

# INFERENCES OF CAUSAL RELEVANCE FROM EXPERIMENTS

Gerd Graßhoff

## *Contents*

1	Theory and Experience . . . . .	1
2	Causal analysis . . . . .	3
2.1	Causal models . . . . .	3
2.2	Theory of causal regularities . . . . .	7
2.3	Principles of causal reasoning . . . . .	7
2.3.1	Method of Difference . . . . .	7
2.3.2	Assumptions . . . . .	8
2.3.3	Inferring a causal factor . . . . .	8
2.3.4	More complex designs . . . . .	10
2.3.5	Other inference patterns . . . . .	11
2.4	Difference tests in practise: notebook entries . . . . .	12
3	Methodology of causal models . . . . .	15

## *1 Theory and Experience*

For many the juxtaposition of theory and experience is governed by the logical properties of a deductive relation. Often it is abbreviated "hypothetico-deductive model of the science" (HD-model). "Theory" is a broad framework of general statements about natural properties and processes. Although theories and conjunctions of scientific statements may lack causal terminology in scientific papers, causal hypotheses are practically always involved. For the following arguments the only relevant aspect is that theories imply propositions that can be compared with empirical data. Here the implication is meant strictly in the sense of a deduction. The HD-model has at first sight attractive advantages. Independent of the logical internal structure of the theories the implication determines the methodological evaluations of the theories by comparing the theory with empirical data according to a simple pattern:

**Thesis 1.** *Theory evaluation according to HD is determined — but not exclusively — by logical consequences from the comparison between prognosis and empirical data. The mechanism is*

the comparison of truthvalues for identical propositional content. The comparison function returns a truth value for the Prognosis.

1. Theory  $\mathcal{E} h_1 \dots h_i \rightarrow$  Prognosis
2. Comparison ( Prognosis == Empirical Data )

Theories in conjunction with additional hypotheses  $h_1 \dots h_i$  derive a prognosis. The inferences are governed by the logical properties of the conditional. It is false only if the antecedent is true and the consequent is false. The empirical evaluation of theories then draws consequences from the comparison of truth values of the predicted proposition and empirical data. In case both propositions are the same and differ in their truth value, one statement is judged to be false: either the empirical data contradict the prediction or the empirical data must be revised.

**Thesis 2.** *The result of comparison (prognosis, empirical data) according to HD has then consequences for the evaluation of the theory (which might be just a hypothesis plus auxiliary hypotheses).*

1. In case the comparison reveals agreement between prognosis and empirical data, support  $(theory \wedge h_1 \dots h_i)$ .
2. In case of disagreement at least one proposition of  $(theory \wedge h_1 \dots h_i)$  is false. A select  $(theory \wedge h_1 \dots h_i)$  could be some sort of procedure to provide criteria for the selection of a proposition to be rejected.

It is the methodological consequence from the HD-model that a complex of hypotheses cannot gain support from a juxtaposition of its prognosis with empirical data. There are always alternative hypotheses in agreement with the data. There is a sense of methodological asymmetry between refutation and confirmation — and not prove — of theories. Still in case of a contradiction between prognosis and empirical data one obtains the information that at least one premise is false. Logically one doesn't know which of the premisses is false. Because of the strong inferential asymmetry Karl Popper draws the conclusion in *Logic of Scientific Discovery* from the logical properties of the HD-model:

“[A] theory of induction is superfluous. It has no function in a logic of science. The best we can say of a hypothesis is that up to now it has been able to show its worth, and that it has been more successful than other hypotheses although, in principle, it can never be justified, verified, or even shown to be probable. This appraisal of the hypothesis relies solely upon deductive consequences (predictions) which may be drawn from the hypothesis: There is no need even to mention ‘induction’”

Some followed Mill's proposal to solve the problem to identify the supported hypothesis by eliminating empirically equivalent alternatives. Then the empirical data could be explained by a specific hypothesis if alternatives could be eliminated by some kind of good reason beyond empirical data. There are two reasons why such procedure doesn't work: (I) each elimination based solely on empirical data faces the lack of specificity of modus tollendo tollens (MTT). (ii) MTT loses its power for refuting specific hypotheses for large or infinite set of premisses, as it is common in scientific reasoning. In case one doesn't want to follow Popper's radical conclusion, one would need criteria beyond the logical consequences of empirical statements. Duhem's *Bons sense* is a vague description of such a selection, but no solution.

## 2 Causal analysis

Instead of the HD-model I propose a different model for the empirical evaluation of theories. The dual relation between theory and empirical data will be replaced by:

1. The canonical statements of prognosis have the form of a general proposition of causal relevance as defined in [Graßhoff and May \(2001a\)](#):  $A \wedge X \vee Y$  are causally relevant for  $Z$ .
2. More complex causal statements can be composed according to rules of causal graphs.
3. Theory evaluation is (should be) done by rules of causal reasoning.
4. Empirical data which are used by rules of causal reasoning have the form of *difference tests* as described below.

In order to exemplify the new model of theory evaluation, the meanwhile classical case of the discovery of the urea synthesis by Hans Krebs and Kurt Henseleit serves as a model of typical experimental research employing causal reasoning.<sup>1</sup>

### 2.1 Causal models

In 1932 Hans Krebs and Kurt Henseleit explained the urea synthesis in animal liver by the urea cycle—the first cyclic metabolic pathway discovered in biochem-

<sup>1</sup> Various aspects of the case study [Graßhoff and May \(1995a\)](#), [Graßhoff and May \(1995d\)](#), [Graßhoff and May \(2003\)](#), [Nickelsen and Graßhoff \(2009\)](#); laboratory notebooks as facsimile and transcription in [Graßhoff and Nickelsen \(2001a\)](#), [Graßhoff and Nickelsen \(2001b\)](#).

istry.<sup>2</sup> This discovery was a milestone in the history of the discipline. For his subsequent studies of a similar process, the tricarboxylic acid cycle, Hans Krebs was later awarded the Nobel Prize.

Already in the mid-19th century new analytical techniques showed that the rate of urea synthesis in living animals increased when they were fed an additional supply of glycine and leucine. Schultzen and Nencki assumed in 1869 that amino acids are intermediates in the reaction chain from proteins to urea.<sup>3</sup> The introduction of the perfusion method marked an essential refinement of the experimental procedures. Reagents are guided through an organ outside the living organism, where the chemical composition of the leaving liquid is determined. In this way one found that not only glycine and leucine but almost all known proteins and amino acids increase the urea production in the liver. Until the twenties one attempted to optimize the utilization of the perfusion method; yet one did not succeed in decrypting the details of the chemical reactions leading to the formation of the urea.

At this time Hans Krebs was working as an assistant in the laboratory of Otto Warburg in Berlin. During these years he conducted basic research and obtained a practical and biochemical knowledge that would play an important role during his discovery of the urea cycle. Krebs owed to Otto Warburg especially the adaptation of the tissue slice method and the use of manometric devices for sensitive measurements of small amounts of substances. In summer 1931, after Krebs had moved to Freiburg, he started his research project on the urea synthesis with his doctoral student Kurt Henseleit. It lasted nearly a full year and took him nearly 200 experiments until the urea cycle could be established.

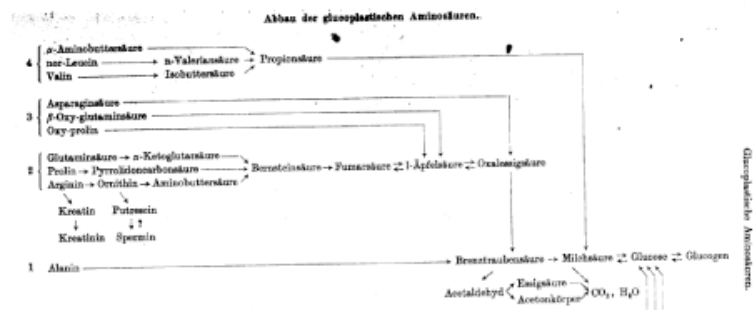


Fig. 1: A certain type of degradation path of amino acids in Neubauer, O. (1928), p. 845.

Krebs is guided by some quite general principles of causal reasoning. Crucial

<sup>2</sup> Shortened from Graßhoff and May (1995b).

<sup>3</sup> (Schultzen and Nencki, 1872).

for this kind of reasoning is a structure called a “causal graph”, which is a complex network of cause-effect relationships. Knowledge of causal relationships is essential for scientific activities such as explaining, predicting and controlling natural processes. The following discussion focuses on the causal aspects. In practice, causal reasoning is of course intertwined with other considerations.

A biochemical pathway is an instantiation of a causal graph. As an example, take the graph from figure 1. It shows the complicated paths of degradation of amino acids, according to Neubauer (1928), a standard textbook available to Hans Krebs. The types of causal relationships in this graph are manifold:

*Directly causally relevant factors:* alanine  $\hookrightarrow$  pyruvic acid

The presence of alanine (Alanin) causes the presence of pyruvic acid (Brenztraubensäure). This is the basic type of a causal relation.

*Causal chains:* alanine  $\leftrightarrow$  pyruvic acid  $\leftrightarrow$  lactic acid

The path from alanine to pyruvic acid and to lactic acid (*Milchsäure*) is a causal chain. Alanine for the formation of lactic acid is mediated by pyruvic acid. The figure contains many such pathways.

*Multiple effects:* acetaldehyde  $\leftrightarrow$  acetic acid, acetaldehyde  $\leftrightarrow$  acetone bodies.

The presence of acetaldehyde has more than one effect: it causes not only the presence of acetic acid, but also that of an acetone body. Different effects of a common cause tend to occur together. This makes them diagnostically relevant for each other.

*Multiple causes:* alanine  $\leftrightarrow$  pyruvic acid, oxaloacetic acid  $\leftrightarrow$  pyruvic acid

There are many different metabolic pathways for pyruvic acid, e.g. by alanine as well as by oxaloacetic acid.

An additional type of relationship are complex causes, where several substances have to be present for an effect; e.g. both hydrogen and oxygen have to be present for the formation of water. All relationships in figure 1 turn out to be complex, if the graph is specified in greater detail. Some of the reactions are such that two substances have to be combined to form a third. And all reactions depend on additional factors such as temperature, pressure and concentration.

A second additional type is an inhibiting factor that inhibits an effect that would be present if the inhibiting factor were absent. An example is a toxic substance that inhibits cellular respiration.

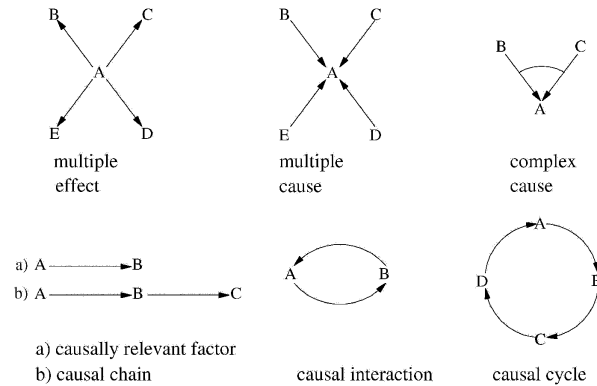


Fig. 2: Types of causal relationships that can be represented in a causal graph.

A third additional type, crucial for our case, is a cyclic process, where some specific event of type A is relevant for the occurrence of another event of the same type. Krebs' discovery was that degradation of amino acids in mammals is a cyclic process with ornithine acting as a catalyst. A reversible reaction, e.g. between glucose and glycogen in figure 1, is also a rudimentary type of a cyclic structure. Fig. 2 summarizes the possible causal relationships representable in a causal graph.

A formal analysis of causal relationship has been one of the major philosophical challenges. Although causal reasoning governs most of our daily life and scientific reasoning, it is all but clear whether one could formulate a set of rules which capture our intuitive understanding of causality. David Hume's analysis is mostly negative and Stuart Mill's concept has been refuted early on. Even the most advanced theories of causal regularity by John Mackie were abandoned by its author.<sup>4</sup> It has been a substantial part of the research project in Hamburg and now in Bern to formulate an adequate causal theory and a theory of causal reasoning that is applicable to well documented cases of scientific reasoning.<sup>5</sup> The graph structures which are discussed here can be given a logical interpretation.<sup>6</sup>

<sup>4</sup> (Mackie, 1980).

<sup>5</sup> The course book developed and used in Bern is Graßhoff and Baumgartner (2003). Recent studies Baumgartner (2008a), Baumgartner (2008b).

<sup>6</sup> (Graßhoff and May, 2001b), (Graßhoff and May, 1995c). Michael May developed the set of rules for causal reasoning described in this paper, (May, 1999a).

## 2.2 Theory of causal regularities

We interpret the graph structures by utilizing a variant of a causal regularity theory. Since details can be looked up elsewhere, our exposure will be brief.<sup>7</sup> The central idea is to represent regularities in a formalism that is similar—but not reducible—to disjunctive normal form in propositional logic. A complex cause is represented by a conjunction; alternative causes are represented by a disjunction of conjunctions; a causal factor is represented by a part of a conjunction, and an inhibiting causal factor by a negation. The presence of a complex cause is assumed to be sufficient for bringing about the effect. ‘To bring about’ implies an asymmetry that cannot be analysed in terms of the material conditional alone (we will give an operational definition below). Furthermore, the complex cause has to be *minimal*, i.e., no part of it is sufficient. This ensures that every factor that is part of the complex cause plays an indispensable part in bringing about the effect. A complex cause is also called *minimal sufficient condition*. Note that a causal factor is normally neither sufficient (since it is only a part of a complex cause) nor necessary (since there may be complex causes of which it is no part) for the effect.

The proposed account is based on two main ideas: uniformity and relevance. A complex cause is represented by a conjunction (“and”); alternative causes by a disjunction of conjunctions (“or”); a causal factor by a part of conjunction, and an inhibiting causal factor by a negation (“not”). The idea of relevance is implemented via the notion of minimality.

Our knowledge of a complex set of regularities is often highly incomplete: What we know are small *fragments* of only some complex causes of a causal network. How can we use incomplete knowledge and even partially wrong beliefs for explaining and predicting events? How can we revise an incorrect hypothesis? How can we generate a better one? Justifying the causal relevance of those partial structures we have found is arguably more important than identifying fully sufficient causes.

## 2.3 Principles of causal reasoning

### 2.3.1 Method of Difference

The basic test for causal relevance of a factor is intimately tied to a procedure called ‘method of difference’. By tying the test for causal relevance to this procedure we get a straightforward *operational definition* for the notion of ‘bringing about’: A factor  $C$  brings about an effect  $E$  if (1) there are two situations  $S_1$  and  $S_2$  which are causally homogeneous except that  $C$  is present in  $S_1$  and absent in  $S_2$ , and (2)  $E$  follows in  $S_1$ , but not in  $S_2$ . John Stuart Mill already studied this method in

<sup>7</sup> Graßhoff and Baumgartner (2003), Graßhoff and May (2003), Graßhoff and May (2001a), Graßhoff and May (2001c), May (1999b).

1843. But despite of this being the method scientists actually explicitly describe in articles—in laboratory notebooks we literally find hundreds of experiments that instantiate this method. Mackie [1974] has reconstructed the logic underlying Mill’s methods; Michael May took this as starting point to extend the basic method for handling much more complex types of experimental design.

This section introduces the basic inference pattern for inferring a positively relevant causal factor.

First, we will discuss the premisses of an inference; then we will show in an example that an application of the inference pattern leads to a deductively valid conclusion; finally we will shortly discuss more complex inference patterns.

### 2.3.2 Assumptions

The key to the proposal is to understand the generation of causal hypothesis as an iterative strategy. There is not one single step of empirical validation of a complex hypothesis.

**Thesis 3.** *Empirical validation (and generation) of complex causal hypotheses is a matter of an iterative process of causal reasoning.*

An inference step to a causally relevant factor of a more complex causal hypothesis is based on three assumptions:

- (i) A complex cause is sufficient for bringing about the effect, and if no complex cause is realized in a situation, the effect does not occur;
- (ii) the test situation is *causally homogeneous*;
- (iii) there is an initial hypothesis.

The first assumption is part of the concept of a deterministic causal regularity. The second assumption is discussed below. An initial hypothesis (assumption (iii)) can have different levels of specificity. On one extreme there is the bare assumption that *some* cause exists for the phenomenon to be explained—no specific factor is mentioned. On the other extreme is a hypothesis already specifying several minimal sufficient conditions. The discovery system starts with the first assumption and ends, if successful, with a justified hypothesis of the second kind.

### 2.3.3 Inferring a causal factor

We will assume that the initial hypothesis already specifies some factors. This is more complex than the most basic case which can be handled by the simple method

of difference. First, we assume that a factor  $A$ , together with an unknown conjunction of factors  $x_1$ , is a complex cause for an effect  $W$ . Second, we assume that a factor  $B$ , together with a second unknown conjunction of factors  $x_2$ , is a second complex cause for  $W$ . Third, we assume that there is an unknown disjunction  $y$  of additional minimal sufficient conditions. This is abbreviated

$$(H_0) \quad (Ax_1) \vee (Bx_2) \vee y \leftrightarrow W.$$

This is called an *incomplete hypothesis* or *causal fragment*. The unknown complete regularity is

$$(R) \quad (ACDE) \vee (B\bar{G}H) \vee (ED\bar{F}) \vee (GHK) \leftrightarrow W$$

(‘ $\bar{G}$ ’ means ‘not  $G$ ’). To test the causal relevance of a factor  $C$  for a condition  $(Ax_1)$  the following experiment is designed: Factors  $A$  and  $C$  are crossed in a factorial design, factor  $B$ , being an alternative intervening cause, is inhibited. An ‘1’ says that the effect  $W$  is present, a ‘0’ that it is absent.

$T_1$		$C$	$\bar{C}$
$A$	$\bar{B}$	1	0
$\bar{A}$	$\bar{B}$	0	0

Tab. 1: Test situation  $T_1$ .

The assumption of causal homogeneity puts a constraint on every causally relevant factor not varied explicitly: If this factor is instantiated in one cell (one of several corresponding test situations in an experiment) it is instantiated in every cell. This premiss has to be justified separately for every single case; there is no *a priori* guarantee for its correctness. Often we can justify this assumption even if we do not know the relevant factors; e.g., if we stir a solution carefully and distribute it among several vessels, we have good reason to suppose the contents of all vessels to be causally homogeneous—even if we do not know exactly what is in the vessel. The assumption of causal homogeneity says that each of the factors  $D, E, G, H, F, K$  is instantiated in every cell or in none. Let us assume that in our example the factors  $DEGHF\bar{K}$  are instantiated. From the result of the experiment, the initial hypothesis  $(H_0)$  and assumptions (i) and (ii), we can infer *deductively* that  $C$  is a part of the complex cause  $(Ax_1)$ .<sup>8</sup> This argument justifies a minimal expansion of  $(H_0)$  to

$$(H_1) \quad (ACx_1) \vee (Bx_2) \vee y \leftrightarrow W.$$

<sup>8</sup> For a proof, see May [1993]; compare also Mackie [1974].

**Proof**

- (1) A minimal sufficient condition is realized in cell 1 ( $AC$ ), but not in 2 ( $A\bar{C}$ ), since the effect  $W$  is present in 1 but not in 2.
- (2) Since cells 1 and 2 are causally homogeneous except for  $C$ , every minimal sufficient condition that does not contain  $C$  or its negation is present in both cases or absent in both cases.
- (3) Since there is no minimal sufficient condition realized in 2, there is *a fortiori* no condition realized that does not contain  $C$  or its negation.
- (4) It follows that no condition not containing  $C$  or its negation is realized in 1 either.
- (5) But since the effect is present in 1, a cause must be present. It follows from (4) that every cause that is present must contain  $C$  or its negation. And it must contain  $C$  unnegated, since  $C$  is instantiated in 1.
- (6) The same kind of argument applies to cells 1 and 3 ( $\bar{A}C$ ). The only difference is in  $A$ ; we can infer that no cause is present in 1 that does not contain  $A$ .
- (7) Since there is no complex cause present in 1 that does not contain  $A$  and no complex cause that does not contain  $C$ , all causes present in 1 must contain  $AC$  (note that this does not preclude more than one cause present at the same time).

This procedure can be iterated: In the next cycle we take ( $H_1$ ) as our initial hypothesis and examine a further candidate factor. To eliminate all alternative hypotheses, the experimental design demands to add an additional row for every newly discovered factor, so that the experiment becomes slightly more complicated. There are additional rules (which we have no space to discuss here) for inferring inhibiting factors, alternative causes and factors without relevance in specific minimal sufficient conditions.

#### 2.3.4 More complex designs

This procedure can be iterated: In the next cycle we take ( $H_1$ ) as our initial hypothesis and examine a further candidate factor. To eliminate all alternative hypotheses, the experimental design demands to add an additional row for every newly discovered factor. For testing a relevance of factor  $D$  we would have to conduct the following experiment:

From the experimental result we can again infer deductively

Tab. 2: Test situation  $T_2$ .

$T_2$		$D$	$\bar{D}$
$AC$	$B$	1	0
$\bar{A}\bar{C}$	$\bar{B}$	0	0
$AC$	$\bar{B}$	0	0

$$(H_2) \quad (ACDx_1) \vee (Bx_2) \vee y \leftrightarrow W.$$

Without the third line, this inference would not have been valid. There is a possibility that  $D$  is not a part of  $AC$ : e.g., if the complete regularity is

$$(R_2) \quad (AD) \vee (ACG) \leftrightarrow W,$$

with  $G$  being constantly absent, we would get the same result in cells 1-4. The effect in 1 is due to  $AD$ , not to  $ACG$ . There is no complex cause that contains  $ACD$ , so our conclusion would be wrong. But if we add the third line this is prohibited, since if  $(R_2)$  or a similar hypothesis were true, then the cause would have been present in 5 (since  $AD$  is present). By adding this line we are able to eliminate a whole class of alternative explanations for the experimental result.

### 2.3.5 Other inference patterns

Depending on the result of an experiment, different inference patterns can be applied. There are additional rules (which we have no space to discuss here) for inferring

- positively relevant factors,
- inhibiting factors,
- factors that are part of alternative causes,
- factors that are relevant through a causal chain,
- factors without relevance.

The procedure discussed in the last section demonstrates how to infer single factors. How can we justify a claim that a set of factors is minimal sufficient? The main principle is that of

- *Explaining anomalies.* If all known factors of a hypothesis currently investigated are present, but the effect is absent, this is noted as an *anomaly*. The goal is to expand the hypothesis in such a way that all anomalies can be explained. This does not have to be done immediately, but can be delayed until suitable experimental data are present. If all anomalies generated in previous experiments can be explained by the causal model—and in complex systems, there are a lot of them—this is good, although not logically conclusive, evidence that the hypothesis is minimal sufficient.
- *Irrelevance.* A claim that a hypothesis is minimal sufficient can be strengthened by proving the non-relevance of factors not contained in the hypothesis. This should not be conflated with a general causal irrelevance of factors for the generation of a specific effect. This cannot be demonstrated by difference tests. When it can be shown that all factors not contained in the hypothesis are non-relevant, its evidence is even conclusive (however, this is a limit we do not normally strive for).
- *Revision.* If no attempt to explain the anomalies is successful, the causal model has to be revised by dropping the hypothesis that gave rise to anomalies. Afterwards, the model can be expanded again.

#### 2.4 Difference tests in practise: notebook entries

Difference tests and rules of causal reasoning is still a philosophical terminology, hardly used by scientists. Do they after all employ these techniques? The case study exhibits the typical scientific terminology in which rules of causal reasoning are embedded.

Figures 3 and 4 show the first and last (out of three) pages of Henseleit's entry on his experiment with ornithine. These pages are typical of the style in which both researchers recorded their experimental work. It begins with a dated, underlined title. In most cases they mention the name of the substances to be tested. It is followed by a characterization of the organism from which cell tissues had been taken for the experiments. A short phrase is added to describe the procedures applied to the cell material before it was planted into the experimental tube (e.g. "Schnitte ausgewaschen").

The next sections on the page record the specific experimental properties such as temperature, the chemical composition of the solution and the time during which the tissues were exposed to substances in the solution.

The lower two thirds of the page display a table structure. The table is divided into two sections: on top there are rows beginning with the amount and name of substances added to the solution (e.g. first row "Ringer Lösung", second

110

Einfluss von Zucker & Säuren auf die Harnstoffbildung am NH<sub>4</sub>Cl

Einfluss von Ornithin.

Mit Kalkphosphat ernährte Ratten, Scheitelle angewendet.

Vorversuch: Temp. 37,5° 4,60 Vol.-% CO<sub>2</sub> → O<sub>2</sub> Versuchsdauer 120 Min.

3 von Fingerzähl → → → →

4,12 Vol.-% → → →

904 fl. Kub. cm  
- 200 mg %

mg Leber	16,40	14,67	13,92	22,47	16,31	16,88
<u>Harnstoffbestimmung: Anordnung wie S. 90.</u>						
	<u>16</u>	<u>17</u>	<u>18</u>	<u>19</u>	<u>20</u>	<u>21</u>
KCO <sub>2</sub>	0,155	0,986	0,928	0,940	1,649	1,405
p	3,1	3,1	3,1	3,1	3,1	3,1
h	7,83	9,31	8,64	8,29	16,90	14,09
h	+8	0	+74	+106	+62	+84,5
xCO <sub>2</sub>	6,84	0	68,6	99,6	102,1	118,8
W <sub>Leber</sub> x H <sub>l.</sub> mm <sup>3</sup>	<u>0,24</u>	<u>0</u>	<u>3,06</u>	<u>2,76</u>	<u>3,98</u>	<u>4,46</u>

Fig. 3: First page of Henseleit's laboratory entry about the ornithine experiments.

row one column to the right “NH<sub>4</sub>Cl”). Little arrows to the right of the name of the substances indicate that the same amount of substance was added to the same experimental setup. Hence each column defines the specific attributes of one experimental arrangement within the comparable series of tests. The lower section of the table starts with the record of the weight of the liver tissue relative to which the amount of urea is measured. Since Krebs and Henseleit used manometric methods for their measuring devices, they recorded differences of pressure as measuring data. Together with auxiliary data they calculated the specific amount of urea formed in the reaction tube. This value comes at the bottom of the column and is typically underlined by either Henseleit, who recorded most of the experiments, or

112

Fortsetzung von S. 111/11.

3ccm Ringer-Lös.	→	→	→	→	→
0,12 NH <sub>4</sub> Cl 1%	→			→	→
0,06 Pyruvat 10% = 200 mg %		0,1 Ornithin (Hoffm. de Roche) 5% = 100 mg %		→	→
<u>15,77</u>	<u>19,57</u>	<u>13,75</u>	<u>20,844</u>	<u>16,17</u>	<u>13,49</u>
<u>32</u>	<u>28</u>	<u>26</u>	<u>28</u>	<u>30</u>	<u>31</u>
1,207	1,528	1,460	1,578	1,391	1,691
3,1	3,1	3,1	3,1	3,1	3,1
12,94	15,59	14,72	15,59	12,92	17,24
+46,5	+49	+6,5	+10	121,5	99,5
60,8	75,4	9,5	15,28	169,1	168,5
<u>2,49</u>	<u>2,45</u>	<u>0,43</u>	<u>0,46</u>	<u>6,74</u>	<u>8,04</u>

Fig. 4: Third and last page of Henseleit's laboratory entry about the ornithine experiments.

by Krebs in his laborator notebook.

In many cases the experimental record in the notebook is concluded with a very short summary titled "Ergebnis". But the historian who had hoped to find here a detailed document of an ongoing thought-process of the researcher will be disappointed. In most cases a result is noted, e.g. that substance A or B leads to an increase in the formation of urea, or else there is a brief conclusion, for instance that next time one should wash the tissues for a longer time.

In the case of the ornithine experiment only the last two columns exhibit an experimental setting, in which ornithine *together* with ammonia produces much more manometric pressure—and hence urea—than all other settings, even with ornithine alone. This is in fact a very puzzling result because according to the stan-

dard hypothesis, with which Krebs was operating at the time, even ornithine alone should have produced urea. This experiment showed the contrary (see the fourth and third last columns in fig. 4); ornithine alone did not produce urea. Further implications of the experiment are not recorded under the title “Ergebnis”. But Krebs and Henseleit did draw conclusions: the notebook carefully records all experimental conditions and those of the measuring process, which led to the measured data and from which they derived the putative findings in terms of the specific formation of urea.

The experiment is set up as a causal test for factors which are either already known to play a causal influence in the formation of urea or which are tested for that role. Typically the notebook entry records such conditions. Krebs set up the experiments as a causal difference test: if two experimental situations are equal in causally relevant aspects and one additional factor exhibits an effect which is missing in the comparable situation without the presence of that factor, then Krebs rightly concludes the causal relevance of that factor. Krebs and Henseleit operated throughout their work with this methodological machinery. The assumption that the comparable situations should be equal to each other in causally relevant aspects is crucial. The experimenter’s skill is revealed in his ability to realize such a condition in an experiment. Should small differences occur, one must be able to control them and account for them. The only way to test the validity of that condition (which we call *condition of homogeneity*) is to repeat an experiment under conditions which one can control, and check whether the observed effects remain the same. If this is not the case, the condition of homogeneity is not satisfied and the experiment does not allow causal conclusions. Krebs had to learn it the hard way, when he conducted experiments with thymine. He measured some increase of pressure with his manometer and jumped to the conclusion that thymine is causally relevant for the formation of urea. It took him a month’s work to discover his error: he had not controlled the validity of the condition of homogeneity with sufficient care. He rectifies his error by repeating the experiment right from the beginning. The notebook entries for the ornithine experiment therefore contain two columns with an identical arrangement. The result is that there are small fluctuations in the outcome but not in the dimension of the “ornithine effect” (i.e. the production of ornithine for the given conditions). For this range of error Krebs could thus treat his experimental procedures as causally equal.

### 3 Methodology of causal models

Instead of using HD-models for justifying causal hypotheses, causal reasoning reverses methodological issues which are often taken for granted.

**Thesis 4.** *A causal hypothesis (claim of causal relevance) cannot be falsified by empirical data. It can be strictly proven that no possible outcome of even a complex difference test allows an inference to the negation of a proposition about causal relevance.*

Just the opposite of what Popper claimed:

**Thesis 5.** *Causal inferences allow the positive justification of a causal relevance on the assumption, (i) homogenous conditions (ii) justified initial hypothesis and (iii) correct empirical readings of the effects.*

A claim without proof

**Thesis 6.** *There is no other empirical justification for causal relevance claims than iterated difference tests.*

### References

- Baumgartner, M. (2008a). Regularity theories reassessed. *Philosophia*, 1–28.
- Baumgartner, M. (2008b). Uncovering deterministic causal structures: a boolean approach. *Synthese*, 1–26.
- Graßhoff, G. and M. Baumgartner (2003). *Kausalität und Kausales Schliessen*. Bern: Bern Studies in the History and Philosophy of Science.
- Graßhoff, G. and M. May (1995a). From historical case studies to systematic methods of discovery. In *Papers from the AAAI Spring Symposium Series 'Systematic Methods of Discovery'* Stanford University, Menlo Park. AAAI Press.
- Graßhoff, G. and M. May (1995b). From historical case studies to systematic methods of discovery. In *AAAI Spring Symposium Series*. AAAI Press.
- Graßhoff, G. and M. May (1995c). Methodische analyse wissenschaftlichen entdeckens. *Kognitionswissenschaft* 5, 51–67.
- Graßhoff, G. and M. May (1995d). Methodische analyse wissenschaftlicher entdeckungen. *Kognitionswissenschaft* 5(2), 51–67.
- Graßhoff, G. and M. May (2001a). Causal regularities. In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current Issues in Causation*, pp. 85–114. Mentis Verlag.
- Graßhoff, G. and M. May (2001b). Causal regularities. In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current Issues in Causation*, pp. 85–114. Paderborn: Mentis.

- Graßhoff, G. and M. May (2001c). Reconstructing the discovery of the urea cycle. In F. L. Holmes, J. Renn, and H.-J. Rheinberger (Eds.), *Reworking the Bench: Research Notebooks in the History of Science*. in Vorbereitung.
- Graßhoff, G. and M. May (2003). Hans krebs' and kurt henseleit's laboratory notebooks and their discovery of the urea cycle - reconstructed with computer models. In H.-J. R. Frederic L. Holmes, Jürgen Renn (Ed.), *Reworking the Bench. Research Notebooks in the History of Science*, Volume Volume 7 of *Archimedes. New Studies in the History and Philosophy of Science and Technology*, pp. 269–294. Dordrecht: Kluwer Academic Publishers.
- Graßhoff, G. and K. Nickelsen (2001a). *Dokumente zur Entdeckung des Harnstoffzyklus. Bd. 1: Laborbuch Hans Krebs und Erstpublikationen*. Bern: Bern Studies in the History and Philosophy of Science. Gerd Grasshoff, Kärin Nickelsen Bd. 1 Laborbuch Hans Krebs und Erstpublikationen Bd. 2 Laborbuch Kurt Henseleit.
- Graßhoff, G. and K. Nickelsen (2001b). *Dokumente zur Entdeckung des Harnstoffzyklus. Bd. 2: Laborbuch Kurt Henseleit*, Volume 2. Bern: Bern Studies in the History and Philosophy of Science.
- Mackie, J. L. (1980). *The Cement of the Universe. A study of Causation* (2nd ed.). Oxford: Clarendon Press.
- May, M. (1999a). *Kausales Schließen*, Volume 64. Berichte des Graduiertenkollegs Kognitionswissenschaft.
- May, M. (1999b). *Kausales Schliessen. Eine Untersuchung über kausale Erklärungen und Theorienbildung*. Ph. D. thesis, University of Hamburg.
- Neubauer, O. (1928). *Intermediärer Eiweißstoffwechsel*, pp. 797–825. Berlin: Julius Springer.
- Nickelsen, K. and G. Graßhoff (2009). Concepts from the bench. krebs and the urea cycle. In G. Hon, J. Schickore, and F. Steinle (Eds.), *Going Amiss in Experimental Research*, Dibner Institute Studies in the History of Science and Technology, pp. 91–117. Boston: Springer.
- Schultzen, O. and M. Nencki (1872). Die vorstufen des harnstoffs im thierischen organismus. *Zeitschrift für Biologie* 8, 124–146.