_3\)’.

As is readily seen, definition of ‘\(Q_3\)’ by (RP) avoids the problem attached to the introduction of ‘\(Q_3\)’ through (D’): given (RP), the falsity of ‘\(Q_1(x,t)\)’ for all \(t\) does not entail the truth of ‘\(Q_3(x)\)’.

There are special instances of (RP), viz. the instance in which \(Q_4\) is the same property as \(Q_1\) and \(Q_5\) equals ‘\(\neg Q_2\)’. In that case (RP) takes the form (RP’):

\[
(r_1'): \, (x) \, (t) \, [Q_1(x,t) \rightarrow (Q_2(x,t) \rightarrow Q_3(x))] \\
(RP') \\
(r_2'): \, (x) \, (t) \, [Q_1(x,t) \rightarrow (\neg Q_2(x,t) \rightarrow \neg Q_3(x))].
\]
The analyticity of the rationality assumption

Since \((r')_2\) is equivalent to

\[(x) (t) \{Q_3(x) \rightarrow (Q_2(x,t) \rightarrow Q_2(x))\},\]

\((RP')\) can be replaced by \((RB)\):

\[(RB): (x) (t) \{Q_3(x) \rightarrow (Q_2(x,t) \leftrightarrow Q_3(x))\}.\]

\((RB)\) is the illustrious bilateral reduction sentence for the term ‘\(Q_3\)’: reduction sentence because the meaning of ‘\(Q_3\)’ is reduced to the meanings of ‘\(Q_1\)’ and ‘\(Q_2\)’, and bilateral because under the condition that ‘\(Q_1\)’, ‘\(Q_2\)’ implies ‘\(Q_3\)’ and vice versa. As in \((RP)\), ‘\(Q_1\)’ and ‘\(Q_2\)’ either belong to the vocabulary \(V\) of \(L\) or are previously introduced into \(L\) by sentences which ultimately use only the vocabulary \(V\); and as in \((RP)\), it is tacitly assumed that the sentence

\[(∃x) (∃t) Q_3(x,t)\]

holds. For if it were false, \((RB)\) would specify neither ‘\(Q_3\)’ nor ‘\(\neg Q_3\)’ for any object, and hence would not introduce or determine the meaning of ‘\(Q_3\)’ at all.

In an even more special case, viz. when ‘\((x) (t) Q_3(x,t)\)’ is universally satisfied, an explicit definition sentence \((D'')\) becomes derivable:

\[(D''): (x) \{Q_3(x) \leftrightarrow (t) Q_2(x,t)\}].\]

Generally speaking, and leaving out operators and arguments for a moment, the difference between introduction by definition and introduction by reduction can be stated in the following way. A definition sentence for a term ‘\(Q\)’ describes a necessary and sufficient condition \(C\) for application of ‘\(Q\)’; it thus enables us to substitute the term ‘\(C\)’ for ‘\(Q\)’ in any sentence whatsoever, since the totality of necessary and sufficient conditions for \(Q\) coincides with \(C\). For example, since in \((D'')\) the necessary and sufficient condition for \(Q_3\) is \(Q_2\), \((D'')\) permits the substitution of ‘\(Q_2\)’ for ‘\(Q_3\)’ in all contexts.

On the other hand, a pair of reduction sentences describes a necessary condition, \(C_1\), and a sufficient condition, \(C_2\), for application of \(Q\),
but $C_i$ and $C_j$ do not coincide.\textsuperscript{32} Therefore a reduction pair does not permit the substitution of the terms ‘$C_i$’ and/or ‘$C$’ for ‘$Q$’. Applied to our pair (RP): (RP) provides both a necessary condition and a sufficient condition for $Q_i$. The sufficient condition is stated in $(r_1)$, which says that ‘$Q_i \land Q_j$’ is sufficient for application of ‘$Q_3$’. The necessary condition is expressed in $(r_2)$, for $(r_2)$ is equivalent to

$$Q_3 \rightarrow \neg(Q_4 \land Q_5),$$

and hence states that ‘$\neg(Q_4 \land Q_5)$’ is necessary for application of ‘$Q_3$’. In other words, according to (RP) the extension of ‘$Q_3$’ contains the extension of ‘$Q_4 \land Q_5$’ and is contained by that of ‘$\neg(Q_4 \land Q_5)$’. Since by assumption the extensions of ‘$Q_4 \land Q_5$’ and ‘$\neg(Q_4 \land Q_5)$’ differ, we may attribute ‘$Q_3$’ to instances of ‘$Q_4 \land Q_5$’ whereas ‘$\neg Q_3$’ may be ascribed to cases of ‘$Q_4 \land Q_5$’. However, the important message is that we stay in the dark about the remaining cases, to wit cases of

$$\neg((Q_4 \land Q_5) \lor (Q_4 \land Q_5)), $$

for (RP) does not tell us whether we should apply the term ‘$Q_3$’ or ‘$\neg Q_3$’ to them. Therefore, and contrary to definition sentences, (RP) is insufficient to render the term ‘$Q_i$’ eliminable from all contexts in which it may occur.

The above-mentioned difference between reduction and definition is usually stated by saying that a definition sentence for ‘$Q$’ specifies the meaning of ‘$Q$’ completely or exhaustively, whereas a reduction pair specifies it only incompletely or partially. In general, reduction pairs cannot be replaced by definition sentences. Only if we succeed in changing and extending a reduction pair to the point where it determines all cases, may we formulate a definition as a substitute for it.

Let us now examine how Hempel uses these Carnapian thoughts on complete and partial definition when he characterises the property of ‘being a rational agent’ as a broad higher-order disposition.

\textsuperscript{32} At this point as well as in the sequel I use subscripts to refer to necessary, and superscripts to refer to sufficient conditions.
The analyticity of the rationality assumption

3.2 Dispositions

According to Hempel, having a belief, having a desire, or acting rationally are all dispositions, i.e. tendencies to behave in a certain way under certain circumstances. Entirely in the manner of Carnap, Hempel partially defines disposition terms in bilateral reduction sentences. That is to say, if ‘M’ is an expression referring to a disposition, then ‘M’ is defined by the bilateral reduction sentence (III.5):

(III.5): (x) \{S(x) \rightarrow (M(x) \leftrightarrow O(x))\}

where ‘x’ ranges over objects, and ‘S’ and ‘O’ denote observable characteristics (I shall follow Hempel in further omitting the time argument t). Again, (III.5) is called a reduction sentence because it reduces the meaning of ‘M’ to expressions referring to observable qualities (‘S’ and ‘O’); and it is called bilateral because, under the assumption ‘S(x)’, ‘M(x)’ implies ‘O(x)’ and vice versa. At this point, three observations are of importance. I shall first list them; in the sequel it will be explained how these observations culminate in the conclusion that ‘being a rational agent’ is an untestable concept.

OBSERVATION 1. By partial definition (III.5), the meaning of ‘M’ is determined only for objects that have property S; for objects which lack S, it cannot be decided whether or not they possess M. This is also expressed by saying that (III.5) represents a conditional definition. In Hempel’s words:

"[(III.5)] specifies that in all cases where the observable characteristic [S] is present, [M] is applicable if and only if the observable characteristic [O] is present as well. In fact, [(III.5)] is an instance of those ... conditional definitions which Carnap calls bilateral reduction sentences (Hempel 1965, 114)."

Since (III.5) is a conditional definition, we must, in order to apply it to a particular x, assume that its antecedent ‘S(x)’ is true. Obviously, under that assumption (III.5) entails that ‘M(x)’ and ‘O(x)’ have equal truth values: either they are both true or they are both false. In the former case, x has the property M, whereas in the latter case it does not.

OBSERVATION 2. Sentence (III.5) is equivalent to the conjunction
of two sentences (III.6) and (III.7):

(III.6): (x) \{M(x) \rightarrow (S(x) \rightarrow O(x))\}
(III.7): (x) \{S(x) \rightarrow (O(x) \rightarrow M(x))\}.

Hempel calls (III.6) and (III.7) *symptom sentences* since they express symptoms of M’s presence (Hempel 1965, 460). Of course, this pair of Hempelian symptom sentences corresponds to the pair (RP’) of Carnapian reduction sentences. For (III.7) is similar to (r’1), whereas (III.6) is equivalent to (III.8):

(III.8): (x) \{S(x) \rightarrow (\neg O(x) \rightarrow \neg M(x))\}

and hence is similar to (r’2).

**OBSERVATION 3.** Sentence (III.6) expresses a *necessary* condition for application of M: it entails that ‘S(x) \rightarrow O(x)’ and hence ‘\neg(S(x) \land \neg O(x))’ is necessary for ‘M(x)’. Sentence (III.7) expresses a *sufficient* condition for application of the term ‘M’: it says that ‘S(x) \land O(x)’ is sufficient for ‘M(x)’. As indicated earlier, if all the necessary and the sufficient conditions for M coincide in one condition C, then the term denoting this condition, ‘C’, may be substituted for ‘M’ in all contexts where ‘M’ occurs. Strictly speaking, the necessary condition stated in ‘\neg(S(x) \land \neg O(x))’ does not coincide with the sufficient condition stated in ‘S(x) \land O(x)’, since the falsity of ‘S(x)’ renders the first formula true whereas it makes the second one false. However, given the fact that (III.5) is a *conditional definition*, the extensions of ‘\neg(S(x) \land \neg O(x))’ and ‘S(x) \land O(x)’ do coincide, whenever (III.5) is applied. For as we have seen in **OBSERVATION 1**, this assumption implies that ‘S(x)’ is true when the definition is applied to x, and thus that ‘\neg(S(x) \land \neg O(x))’ and ‘S(x) \land O(x)’ have the same extension: both are true if ‘O(x)’ and both are false if ‘\neg O(x)’.

The three observations above yield the following conclusion: if a disposition term ‘M’ is partially defined in exactly one bilateral reduction sentence (III.5), then there exists only one necessary condition and only one sufficient condition for M. The former is stated in (III.6), the latter in (III.7). Moreover, since (III.5) is also a conditional definition, the necessary condition and the sufficient condition coincide over the area of application.

However, most dispositions are defined in several bilateral reduction
sentences; they thus pertain to several necessary and sufficient conditions which do not coincide. This will be explained in Section 3.3.

3.3 Broad dispositions

Some disposition terms are defined by one single bilateral reduction sentence. ‘Being soluble in water’ is a case in point. We might define this term in a unique bilateral reduction sentence, which states that if a substance is put in water, then it is water-soluble if and only if it dissolves. (To be accurate, we might stipulate that the water be pure, and that the quantity of the substance added be inferior to the amount required to produce saturation.) Following Ryle, who talks about "simple, single-track dispositions", I take such disposition terms to denote *simple* dispositions.\(^{33}\) Clearly, the ‘M’ out of the previous section refers to a simple disposition.

However, most dispositions are not simple, since they manifest themselves in a variety of ways, depending on the circumstances. To mention again an already over-worked example: being magnetised is not a simple disposition since the magnetisation of, for instance, an iron bar can be detected by looking whether iron filings will cling to its ends, but also by watching whether one end attracts the north pole and the other the south pole of a compass needle, or by trying to find out whether, if the bar is broken in two, each of the parts exhibits the two tendencies just described for the whole bar. Dispositions like these are called *multiple* (by Carnap) or *broad* (by Hempel).\(^{34}\)

---

\(^{33}\) "In discussing dispositions it is initially helpful to fasten on the simplest models, such as brittleness of glass or the smoking habit of a man. For in describing these dispositions it is easy to unpack the hypothetical proposition implicitly conveyed in the ascription of the dispositional properties. To be brittle is just to be bound or likely to fly into fragments in such and such conditions; to be a smoker is just to be bound or likely to fill, light and draw on a pipe in such and such conditions. These are simple, single-track dispositions, the actualizations of which are nearly uniform." (Ryle 1949, 43).

\(^{34}\) Ryle calls them *higher-grade* dispositions:

"There are many dispositions the actualizations of which can take a wide and perhaps unlimited variety of shapes ... the higher-grade dispositions of people ... are, in general, not single-track dispositions, but dispositions the exercises
Often the difference between broad and simple dispositions is a matter of description. Many dispositions can be described as being either simple or broad. Fragility may serve as an example. We might describe fragility as a simple disposition; if we do that, we will say something like "An object, \( x \), is fragile if \( x \) breaks under a relatively weak blow". However, we can also describe it as a broad disposition. In that case we will specify both the expressions "to break" and "a relatively weak blow". Accordingly, we will distinguish between different ways in which \( x \) can break, and between different ways in which \( x \) can be struck by weak blows. I come back to this point in Chapter IX.

A term ‘\( T \)’ denoting a broadly dispositional trait cannot be defined in one single bilateral reduction sentence. Instead, the definition of ‘\( T \)’ consists in a set \( U \) of \( n \) different bilateral reduction sentences \( (u-1), \ldots, (u-n) \), each of which describes a specific way in which \( T \) manifests itself. Let ‘\( S-1 \)’, \ldots, ‘\( S-n \)’ describe \( n \) different situations \( S-1, \ldots, S-n \) and let ‘\( O-1 \)’, \ldots, ‘\( O-n \)’ describe \( n \) reactions corresponding to \( S-1, \ldots, S-n \). Then the partial definition of \( T \) is the set \( U \):

of which are indefinitely heterogeneous. When Jane Austen wished to show the specific kind of pride which characterized the heroine of *Pride and Prejudice*, she had to represent her actions, words, thoughts, and feelings in a thousand different situations. There is no standard type of action or reaction such that Jane Austen could say ‘My heroine’s kind of pride was just the tendency to do this, whenever a situation of that sort arose.’” (Ryle 1947, 43-44).

I myself shall avoid the term ‘higher-grade disposition’. Using this term could easily cause confusion with Hempel’s usage of ‘higher-order-disposition’. As I explain in the text, the latter term denotes a broad disposition of second order. Cf. Føllesdal, who calls a higher-order-disposition a "second-order" disposition (Føllesdal 1986).
The analyticity of the rationality assumption

\begin{align*}
(u-1): \ & (x) \{S-1(x) \rightarrow (O-1(x) \leftrightarrow T(x))\} \\
(u-2): \ & (x) \{S-2(x) \rightarrow (O-2(x) \leftrightarrow T(x))\} \\
\vdots \\
U \\
\vdots \\
(u-n): \ & (x) \{S-n(x) \rightarrow (O-n(x) \leftrightarrow T(x))\} \\
\end{align*}

(Hempel 1965, 113ff, 459ff, 472ff; cf. Carnap 1936-1937, 450ff). Every \((u-i)\) \((1 \leq i \leq n)\) is equivalent to the conjunction of a sentence \((u_i)\) and a sentence \((u^i)\):

\begin{align*}
(u_i): \ & (x) \{T(x) \rightarrow (S_i(x) \rightarrow O_i(x))\} \\
(u^i): \ & (x) \{S_i(x) \rightarrow (O_i(x) \rightarrow T(x))\},
\end{align*}

where \((u_i)\) expresses a necessary condition, \(C_i\), and \((u^i)\) expresses a sufficient condition, \(C_i\), for application of \('T'\), \(C_i\) being described by \('S_i(x) \rightarrow O_i(x)'\) and \(C_i'\) by \('S_i(x) \land O_i(x)'\). 35

There exists an interesting difference between simple dispositions and broad dispositions. Consider, for example, the disposition term \('M'\) mentioned in Section 3.2. \('M'\) denotes a simple disposition \(M\), for its definition contains one and only one bilateral reduction sentence, viz. (III.5). As we have seen in OBSERVATION 1, (III.5) is a conditional definition, which, see OBSERVATION 2, truth-functionally equals the conjunction of (III.6) and (III.7). (III.6) states the only necessary condition and (III.7) expresses the only sufficient condition for \(M\), and according to OBSERVATION 3 this implies that the two conditions coincide in the domain of application. But if they do so, then, see Section 3.1, all occurrences of \('M'\) may be replaced by \('S'\) and \('O'\), and this boils down to

35 Recall that I use subscripts to refer to necessary, and superscripts to refer to sufficient conditions: \(C_i\) and \(C_i'\). When a situation \(S\) is part of a necessary condition, I shall attach to it a subscript: \(S_i\); when \(S\) is part of a sufficient condition, a superscript is added: \(S_i'\). I will stick to this convention, regardless of the fact that \(S_i\) and \(S_i'\) are situations rather than conditions, and disregarding that they may refer to the same situation, in which case \(S_i=S_i'\). The same goes \textit{mutatis mutandis} for \(O\), \(O_i\), and \(O_i'\).
saying that (III.5) is, within its domain of application, an analytical sentence.36

While simple dispositions are defined in analytical sentences, broad dispositions are not. A broad disposition is defined by a set of bilateral reduction sentences which jointly have empirical implications. Consider again the term ‘$T$’. ‘$T$’ is defined by $U$, the members of which together entail that any $x$ which satisfies the sufficient condition implied by sentence (u-i), also satisfies the necessary condition implied by sentence (u-j) ($i \neq j$). In other words, the members of $U$ jointly entail the sentence ($u^i u_j$):

$$(u^i u_j): (x) \left\{ (S'(x) \rightarrow O'(x) \rightarrow T(x)) \land (T(x) \rightarrow (S_j(x) \rightarrow O_j(x))) \right\}.$$  

This sentence implies ($u^i u_j^*$):

$$(u^i u_j^*): (x) \left\{ (S'(x) \land O'(x)) \rightarrow (S_j(x) \rightarrow O_j(x)) \right\}$$

from which ‘$T$’ has disappeared. The sentences ($u^i u_j$) and ($u^i u_j^*$) are not analytical truths, but statements that have an empirical character. This can be seen if we take ‘$T$’ to mean ‘is magnetised’ and let $x$ range over bars. In that case ($u^i u_j^*$) might be read as: any bar that satisfies the iron filings condition, also satisfies the compass needle condition. Clearly, the latter statement is not an analytical truth, but an empirical one. Therefore, the sentences of $U$ count as empirical statements, whereas single bilateral reduction sentences may be regarded as analytical statements (within their domain of application).

According to Hempel the property of being rational is a broad disposition. Notwithstanding the difference between broad and simple dispositions, I shall, however, treat the property of being rational as a simple disposition. Such an approach does not seem to interfere with the procedure that Carnap and Hempel have in mind. For as Carnap has explained, broad dispositions "can still be expressed by a conjunction of simple dispositions" (Carnap 1956, 64). Therefore, if it turns out that Hempel’s analysis fails in the simple variant, it most likely will do so in the more complex versions.

36 Carnap already noticed the analytical character of sentences like (III.5): "... a bilateral reduction sentence ... has no factual content” (Carnap 1936-1937, 444; cf. 451ff). Hempel also alluded to it (Hempel 1965, 459-460).
as well. (See also the first objection in Section 5, below.)

3.4 Higher-order dispositions

In Hempel’s view, the property of being rational (from now on: \( R \)) is not only a broad disposition, but also a higher-order disposition, i.e. a disposition which manifests itself in response to (lower-order) dispositions. In the case of \( R \), these (lower-order) dispositions are beliefs and desires:

"... the dispositions implied by attributing rationality to a person are higher order dispositions; for the beliefs and ends-in-view in response to which, as it were, a rational agent acts in a characteristic way are not manifest external stimuli but rather, in turn, broadly dispositional features of the agent" (Hempel 1965, 473 - emphasis by Hempel).

In conformity with my decision to regard \( R \) as denoting a simple rather than a broad disposition, I shall take \( R \) as referring to one belief (\( B \)) and one desire (\( D \)). \( B \) and \( D \) are normal, lower-order dispositions. Thus it seems that we might define the terms ‘\( B \)’ and ‘\( D \)’ respectively in the bilateral reduction sentences (III.9) and (III.10) respectively:

\[
\text{(III.9): } (x) \{ S(x) \rightarrow (B(x) \leftrightarrow O(x)) \}
\]
\[
\text{(III.10): } (x) \{ S(x) \rightarrow (D(x) \leftrightarrow O(x)) \}
\]

\[37\] Carnap’s belief that broad dispositions can be expressed by a conjunction of simple ones has the paradoxical consequence that a conjunction of analytic sentences may have empirical content. Hempel and Pap drew attention to this paradox (Pap 1958a, 335; Pap 1963; Hempel 1963). In turn, Carnap tried to evade this paradox by dividing terms that denote broad dispositions into two classes: the pure disposition terms and the theoretical terms. The former are expressible by conjunctions of simple dispositions but lack empirical content, whereas the latter have empirical content but cannot be expressed as such conjunctions (Carnap 1956; Carnap 1963b). However, this whole discussion appears to be irrelevant to my point which is that Hempel’s use of ‘is rational’ deprives this term of any content whatsoever, no matter how it is interpreted. Carnap’s distinction between pure disposition terms and theoretical terms is discussed in Chapter VIII, Section 2.4.
where \( x \) ranges over persons, ‘\( S(x) \)’ means ‘\( x \) is in such and such an observable situation’, and ‘\( O(x) \)’ means ‘\( x \) performs such and such an action’. 38

However, (III.9) and (III.10) do not really do justice to Hempel’s argument. For according to Hempel \( B \) and \( D \) are epistemically interdependent. In order to explain how that disqualifies (III.9) and (III.10) as definitions of the terms ‘\( B \)’ and ‘\( D \)’, let us make a brief excursion into the field of epistemic interdependence.

3.5 Excursion: epistemic interdependence

What does it mean to say that there exists an "epistemic interdependence of belief attributions and goal attributions" (Hempel 1965, 475)? More particularly, how does the epistemic interdependence of \( B \) and \( D \) affect the definitions of ‘\( B \)’ and ‘\( D \)’? As it turns out, Hempel answers this question in two different ways.

Suppose that we attribute to a person \( P \) the belief that it is raining outside. Does this attribution imply anything concerning \( P \)’s observable actions? Does it for instance imply that \( P \) will take an umbrella when he goes out? Obviously not. The mentioned attribution implies that \( P \) will take an umbrella only on suitable assumptions about \( P \)’s desires, such as that he wants to protect himself from getting drenched. Of course, the same holds for attributions of desires. The statement that \( P \) wants a drink of wine implies \( P \)’s drinking a certain liquid only on the assumption that \( P \) believes this liquid to be wine. So it appears that an "epistemic interdependence of belief attributions and goal attributions" means that hypotheses about beliefs are testable only when taken in conjunction with hypotheses about desires, and vice versa. It looks as if Hempel has something like that in mind when he writes:

"... ascription of an objective ... has implications concerning characteristic overt behavior only when taken in conjunction with ascriptions of appropriate beliefs."

38 Here \( B \) and \( D \) are of course single dispositions, not, as in Chapter II (Section 3.1), sets of dispositions.
Similarly, in our earlier example, the hypothesis that \( P \) believes the streets are slushy implies the occurrence of characteristic overt behavior only when taken *in conjunction* with suitable hypotheses about \( P \)'s objectives.

Indeed, it seems that a hypothesis about an agent’s objectives generally can be taken to imply the occurrence of specific overt action only when conjoined with appropriate hypotheses about his beliefs, and vice versa ...

That is, belief attributions and goal attributions are epistemically interdependent" (Hempel 1965, 475 - my emphasis).

This quotation indicates why (III.9) and (III.10) are inappropriate as definitions of ‘\( B \)’ and ‘\( D \)’: (III.9) and (III.10) wrongly presuppose that assumptions about beliefs and desires can be defined without taking them in conjunction. Furthermore, the quotation suggests what in fact might count as a decent (partial and/or conditional) definition of ‘\( B \)’ as well as of ‘\( D \)’, to wit:

\[
(\text{III.11}): (x) \{ S(x) \rightarrow (B(x) \land D(x)) \leftrightarrow O(x)) \}.
\]

However, Hempel’s discussion of epistemic interdependence also suggests definitions different from (III.11). This takes us to the second way in which Hempel answers the question to what extent the epistemic interdependence of \( B \) and \( D \) affects the definitions of ‘\( B \)’ and ‘\( D \)’.

Hempel explicitly denies that epistemic interdependence renders it impossible to ascertain a person’s beliefs or desires. His main reason is that:

"...often we have good *antecedent information* about one of the inter[de]pendent items, and then a hypothesis about the other may be tested by ascertaining how the person acts in certain situations. For example, if we have good grounds for the assumption that our man is subjectively honest, that he endeavors to ‘tell the truth’, then his answers to our questions may afford a reliable indication of his beliefs. Conversely, we are often able to test a hypothesis about a person’s objectives by examining his
behavior in certain critical situations because we have good reason to assume that he has certain relevant beliefs" (Hempel 1965, 475 - my emphasis).

Judging from this quotation, hypotheses about beliefs are testable, not in conjunction with, but only under the assumption that suitable hypotheses about desires are true, and vice versa. The quotation thus gives rise to definitions of ‘B’ and ‘D’ which differ slightly from definition (III.11). For it suggests that ‘B’ is properly defined in:

\[
(\text{III.12}): (x) \left( S(x) \rightarrow \{ D(x) \rightarrow (B(x) \leftrightarrow O(x)) \} \right)
\]

and that (III.13) is a suitable definition of ‘D’:

\[
(\text{III.13}): (x) \left( S(x) \rightarrow \{ B(x) \rightarrow (D(x) \leftrightarrow O(x)) \} \right)
\]

The difference between (III.12) and (III.13) on the one hand and (III.11) on the other is evident: if both ‘S(x)’ and ‘O(x)’ are true when both ‘B(x)’ and ‘D(x)’ are false, then the corresponding instance of (III.12) is true, as is that of (III.13), but that of (III.11) is false.

As is readily seen, (III.12) as well as (III.13) exclude that ‘S(x)’, ‘D(x)’ and ‘B(x)’ are true whereas ‘O(x)’ is false: so if ‘O(x)’ is false, then ‘S(x)’ or ‘B(x)’ or ‘D(x)’ is false. Now ‘S(x)’ is in the domain of application not allowed to be false because (III.12) and (III.13) are conditional definitions. As we have seen in Section 3.2., OBSERVATION 1, this means that they define ‘B’ and ‘D’ solely under the condition that x has the property S. So ‘B(x)’ or ‘D(x)’ is false. In other words, if the universal quantifier is restricted to the domain of application, then (III.12) and (III.13) imply (III.14):

\[
(\text{III.14}): (x) \left( \neg O(x) \rightarrow (\neg B(x) \lor \neg D(x)) \right)
\]

Evidently, with the same restriction (III.14) is also implied by (III.11). For (III.11) too is a conditional definition; it too defines ‘B’ as well as ‘D’ solely under the condition that ‘S(x)’ is true. So if ‘O(x)’ is false, according to (III.11) ‘B(x)’ or ‘D(x)’ must be false.

To summarise, Hempel’s characterisation of the epistemic interdependence of B and D indicates two different ways of defining ‘B’ and
‘D’: according to the first way ‘B’ is defined in (III.11) and so is ‘D’, according to the second way ‘B’ is defined in (III.12) and ‘D’ is defined in (III.13). The crucial point is that either of the two ways implies (III.14). In order to understand why this point is essential, let us return to the discussion of higher-order dispositions.

3.6 Higher-order dispositions again

As we have seen in Section 3.4, the higher-order disposition term ‘R’ is defined by making an appeal to the lower-order disposition terms ‘B’ and ‘D’. Hence we might specify ‘R’ in the following bilateral reduction sentence:

\[(III.15)\]: \[(x) \{(S(x) \land B(x) \land D(x)) \rightarrow (R(x) \leftrightarrow O(x))\}\].

Since (III.15) is a conditional definition of ‘R’, we must in each context of application assume that its antecedent is true. Under that assumption, (III.15) entails that ‘R(x) ↔ O(x)’ and, accordingly, that ‘R(x)’ and ‘O(x)’ are either both true or both false. In the former case, x has the property R, whereas in the latter case it does not. However, the problem is that the latter case cannot possibly arise. This impossibility results from the way in which B and D are described. As has been explained in the previous section, Hempel indicates that there are two possible ways in which ‘B’ and ‘D’ might be defined: either ‘B’ as well as ‘D’ are defined in (III.11) or ‘B’ is defined in (III.12) and ‘D’ is defined in (III.13). Each of the two ways implies (III.14), and hence entails that ‘B(x)’ or ‘D(x)’ is false if ‘O(x)’ is false. Therefore, each of them excludes that the antecedent of (III.15) is true while ‘O(x)’ is false. Consequently, assuming (III.15) to hold, each of the two excludes that the antecedent of (III.15) is true while ‘R(x)’ is false.

From this it follows that, on (III.15), every agent in the domain of application is a rational agent. This of course deprives the property R of any empirical content whatsoever. As a result, Hempel’s rationality assumption (II.18) in schema (AE\(^{B}\)) becomes untestable, whereby adequacy condition (CA3) is violated.
III: The Logical Connection Argument

4. The analyticity of the empirical law assumption or the pseudo-character of DIF-1

As has been stated above, the LCA claims that statement (III.1) in (AE\text{\textasciicircum}lca) is a tautology and consequently can be missed; as a result, we get (AE\text{\textasciicircum}lcaa). If this is a sound argument, then Hempel’s supposed empirical law (II.17) in (AE\text{\textasciicircum}hs) is not empirical but analytical in character. For the analyticity of (III.1) entails that P’s action analytically follows from his reasons, i.e. from (the descriptions of) his beliefs and desires. But then of course P’s action will still follow from his reasons if P is a rational agent as well - to say nothing of the nugatory nature of the latter property.

It is not my aim to adjudicate upon the validity or invalidity of the LCA at this juncture. Rather, I merely maintain that its validity is implied by Hempel’s very own argument. After all that has been said above, this point can without much effort be explained in a few words. For it will turn out that the analyticity of (III.1) and the untestability of (II.18) are two sides of the same coin, and thus that the violation of condition (CA2) goes hand in hand with the breach of (CA3). The pseudo-character of DIF-2 and that of DIF-1 thus appear to mirror each other.

Let us suppose that a person P has one reason, and hence one belief and one pro-attitude: he wants such and such to be the case (desire D-i) and he believes that performing action O brings about such and such (belief B-j). Then, on Hempel’s argument explained in Section 3.5 above, there are no more than two possibilities. Either both ‘B-j’ and ‘D-i’ are defined in (III.16):

\[(\text{III.16): } (x) \{S(x) \rightarrow ((B-j(x) \land D-i(x)) \leftrightarrow O(x))\}\]

or ‘B-j’ is defined in (III.17) whereas ‘D-i’ is defined in (III.18):

\[(\text{III.17): } (x) \{S(x) \rightarrow \{D-i(x) \rightarrow (B-j(x) \leftrightarrow O(x))\}\}\]

\[(\text{III.18): } (x) \{S(x) \rightarrow \{B-j(x) \rightarrow (D-i(x) \leftrightarrow O(x))\}\}\].

We have seen that (III.17) and (III.18) together imply (III.19):

\[(\text{III.19): } (x) \{\neg O(x) \rightarrow (\neg B-j(x) \lor \neg D-i(x))\}\]
and that (III.19) is also entailed by (III.16). Sentence (III.19) can be transformed into (III.20):

\[(x) \{(B-j(x) \land D-i(x)) \rightarrow O(x)\},\]

and (III.20) licenses deduction of the following form:

\[(B-j(a) \land D-i(a) = O(a))\]

where ‘\(a\)’ refers to a particular person, \(P\), and where ‘\(=\)’ denotes logical implication restricted to the conditional definition(s) of ‘\(B-j\)’ and ‘\(D-i\)’, i.e. to their domain of application. Now (III.21) precisely formulates the LCA. For (III.21) says that the performance of action \(O\) is deductively entailed by having the dispositions \(B-j\) and \(D-i\). In other words, the way in which Hempel on the basis of Carnap’s analysis partially and conditionally defines dispositions not only renders the rationality assumption untestable, but also leads to the LCA. As a result, not only is condition (CA3) broken, but condition (CA2) is violated as well.

5. Objections

My argument in the previous two sections seems open to three objections, which I shall now consider in turn. The first objection is that my argument follows from my decision to consider \(R\) as a simple disposition (instead of a broad one). For as opposed to sentences which jointly define broad dispositions, sentences which define simple dispositions do not have the character of empirical laws. Small wonder, then, that my analysis of the disposition \(R\) results in the conclusion that the term ‘\(R\)’ is an empty concept.

This objection, however, does not carry weight. This can be shown by comparing two sentences, each of which defines a simple disposition, namely

\[(x) \{S(x) \rightarrow (M(x) \leftrightarrow O(x))\}\] and
\[(x) \{(S(x) \land B(x) \land D(x)) \rightarrow (R(x) \leftrightarrow O(x))\}.

(III.5) implies that, if its antecedent is true, there are two possibilities:
(a): ‘M(x)’ and ‘O(x)’ are true
(b): ‘M(x)’ and ‘O(x)’ are false.

If (a) is the case, M is applicable to x; in the case of (b), M is not applicable to x. Like (III.5), (III.15) implies that, if its antecedent is true, there are two possibilities:

(a*): ‘R(x)’ and ‘O(x)’ are true
(b*): ‘R(x)’ and ‘O(x)’ are false.

But unlike possibility (b), possibility (b*) cannot occur. This means that my argument to the effect that R is empty does not follow from the assumption that R is a simple disposition.

I now turn to the second objection, which, in fact, is less an objection than a remark. As we have learnt from Hempel, B and D are epistemically interdependent. However, it might be argued that there exists another epistemic interdependence, which I have not taken into account, viz. the epistemic interdependence between B and D on the one hand, and R on the other (Hempel 1965, 476). Further, it might be remarked that my conclusion merely results from the latter epistemic interdependence.

In order to show that this is not a telling objection either, I shall compare (III.15) to (III.12), both of which are subject to epistemic interdependence. Clearly, (III.12) is equivalent to (III.12*):

(III.12*): (x) {((S(x) ∧ D(x)) → (B(x) ↔ O(x)))

which says that ‘B(x)’ and ‘O(x)’ are both true or both false if its antecedent is true. Likewise, as we have repeatedly seen, (III.15) implies that ‘R(x)’ and ‘O(x)’ are both true or both false if its antecedent is true. Nevertheless (III.15) and (III.12*) differ in an important respect. The falsity of ‘R(x)’ and ‘O(x)’ contradicts the assumption that the antecedent of (III.15) is true. On the other hand, the falsity of ‘B(x)’ and ‘O(x)’ by no means contradicts the assumption that the antecedent of (III.15) is true. In other words, while (III.12*) allows that ‘¬B(a)’ is the case, (III.15) excludes the possibility of ‘¬R(a)’. Ergo, the epistemic interdependence cannot by itself be the reason for my conclusion that R is devoid of meaning.

The third objection I borrow from Hempel. According to Hempel "there are various kinds of circumstances in which we might well retain our
Objections

assumptions about the agent’s beliefs and objectives and abandon instead the assumption of rationality” (Hempel 1965, 476). Examples of these "various kinds of circumstances" are situations in which a person "overlook[s] certain relevant items of information which he clearly believes to be true", or "overlook[s] certain aspects of the goal he is seeking to attain" (Hempel 1965, 476). As a result of these circumstances the sentence (III.22):

\[
S(a) \land B(a) \land D(a) \land \neg R(a) \land \neg O(a)
\]

might well be true. Hence, the claim that a definition of ‘B’ and ‘D’ implies (III.14) is false.

Evidently, this objection is not directed against (III.15), not against the way in which ‘R’ is specified. It is basically an objection against the two possible ways in which ‘B’ and ‘D’ might be defined. For it is by virtue of these ways that (III.14) is claimed to be true, and hence that (III.22) is said to be false. Consequently, if we define the agent’s beliefs and desires by either of these two ways, it is impossible to abandon the assumption of rationality. Yet, as we have seen, both ways are suggested by Hempel’s own argument, wherefore we must conclude that Hempel contradicts himself at this point.

6. The return of the akrasia problem

In the present chapter I have argued for the following three claims:

(1) The rationality assumption (II.18) in Hempel’s schema of action explanation (AEhs) cannot be tested by experiment or observation; to this extent, DIF-2 is a pseudo-difference.

(2) The supposed empirical law (II.17) in (AEhs) is in fact derivable from the analytical statement (III.1); to this extent DIF-1 is a pseudo-difference.

(3) The untestability of (II.18) and the analyticity of (II.17) branch from the same trunk; they both result from considering the properties of having a belief, having a desire, and acting rationally as specific dispositions. Hence, the pseudo-character of DIF-1 and that of DIF-2 mirror one another.

What follows from these claims? Not that all talk about rational action is insignificant and futile. Nor that all talk about rational acting as a disposition is meaningless either. It does not even follow that talking about acting rationally as a disposition which can be couched in bilateral reduction
sentences is senseless. However, a conclusion that can be garnered from our claims is that Hempel’s specific approach to what it means to be a rational agent deprives the very concept of any empirical meaning whatsoever. Starting from Hempel’s theory on action explanation, every action is a rational action and irrational actions cannot possibly occur.

This means that Hempel’s position in the debate on action explanation finally faces us with the old akrasia problem. As it did for the Greek philosophers, akratic behaviour constitutes a mystery for Hempel. No more than Socrates, Plato, Aristotle or the early Stoics is Hempel able to explain how such behaviour can occur.

However, not only does Hempel’s position run up against akrasia, but that of the LCA adherents stumbles upon it too. This can be explained in two simple ways. The first is based on the identification of Hempel’s schema for action explanation with that of the LCA adherents. For another conclusion to be gathered from our three claims is that (AElcaa) is the same as (AE\textsuperscript{hs}): if (II.17) and (II.18) indeed are tautologies, then (AE\textsuperscript{hs}) boils down to (AElcaa). Ergo, if the former confronts us with the akrasia problem, then so does the latter. The second way of explaining how the LCA position impinges the akrasia problem is based on the similarity between (AElcaa) on the one hand and Von Wright’s schema for action explanation, (AE\textsuperscript{wright}), on the other. I have stressed this similarity before; I even decided to regard (AElcaa) as a formalised version of (AE\textsuperscript{wright}) (see Section 2). But we have also seen that Von Wright, and for very good reasons, calls his (AE\textsuperscript{wright}) a practical syllogism, thereby paying homage to the Stagirian sage. It therefore should not come as a surprise that both (AE\textsuperscript{wright}) and (AElcaa) run against the same difficulty as that which jeopardises an Aristotelian practical syllogism. The akrasia problem, which constituted a vexing question for Aristotle and other Greek philosophers, thus presents itself again in the context of the contemporary debate about the explanation of actions. For at least two factions in the debate, the Hempelians and the LCA adherents, finally have to face the question again. In the next chapter we will learn that the same holds for the third faction in the debate, headed by Donald Davidson.

In retrospect, the return of the akrasia problem could almost have been predicted. It was so to speak written in the stars. To begin with, philosophers such as Inwood and Cooper stressed that Aristotelian practical syllogisms were primarily meant, not as arguments or inferences of which the conclusions happen to describe actions, but as models for explanation of actions (Inwood 1985; Cooper 1968; Cooper 1971). From that point of view,
The return of the akrasia problem

it was to be expected that problems generated by Aristotelian practical syllogisms would reappear in any modern discussion about action explanation.

Furthermore, for a long time the akrasia problem has been tacitly lingering under the surface of argumentations about human actions in general. It was there as it were incognito, not being recognised as the very difficulty that annoyed Greek philosophers from Socrates to Chrysippus. By way of example I refer once more to the final paragraphs of Anscombe’s Intention. To me it seems that the quintessence of Anscombe’s famous study is on the very last page, where a pur sang akratic action is described. Yet no matter how exemplarily akratic St. Peter’s denial of Christ in fact is, Anscombe abstains from calling it akratic, irrational, incontinent or weak-willed. It is as if she encounters our problem, without truly identifying it. Another example of the surreptitious presence of the problem is provided by Von Wright 1971. In the midst of a discussion about the LCA, Von Wright tells us about a revolutionary who aspires to free his compatriots by killing a tyrant. Here is the entire story:

"We have the premises of a practical argument: an agent intends to bring about something and considers the doing of something else necessary for this end. It is time for him to act. He thinks so himself. Perhaps he was resolved to shoot the tyrant. He stands in front of the beast, aiming at him with his loaded revolver. But nothing happens. Must we say that he is ‘paralyzed’? He is subjected to medical examination and nothing is found which would indicate that he was physically prevented from carrying his intention into effect. Must we say that he gave up his intention or that he revised the requirements of the situation? He refuses to admit either alternative. Must we say that he is lying? These questions aim at constructing a case in which to say that he was prevented, or forgot about the time, or gave up his intention, or reassessed the requirements of the situation would have no other foundation than the mere fact that he did not set himself to action in accordance with the premises. This is an extreme case, to be sure. But I do not see that it could not occur." (Von Wright 1971, 116-117).
III: The Logical Connection Argument

Von Wright here describes a full-blooded akatic action without, however, calling it by its name: he does not identify the rebel’s not-shooting as irrational, incontinent or akatic. But identified or not, the problem lies there in its full glory, and Von Wright has to handle it. That is, he must match his own story about the rebel with his belief in the validity of the LCA. Earlier we have seen that Von Wright considers that validity as an epistemological affair: "In this mutual dependence of the verification of premises and the verification of conclusions in practical arguments consists, as I see it, the truth of the Logical Connection Argument." (Von Wright 1971, 117). As a way out, Von Wright now concludes that "despite the truth of the Logical Connection Argument, the premises of a practical inference do not with logical necessity entail behavior" (ibid.). The necessity of the practical syllogism is, Von Wright tells us, a necessity *ex post actu*: "It is only when action is already there and a practical argument is constructed to explain or justify it that we have a logically conclusive argument." (ibid.). This means that in Von Wright’s story, where the rebel does literally nothing, an event occurs which is unexplainable in principle. In my opinion, that price is too high for a justification of akatic actions.