Laws of Nature and Causality in the Special Sciences

A Philosophical and Formal Analysis

Bert Leuridan

Proefschrift voorgedragen tot het bekomen van de graad van
Doctor in de Wijsbegeerte

2008

Promotor:	Prof. dr. Joke Meheus Vakgroep Wijsbegeerte en Moraalwetenschap
Copromotor:	Prof. dr. Erik Weber Vakgroep Wijsbegeerte en Moraalwetenschap
Decaan:	Prof. dr. Freddy Mortier
Rector:	Prof. dr. Paul Van Cauwenberge

All scientific work is incomplete – whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time. (Hill, 1965, 300)

Preface

This dissertation concerns laws of nature and causality in the special sciences. The label 'special sciences' covers many scientific disciplines, ranging from genetics over occupational medicine and epidemiology up to economics and the social sciences. I will perforce restrict myself mainly to the theory of classical genetics, but on occasion I will also glance at other disciplines.

Are there any laws of nature in the special sciences? It has often been argued there are not. I will argue by contrast that this view is mistaken. There are laws of nature in the special sciences ... provided we adopt an appropriate conceptual framework. Given such a framework, we may rightly call the principles of e.g. the theory of classical genetics 'laws of nature'.

Laws of nature and causality are closely linked with many others topics from philosophy of science. I will go more deeply into the following issues: mechanisms and scientific theories, explanation and policy, experimentation and causal discovery, ... Thus I will not only examine what laws of nature are, but also what they can be used for and how they can be discovered.

As such, this dissertation should be looked upon as *philosophy of science*. More precisely, it should be considered *formal* philosophy of science. Throughout I will make use of the language of causal modelling and Bayesian nets to tackle the above issues. And I will present a formal logic for causal discovery. Finally, it should be deemed *historically based* philosophy of science. A large part of it will rely on papers and monographs from the history of classical genetics and rival theories of inheritance.

By calling this dissertation historically based formal philosophy of science, I purposively exclude two things. Firstly, it is not intended as history of science. Secondly, it is not pure metaphysics. Laws of nature have been amply discussed in metaphysics, where they are often understood as metaphysical entities underlying the regularities observed in the world. I will define laws of nature as generalizations describing these regularities. What brings about such regularities is a question I will mainly sidestep. Any philosophy of science involves metaphysical commitments – this I certainly do not want to dispute –, but still we should distinguish between these two domains.

Acknowledgements

I am greatly indebted to many people. Firstly, I want to thank Joke Meheus, my supervisor, and Erik Weber, my co-supervisor. Their support during the past four years was worth a lot and their comments on the drafts of this dissertation have helped a great deal.

I also want to thank Sandra Mitchell, Jon Williamson, Diderik Batens, Maarten Van Dyck, Jan-Willem Romeijn, Gert de Cooman, Filip Hermans, Eric Schliesser, Lindley Darden, Carl Craver, Stuart Glennan, Marcel Weber, Albrecht Heeffer, several anonymous referees, and many others, for commenting on the ideas to be presented, and for agreeing with my views, or for strongly disagreeing with them ...

I want to thank Joke, Erik, Maarten and Albrecht for proofreading this text, and Gitte and Marc, my father, for their technical support.

And last but not least, I am enormously grateful to you, Annelies. Your patience, your empathy and your support during the past weeks, months and years have been more important than anything else.

Contents

Chapter 1

Introduction. What are laws of nature?

For years on end, 'laws of nature' have played an important role in philosophy of science. What laws of nature are, and what criteria they satisfy, thus was an important philosophical issue. Within the logical empiricist tradition, criteria have been developed which gave rise to what now can be labeled 'the received view'. With much rigour, and with the help of first order predicate logic, lawful generalizations were distinguished from mere accidental ones. I called this approach 'the received view', but an even better label would be 'the once received view', since it has come under attack ever more. In section 1.1 I will briefly present the criteria in question and show why they are wanting as regards the theories and generalizations of science – especially of the special sciences! In sections 1.2 and 1.3 I will present two alternative approaches to laws of nature in the special sciences. The first is Sandra D. Mitchell's pragmatic approach. The second is Jim Woodward's concept of 'invariance' as the hallmark of causal and/or explanatory relations. In section 1.4 I will synthesize their ideas and propose four related concepts: P-laws, P-regularities, cP-laws and cP-regularities. These concepts, it will be seen, are very fruitful for tackling the problem of laws of nature and causality in the special sciences.

1.1 The received view on laws of nature

For decennia, 'laws of nature' (or 'scientific laws') have played an important role in philosophy of science. The discovery of laws was deemed the hallmark of good science and the laws themselves were deemed indispensable for prediction, explanation and the evaluation of counterfactual statements.

Some statements were considered paradigmatic examples of scientific laws, such as (1.1), (1.2), or (1.3). Others were considered paradigmatic examples of accidental generalizations, such as (1.4), or (1.5).

Newton's laws of motion (1.1)

Kepler's laws of planetary motion (1.2)

"All copper when heated expands" (1.3)

"All the screws in Smith's car are rusty" (1.4)

"All the coins in my pocket to-day are made of silver" (1.5)

With both the alleged functions of prediction, explanation and support of counterfactuals and these paradigmatic examples in mind, the necessary and sufficient conditions for lawlikeness were sought. I will list seven criteria that were frequently cited, e.g. by Nagel (1961, 29–78), Goodman (1973, 17–27, 73–81) and Hempel (1965, 231, 265–268). These criteria were strongly shaped by the language, as well as the inferential power of first order classical logic (see also Mitchell, 2000 and Bechtel and Abrahamsen, 2005). To be sure, this list is not exhaustive and both the relative importance and the precise interpretation of these criteria were heavily discussed. But they strongly influenced the debate on laws of nature in philosophy of science and are rightly called the 'received' view.

First, however, some conceptual clarifications are in order. I will treat 'law of nature' and 'scientific law' on a par.[1] By 'law of nature' or one of its synonyms, I will always mean a *generalization*. Laws are generalizations *describing* regularities. Hence they should not be confused with these regularities. Nor should they be interpreted as metaphysical entities that *produce* or *are responsible for* these regularities.[2]

Let us now have a look at the conditions to be satisfied by laws or lawful generalizations. Generalizations that satisfy these conditions, I will call *strict* laws. These will be contrasted with weaker laws in subsequent sections. The first condition is *universality*. Laws of nature are represented as universal conditionals $(\forall x)(Ax \supset Bx)$, where '$\forall$' is the classical universal quantifier and '$\supset$' is material implication. (In Hempel (1965, 266) conditional form is considered less important.) The second criterion is *truth* or *high probability*. Sentences describing lawful regularities should be true (Hempel, 1965,

[1] But I will distinguish between 'strict laws' and 'pragmatic laws'.

[2] A different view, viz. that laws of nature are different from both regularities and generalizations and that they produce the former can be found in Mumford (2004). (However, Mumford argues that no such laws can exist in nature.)

265), or they should have adequate empirical support, or high probability (Nagel, 1961, 42–43). Thirdly, laws of nature are *not vacuously true*. A universal generalization $(\forall x)(Ax \supset Bx)$ is vacuously true if nothing satisfies its antecedent, i.e. if no object or event is A. Vacuous truth was mostly considered an unacceptable feature for scientific laws. Otherwise, assuming that unicorns do not exist, an infinite number of 'laws' such as 'All unicorns are black', 'All unicorns are red', ... would have to be accepted. (Nagel, 1961, 50–52) The fourth criterion is that laws of nature should have *no genuine exceptions* (i.e. exceptions which are not the result of poor observation, careless experiments, ...). This is implied by the conditions of universality and truth. (The condition of exceptionlessness was not that strict, however. Nagel (1961, 65–66) allowed for genuine exceptions if the law concerned is sufficiently supported by indirect evidence.)

The conditions given so far do not suffice to rule out accidental universal generalizations such as (1.4) or (1.5). Indeed, both are universal generalizations and the world may be such that they are non-vacuously true and have no genuine exceptions. But to many philosophers it seemed (and does seem) unacceptable to classify them as laws. Three more criteria were introduced to remedy this problem. The fifth criterion states that laws should be *general* or *non-local*. In a lawlike generalization, no reference to particular objects or specific spatio-temporal locations (either explicitly or implicitly) is allowed. However, this condition has to be restricted to fundamental laws such as (1.1). Otherwise, derivative laws such as (1.2) threaten to loose their status as scientific laws.[3] (Nagel, 1961, 57–58; Hempel, 1965, 267) According to the sixth condition, laws of nature have to be *projectible* and to have *unlimited scope*. If the evidence supporting a generalization is known or supposed to exhaust that generalization's scope of predication, or if its scope of predication is known or supposed to be closed for further augmentation, the generalization in question would not be appropriate for prediction and hence not be considered a law of nature. Consequently, a law had to be acceptable prior to the determination of all its instances. (Nagel, 1961, 62–64; Hempel, 1965, 267; Goodman, 1973, 20–21) Finally, one of the most striking features of a law of nature is that it expresses not merely *de facto* generality, but some strong or 'necessary' connection between properties or kinds. Laws of nature have some kind of *necessity*. This condition is closely connected with the conviction that laws of nature support counterfactuals. Consider a screw s that never was in Smith's car and is destroyed, so that it never

[3] A law is *fundamental* if it contains no individual names and if all its predicates are purely qualitative (i.e. do not require reference to particular objects or spatio-temporal locations). A law is *derivative* if it is a logical consequence of a set of fundamental laws. (Nagel, 1961, 57)

will be in Smith's car. Generalization (1.4) cannot guarantee that, had s been in Smith's car, it would have been rusty. Suppose, by contrast, that some piece of copper c was never heated and is destroyed too. Generalization (1.3), it is assumed, allows us to accept the counterfactual claim that if c had been heated it would have expanded. The relation between temperature and volume is deemed 'necessary', while that between 'being is Smith's car' and 'being rusty' is not.[4] (Nagel, 1961, 50–56, 68–73)

For decades, these criteria for lawfulness have dominated philosophical discussions about laws of nature. But in the 1980's and the 1990's, many problems cropped up. If the criteria are accepted at face value, we end up having little or no laws in the sciences at all. This holds for physics (Cartwright, 1983), as well as for chemistry (Christie, 1994), biology (Beatty, 1995, 1997; Brandon, 1997; Sober, 1997) and the social sciences (Beed and Beed, 2000; Roberts, 2004). The generalizations that are put forward in these sciences often fail to correctly describe reality since they involve many idealizations. Hence they violate the criterion of nonvacuous truth. Often they have genuine exceptions. Many refer to particular objects or spatio-temporal locations either implicitly or explicitly. And most lack necessity since they are contingent upon historically evolved conditions. Given that laws are traditionally deemed indispensable for prediction, explanation and the assessment of counterfactuals (and in recent approaches also for manipulation, cf. infra), this is highly problematic.

This situation has led some philosophers to abandon the concept of 'law of nature' and to search for alternatives. Others have sought for an alternative interpretation of 'law' itself. In the following sections I will introduce Sandra Mitchell's concept of 'pragmatic laws' (section 1.2) and Jim Woodward's concept of 'invariant relations' (section 1.3). Then, in section 1.4 I will propose the overarching concepts of 'P-law', 'P-regularity', 'cP-law' and 'cP-regularity'.

1.2 Pragmatic laws

Instead of rejecting the concept of law of nature, Sandra Mitchell (1997, 2000) sets out to refine it. She starts from the findings of the previous sec-

[4]Note that the condition of necessity was not unproblematic. Firstly, there are many kinds of necessity (logical, causal, physical, ...). While some (e.g. conventionalists) interpreted scientific laws as logically necessary, most did not. Secondly, since the concept of law of nature was mostly discussed within the neopositivistic framework, the problem was how to accommodate the idea of necessity (support of counterfactuals) with an empiricist world view.

tion, viz. that the existing criteria for lawfulness are too restrictive, at least with respect to biology and perhaps also with respect to other sciences. Biological generalizations are contingent, they have exceptions, etc. Therefore, she proposes a pragmatic approach to the question whether there are laws in biology. (Again, the term 'laws' refers to (some special kind of) generalizations describing regularities, not to some alleged metaphysical entities producing these regularities.)

> The pragmatic approach focuses on the *role* of laws in science, and queries biological generalizations to see whether and to what degree they function in that role. (Mitchell, 1997, S469, original emphasis)

The roles of laws that Mitchell focuses on are prediction, explanation and manipulation. If a generalization is used for one or several of these tasks, it qualifies as a pragmatic law. (From this it does not follow, however, that laws are sufficient for these tasks. Perhaps some extra ingredient might be necessary.)

Mitchell contrasts the pragmatic approach for evaluating the lawfulness of biology both with the normative and the paradigmatic approach. In the normative approach one begins with a norm or *definition* of lawfulness, more specifically the traditional criteria for strict lawfulness (cf. supra), and reviews the candidate biological generalization to see if it meets the specified conditions. The paradigmatic approach begins with a set of *exemplars* of laws (characteristically in physics) and compares these to the generalizations in biology. If a match is found, the generalization is considered a law. (Mitchell, 1997, S469; Mitchell, 2000, 244–250) As we saw in section 1.1, paradigmatic and pragmatic considerations also played an important role in the works of Hempel, Nagel and Goodman. Criteria for lawfulness were assumed to rank Newton's paradigmatic laws of motion as laws and statements about the screws in Smith's car as accidental generalizations. The criteria also had to be such that laws are the vehicles for prediction (Goodman) and explanation (Hempel) par excellence. So Mitchell's approach does not differ radically in spirit from the traditional one. The main difference, and also the most interesting one, concerns the new, gradual criteria she proposes for the ranking of lawful generalizations. (Mitchell, 1997, S475–S478; Mitchell, 2000, 259–263)

Generalizations are laws if and to the extent that they can be used for prediction, explanation or manipulation. Therefore, they must be projectible.

> The function of scientific generalizations is to provide reliable expectations of the occurrence of events and patterns of properties. The tools we use and design for this are true generalizations that

> describe the actual structures that persist in the natural world. (Mitchell, 1997, S477)

Given that these generalizations will seldom be universal, we need to know when (in what contexts) they hold and when they don't. The interesting problem is not *that* biological generalizations are contingent, but how and to what extent. Therefore, if we want to use a generalization, we need to assess the stability and strength of the relation or regularity it describes. Stability and strength are two very important ontological parameters for the evaluation of a generalization's usefulness. (To these, Mitchell also adds several gradual representational parameters, such as degree of accuracy, level of ontology, simplicity, and cognitive manageability (Mitchell, 1997, S477–S478; Mitchell, 2000, 259–263). I will not discuss these parameters here.)

Stability What are the conditions upon which the regularity under study is contingent? How spatio-temporally stable are these conditions? And what is the relation between the regularity and its conditions (is it deterministic, probabilistic, etc.)?

Stability is a gradual parameter. All regularities are contingent in that they rest on certain conditions. These conditions are historically shaped and are to a certain extent spatio-temporally stable. Stability does not bear solely on the laws of physics. Only if contingency is interpreted gradually, Mitchell claims, our conceptual framework will be rich enough to account for the diversity of types of regularities and generalizations and for the complexity found in the sciences (Mitchell, 1997, S469–S477; Mitchell, 2000, 250–259). Strength too is a gradual parameter.

Strength How strong is the regularity itself? Does it involve low or high probabilities? Or is it deterministic? Does it result in one unique outcome? Or are there multiple outcomes?

Mitchell's pragmatic approach raises two questions which I should deal with first. Firstly, the approach is very liberal and one may urge that it qualifies too many generalizations as lawful. More specifically, Mitchell's approach is often criticised for allowing for very weak and/or unstable laws. Secondly, one may question whether it sufficiently allows for distinguishing between causal laws and non-causal laws.

Is Mitchell's approach too liberal? Obviously it is from the traditional point of view. Few pragmatic laws satisfy the criteria for strict lawfulness. But does this provide us with sufficient reason to deprive them of their honorific label 'law'? And more specifically, should the fact that Mitchell's approach allows for very weak and/or unstable pragmatic laws count as a

shortcoming? I am of course willing to give up the word 'law', but I doubt this would be of any help. Moreover, there are two good reasons to stick to Mitchell's approach.

A first reason is that many scientific generalizations (in many different scientific disciplines) are called laws, while failing to satisfy the criteria for strict lawfulness. By contrast, their status as a law and their usefulness in practice can be easily acknowledged within Mitchell's framework. Theories of lawfulness that apply more stringent criteria run the risk of selling short these generalizations. The history of classical genetics provides us with a nice example of such non-strict scientific laws. In 1900 William Bateson was deeply convinced that it would be both useful and possible to discover the laws of heredity (Bateson, 1900). This conviction was mainly inspired by the works of Francis Galton, who formulated the law of regression and what would later be known as the law of ancestral inheritance (Galton, 1889, 1897). But at that time Bateson also got acquainted, via Hugo de Vries, with the works of Gregor Mendel (Mendel, 1933; de Vries, 1900a). What is particularly interesting is the way Bateson conceived of the laws of heredity. He acknowledged that both Galton's laws and Mendel's law (at that time, Bateson did not distinguish between the law of segregation and the law of independent assortment) are subject to exceptions and have a limited scope of application. However, this did not dissuade him from holding to the label 'law'. Nor did he later change his mind, when ever more exceptions to Mendel's laws were adduced by the biometricians, e.g. Weldon (1902), who rejected Mendel's theory in favour of Galton's (Bateson, 1902).

The idea that 'laws' may have exceptions, or limited scope of application was not new. Some 35 years before, Gregor Mendel had explicitly declared that before the time he started his investigations, "no generally applicable *law* of the formation and development of hybrids ha[d] yet been successfully formulated"[5] (Mendel, 1865, 2, my emphasis) and that the purpose of his experiments was to observe the changes to which pairs of opposing traits are subject in the offspring of hybrids, "and to deduce the *law* according to which they appear in successive generations."[6] (Mendel, 1865, 5, my emphasis) Mendel found many regularities and used various labels to describe them: the 'law of combination of differing traits', the 'law of development', the 'law of simple combinations', ... From the fact that he spoke in terms of laws it should not be concluded he *a priori* considered them to be universal or

[5] "[Es ist] noch nicht gelungen [...], ein allgemein gültiges *Gesetz* für die Bildung und Entwicklung der Hybriden aufzustellen [...]." (Mendel, 1933, 1, my emphasis)

[6] "Diese Veränderungen für je zwei differirende Merkmale zu beobachten und das *Gesetz* zu ermitteln, nach welchem dieselben in den aufeinander folgenden Generationen eintreten, war die Aufgabe des Versuches." (Mendel, 1933, 6, my emphasis)

exceptionless. Instead, their scope of application and their projectibility had to be determined by further research. "The object of further experiments will be to determine whether the *law* of development discovered for *Pisum* is also valid for hybrids of other plants."[7] (Mendel, 1865, 32, my emphasis)

In the works of Thomas Hunt Morgan and his co-workers (Morgan et al., 1915; Morgan, 1919, 1928), Mendel's findings of segregation and independent assortment were also called laws, even if they were complemented with systematic explanations of their failures (coupling and crossing-over, sex-linked inheritance, failure of dominance, ...). And even today, textbooks in modern genetics start with an overview of Mendel's laws (Klug and Cummings, 1997, chapter 3).

Mendel's findings were certainly not strict laws,[8] but their usefulness can be acknowledged within the pragmatic approach, as can their status as 'laws'. Much research in classical genetics aimed at uncovering the conditions for the different regularities, assessing their stability, specifying their strength, etc. Nothing is gained by merely claiming these regularities are not lawful. This completes the first reason to stick to Mitchell's approach.

A second reason for sticking to Mitchell's approach is that it also nicely fits actual scientific practice. Scientists invest plenty of time and money to discover (statistical) regularities that can be used for prediction, explanation or interventions. Granted, few of the resulting descriptions are called laws. But what is more interesting is the fact that the criteria used nicely fit Mitchell's liberality. In 1965, Sir Austin Bradford Hill famously addressed the problem of causal inference (Hill, 1965, 295). Hill's paper is still very influential today (at least it is cited frequently). He envisaged situations in occupational medicine in which our observations reveal a statistically significant association between two variables (a disease or injury A and conditions of work B) but where our background knowledge (the general body of medical knowledge) does not suffice to determine whether the relation is causal. His paper was unquestionably motivated pragmatically:

> In occupational medicine our object is usually to take action. If this be operative cause and that be deleterious effect, then we shall wish to intervene to abolish or reduce death or disease. (Hill, 1965, 300)

[7] "Es wird die Aufgabe weiterer Versuche sein, zu ermitteln, ob das für Pisum gefundene *Entwicklungsgesetz* auch bei den Hybriden anderer Pflanzen Geltung habe." (Mendel, 1933, 32, my emphasis)

[8] Mendel's laws fail to satisfy the criteria for strict lawfulness in yet another respect. They are evolutionarily contingent. They are contingent on conditions that are the product of evolution. So they violate the criterion of *necessity*. (Beatty, 1995).

To be useful in reducing death or disease, an association needn't be strong:

> We may recall John Snow's classic analysis of the opening weeks of the cholera epidemic of 1854 [...]. The death rate that he recorded in the customers supplied with the grossly polluted water of the Southwark and Vauxhall Company was in truth quite low – 71 deaths in each 10,000 houses. What stands out vividly is the fact that the small rate is 14 times the figure of 5 deaths per 10,000 houses supplied with the sewage-free water of the rival Lambeth Company. (Hill, 1965, 296)

The weakness of the relation between sewage and cholera – $P(cholera \mid sewage)$ is very low – doesn't make it unusable for occupational (preventive) medicine. It underlay interventions to improve public health. To be useful in reducing death or disease, an association also needn't be very stable:

> Arsenic can undoubtedly cause cancer of the skin in man but it has never been possible to demonstrate such an effect on any other animal. (Hill, 1965, 298)

Whether arsenic causes cancer in non-human animals is of little interest if the intended domain of application consists of humans. Evidence from humans should suffice. To conclude, the case of Hill shows that Mitchell's approach nicely fits the pragmatic slant of occupational medicine (which, after all, is part of the life sciences). And it shows that Mitchell's liberality regarding very weak and/or unstable pragmatic laws is a strength, rather than a weakness.

The case of Hill brings us to the second question that is raised by Mitchell's approach. Hill explicitly intended to distinguish causal regularities from mere association, but Mitchell's framework provides no means for making such a distinction. A regularity can be very stable or very strong, even if it is a spurious association. The distinction is favourable for two reasons. Firstly, causalists regarding explanation allege that all explanantia should cite (at least some of) the explanandum's causes. Secondly, there is widespread agreement among philosophers that manipulation requires causal relations. I will not take up a position regarding the indispensability of causes in either explanation or manipulation in this chaper. But in order not to lose the causalists of explanation and/or manipulation, I will distinguish between causal regularities and non-causal ones.

1.3 Invariant generalizations

The distinction between causal regularities and non-causal ones can be drawn with the help of Jim Woodward's theory. His theory is closely related to the status of causality in the special sciences and the formal framework he adopts is connected to actual statistical techniques (such as structural equation modelling). In Woodward's view, a generalization is causal and explanatory if and only if it is invariant. And it is invariant to the extent that it "remains stable or unchanged as various other changes occur" (Woodward, 2003b, 239). Different senses of invariance can be distinguished (Woodward, 2003b, 242). For example, a generalization can be invariant under changes to variables not occurring in the generalization itself. Or it can be invariant under changes to variables occurring in the generalization itself, where these changes are not the result of interventions (*non-I-changes*). But the most important sense of invariance is invariance under interventions.

Before I present Woodward's framework in more detail, let me give an intuitive example. Consider the relation between 'Physical Exercise' on the one hand, and 'Stamina' on the other hand. 'Physical Exercise' and 'Stamina' can be considered as *random variables*. Intuitively, a random variable represents some feature of an entity or set of entities. Each random variable may have several values. Each value represents a different state that the feature can take. For example, we may distinguish between two possible values for 'Physical Exercise': 'Physical Exercise = less than 2 hours/week', and 'Physical Exercise = more than 2 hours/week'. Likewise, we may distinguish between two possible values for 'Stamina': 'Stamina = bad', and 'Stamina = good'. (Note that variables may have more than two possible values. For the notion of variables and their values, see page 41 and definition 3.19.) Suppose now that there is a correlation between 'Physical Exercise' and 'Stamina' in the Belgian population. For example, that the probability of 'Stamina = good' is higher in the subpopulation for which 'Physical Exercise = more than 2 hours/week' than it is in the subpopulation for which 'Physical Exercise = less than 2 hours/week'. This relation may be stable or invariant under changes to other variables. For example, the relation between 'Physical Exercise' and 'Stamina' may remain unchanged even if Belgium's GNP changes (from 'GNP = n euros' to 'GNP = n' euros'). The relation could also be invariant under changes that are due to a process affecting 'Physical Exercise' and 'Stamina' at the same time (*non-I-changes*). But the interesting point is whether it would be invariant under a manipulation of 'Physical Exercise' that does not influence 'Stamina' – except, perhaps, through 'Physical Exercise'. For example, it is plausible that by increasing the number of gym classes from less to more than 2 hours/week for all pupils, the number

of pupils with good stamina may be raised. By contrast, it is not plausible that directly improving these pupils' stamina, e.g. by administering drugs, would help to increase their amount of physical exercise. This asymmetry Woodward deems the hallmark of causal relations.

When are generalizations invariant under interventions? A necessary condition is that they are change- or variation-relating,[9] "in the sense that they purport to describe a relationship between changes or variations in the value of one or more variables and changes or variations in the values of another variable." (Woodward, 2003b, 245) However, not all change-relating generalizations are invariant under interventions (spurious correlations are not). For example, changes or variations in the barometer reading are accompanied by changes or variations in the weather condition. But the relation between barometer status and weather conditions is not invariant: you cannot influence the weather condition by actively changing or manipulating the barometer's pointer.

When is a change-relating generalization G between X and Y invariant under interventions? Interventions are informally defined as follows:

> an intervention on some variable X with respect to some second variable Y is a causal process that changes the value of X in an appropriately exogenous way, so that if a change in the value of Y occurs, it occurs only in virtue of the change in the value of X and not through some other causal route. (Woodward, 2003b, 94)

The precise definition of *intervention* is as follows:

(IN) I's assuming some value $I = z_i$, is an *intervention* on X with respect to Y if and only if I is an intervention variable for X with respect to Y and $I = z_i$ is an actual cause of the value taken by X. (Woodward, 2003b, 98)

I is an *intervention variable* for X with respect to Y if it satisfies the following conditions:

(IV) I1 I causes X

[9]The variables in a change-relating generalization are repeatable. A variable is repeatable (or repeatably instantiatable, or type-level) if it can be assigned a value more than once. Repeatable variable are to be contrasted with single-case variables (token level variables) which can only be assigned a value once. (Williamson, 2005, 7) By 'variable' I will always mean 'repeatable variable'.

I2 I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.

I3 Any directed path from I to Y goes through X.[10] That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the $I - X - Y$ connection itself; that is, except for (a) any causes of Y that are effects of X (i.e., variables that are causally between X and Y) and (b) any causes of Y that are between I and X and have no effect on Y independently of X.

I4 I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X.

I5 I does not alter the relationship between Y and any of its causes Z that are not on any directed path (should such a path exist) from X to Y. (Woodward, 2003b, 98–99)

Consider a generalization or equation G relating the variable X to another variable Y. Let the actual value of X be x_0 and let the actual value of Y be $G(x_0) = y_0$. Consider now a (real or ideal) intervention that changes the value of X from x_0 to some different value x_1, where the difference between x_0 and x_1 is such that G predicts that a change will occur in Y (from y_0 to $G(x_1) = y_1 \neq y_0$). G, then, is invariant under this intervention if and only if it is indeed the case that $G(x_1) = y_1$ under this intervention. (Woodward, 2003b, 250)

Invariant relations must be stable under some interventions, but it is perfectly possible that they fail to be stable under others. Some generalizations are more invariant than others, depending on the range and importance of the interventions under which they are invariant. Invariance comes in degrees. Like stability and strength it is a gradual notion. (Woodward, 2003b, 251) But it also involves a threshold. If a generalization isn't stable under any interventions, it is noninvariant, and hence neither causal nor explanatory. (Woodward, 2003b, 248–249)

[10] A directed path is a series of direct causal relations that has no colliders $X \rightarrow Y \leftarrow Z$ or forks $X \leftarrow Y \rightarrow Z$.

1.4 (c)P-laws and -regularities

With the help of Woodward's conceptual framework, Mitchell's concept of pragmatic law can be refined. Admittedly, Woodward defines 'laws' traditionally and he argues that lawfulness isn't of any help regarding scientific explanation (Woodward, 2003b, 166–167). But this does not preclude us from joining the concepts of pragmatic law and invariance. (For more detailed comparisons between both frameworks, see Mitchell (2000, 258–259) and Woodward (2003b, 295–299).) In the following chapters, I will repeatedly use the following four concepts:

(*P*-regularity) A regularity is a pragmatic regularity (a *P*-regularity) if it has some degree of stability and strength.

(*P*-law) A generalization is a pragmatic law (a *P*-law) if it describes a *P*-regularity. It has stability and strength to the extent that the regularity it describes is stable and strong. It allows one to a certain extent to predict, to explain and/or to manipulate the world. It may, but need not, satisfy the criteria for strict lawfulness.

(*cP*-law) A generalization is a *causal* *P*-law (a *cP*-law) if it is a *P*-law and if it is invariant under some range of interventions. It allows one to a certain extent to predict, to explain and/or to manipulate the world.[11] It may, but need not, satisfy the criteria for strict lawfulness.

(*cP*-regularity) A *P*-regularity is a *causal* *P*-regularity (a *cP*-regularity) if it is described by a causal *P*-law.

In the interest of readability I will often write '(*c*)*P*-regularity' (respectively '(*c*)*P*-law') instead of the phrase '*P*-regularity *or* *cP*-regularity' (respectively, '*P*-law *or* *cP*-law').

1.5 Outline

Up till now we have seen that the concept of 'law of nature' occupied a central place in twentieth century philosophy of science, but that the traditional criteria for lawfulness are too strict. They hardly allow to classify generalizations in physics (let alone in biology and the other special sciences) as laws. We have seen that a relaxed notion of 'law of nature', viz. Mitchell's concept of 'pragmatic laws' may be of help. It fits the use of the label 'law' in the theory of classical genetics and it dovetails with actual scientific practice in

[11] I'm assuming that every regularity can be described by some generalization. Note also that *cP*-laws should not be mistaken for *ceteris paribus* laws.

e.g. occupational medicine. Finally, we saw that extra conceptual strength is gained by explicitly distinguishing between causal regularities and generalizations, and non-causal ones. To that end, I briefly presented Woodward's concept of 'invariance'. In section 1.4 I combined 'pragmatic laws' and 'invariance' and introduced the terms 'P-regularity', 'cP-regularity', 'P-law', and 'cP-law'.

It can of course be objected that the definitions of these terms are non-committal. They hardly tell us how these regularities look like, how they can be represented by laws, or how these laws can be expressed. Moreover, the definitions of 'P-law' and 'cP-law' invoke crucial other concepts, such as prediction, explanation and manipulation, but these concepts are left indeterminate. Finally, the definitions of P-regularities and cP-regularities leave open the question how such regularities are discovered. In the following chapters, I will remedy these shortcomings.

In chapters 3 and 4 I will tackle the first shortcoming and show how $(c)P$-laws and the scientific theories in which they figure, can be expressed or represented. In chapter 3, "The Language of Causal Modelling", I will introduce the formal framework of causal modelling. A large part of the definitions presented there are well-entrenched in the literature, but I will add several concepts that will prove valuable in the subsequent chapters. In chapter 4, "The Causal Structure of Classical Genetics", I will use this framework to analyse and represent the laws and theory of classical genetics in a way that resembles (but not strictly follows) the structuralist approach of Balzer et al. (1987). It will be seen how the law of segregation, the law of independent assortment, linkage, crossing-over, etc. fit together in this causal structure. We will also see that they all are faced with exceptions and lack strict universality, that the history of classical genetics is replete with changes to account for these anomalies, that the theory of classical genetics consists of a number of theory-elements that are the product of a set of exemplars and that they each have a limited intended domain of application, etc.

In chapters 5 and 6 I will tackle the second shortcoming. How can non-strict laws be used in prediction, explanation or policy? I will not discuss the case of prediction since I am convinced that the central problems involved in prediction from non-strict laws also surface in the case of explanation and the case of policy. In chapter 5, "Explanation in Genetics: a Many-Headed Monster", I will show how the laws of classical genetics can be used for explanation and thus meet one of major requirements of the pragmatic approach to laws of nature. Several types of explananda were addressed in classical genetics. I will mostly zoom in on the explanation of phenotypic distributions. By calling explanation in genetics a many-headed monster, I mean that it incorporates many of the central aspects often attributed to explanation in the

literature. The explanation of phenotypic distributions can be considered a kind of *deductive-statistical explanation* in which heavy stress is laid on the role of *causal* generalizations (and in which the straightjacket of the received view's criteria for strict lawfulness is of course relinquished). In other words, phenotypic distributions are explained by means of genetics' causal P-laws. The cP-laws that are invoked together describe a '*complex-system mechanism*', and thus provide a kind of mechanistic explanation. Finally, (explanation on the basis of) the theory of classical genetics provides *ontologically grounded derivational unification* as a bonus. I will conclude by examining other types of explananda, such as singular events or the nature of the relation between genotypes and phenotypes. In chapter 6, "P-laws, cP-laws and Policy", I will address the question to what extent P-laws and cP-laws can be used for policy and thus meet another major requirement of the pragmatic approach. In the past few decades, consensus has grown in the literature that causal relations are a *conditio sine qua non* for manipulation. But then it can be asked whether non-causal P-laws (more precisely, spurious correlations) are of any use in policy. If not, it would seem they have little value apart from their evident applicability for prediction. But then the lack of any clear distinction between causal pragmatic laws and non-causal ones should be counted heavily against Mitchell's framework and it would become hard to see what this framework adds to the theory of Woodward.

In chapters 7, 8, and 9 I will tackle the third shortcoming. How are pragmatic regularities discovered? The discovery of causal and non-causal relations is studied from many different angles in the philosophical (and the non-philosophical!) literature. Some recurrent themes are: the superiority of experimental over merely observational data for causal discovery, the development of algorithms to sidestep this difficulty in case only non-experimental data are available, the use of reliable background knowledge to complement these algorithms, and the omnipresence of statistical techniques in the discovery of causal and non-causal relations. In chapter 7, "Experiments in Classical Genetics", I will show that the claims of the leading geneticists notwithstanding, real experiments were very rare in the history of classical genetics. This emerges from the distinction between experimental designs and e.g. prospective designs as it is made in contemporary methodology of many special sciences (the social sciences, epidemiology, ...). It is of course not my intention to chide these geneticists for their misapprehension. I rather want to show how this hampered causal discovery and how it kindled objections by opponents of Mendelian genetics. In chapter 8, "Causal Discovery and the Problem of Ignorance. An Adaptive Logic Approach", I will further discuss the topic of causal discovery from non-experimental data. Instead of reviewing all the algorithms that are currently on the market, I will focuss on

one of them: Judea Pearl's **IC**-algorithm. The **IC**-algorithm is well-known and its merits can hardly be overrated, but it cannot handle the *problem of ignorance*, a problem that is both ubiquitous, persistent and far-reaching in scientific practice. In short, this problem comes down to the following: for some pairs of variables A and B, and for some disjoint (and possibly empty) sets of variables $\mathbf{Q}$, it is not known whether or not A and B are independent conditional on $\mathbf{Q}$. I will also present an adaptive logic, **ALIC**, which is based on the **IC**-algorithm and which adequately handles cases of ignorance. The existing algorithms for causal discovery, such as **IC**, and the existing methods for the discovery of causal and non-causal relations in general, rely heavily on – and are strongly tied to – statistical techniques. This is fine, but it also requires that these techniques are themselves as neutral as possible regarding their domain of application (i.e. the theory being developed or tested). In chapter 9, "Galton's Blinding Glasses. Modern Statistics Hiding Causal Structure in Early Theories of Inheritance", I will argue that present-day statistical techniques such as structural equation modelling may indeed be perniciously biasing. (From this it should not be concluded that these techniques are to be renounced, only that we should apply them carefully.) My arguments are based on a case study, viz. Galton's theory of ancestral inheritance. Although Francis Galton used (and introduced!) more novel statistical techniques than Mendel, his theory of heredity was surpassed by the latter's. I will show that his failure was due (at least in part) to his knowledge of statistical techniques. Finally, in chapter 10, "Concluding Remarks: the Biassing Role of Causal Models", I will examine whether and to what extent the language and inferential framework of causal modelling may be biassing as well. It may be feared that they constrain both the discovery of causal relations and our concepts of causality and laws of nature in the special sciences.

First, however, I will consider one major alternative, or alleged alternative, for laws of nature. In the past decades, the concept of 'complex-system mechanism' has been repeatedly put forward in philosophy of science as an alternative to laws of nature, especially with an eye to explanation in the life sciences. In chapter 2, "Can Mechanisms Really Replace Laws of Nature?", I will argue that as valuable as the concepts of mechanism and mechanistic explanation are, they cannot replace regularities nor undermine their relevance regarding explanation. It goes without saying that this argument is based on a relaxed notion of laws and regularities, viz. $(c)P$-laws and -regularities.

Chapter 2

Can Mechanisms Really Replace Laws of Nature?

Today, complex-systems mechanisms and mechanistic explanation are very popular in philosophy of science and are deemed a welcome alternative to the decried laws of nature we met in section 1.1 and to D-N explanation.[1] However, starting from the relaxed, pragmatic notion of regularities and laws of sections 1.2, 1.3 and 1.4, I will cast doubt on their status as a genuine alternative. I will argue that (1) all complex-systems mechanisms ontologically must rely on stable regularities, while (2) it is not obvious that all such regularities must rely on an underlying mechanism. Analogously, (3) models of mechanisms must incorporate (and hence are epistemologically dependent on) pragmatic laws, while (4) such laws themselves needn't always represent mechanisms. As valuable as the concepts of mechanism and mechanistic explanation are, they cannot replace regularities nor undermine their relevance regarding explanation.

2.1 Introduction

Today, mechanisms and mechanistic models are very popular in philosophy of science, in particular in philosophy of the life sciences. Mechanicist philosophers like Machamer et al. (2000) and like Bechtel and Abrahamsen (2005)

[1] In deductive-nomological explanation (D-N explanation), the *explanandum* is either a sentence describing a particular fact, or a general law (a universal conditional in first order logic, i.e. a strict law). The *explanans* must consist of (true!) sentences describing particular facts and at least one general law and it must have empirical content (i.e., it must be capable, at least in principle, of test by experiment or observation). Finally, the explanandum must be a logical consequence of the explanans. (Hempel, 1965, 247–249, 336)

set their face against the dominant position that strict laws of nature and D-N explanation have occupied for years on end. (As I stated in chapter 1, by 'law of nature' I mean a generalization *describing* a regularity, not some metaphysical entity that *produces* or *is responsible for* that regularity.) Their opposition is not groundless. The criteria for lawfulness that have been advanced by e.g. Nagel (1961), Hempel (1965), and Goodman (1973) and that are considered the received view are highly problematic. If we would hold on to these criteria, we should classify almost all scientific generalizations as contingent, accidental generalizations. This holds for the special sciences, such as biology and the social sciences, as well as for chemistry and physics. If there are no strict laws, there are no D-N explanations in Hempel's sense. Hence the mechanicist alternative, which states that explanation involves *mechanistic models* (i.e. descriptions of *mechanisms*) instead of strict laws, might be very welcome.

As we saw in section 1.2, the received view has been attacked from other sides as well. Instead of abandoning the concept of law of nature, Sandra Mitchell (1997, 2000) proposed to revise it. In her view, laws of nature should be interpreted pragmatically. A generalization is a *pragmatic law* if it allows of prediction, explanation and/or manipulation, even if it fails to satisfy the traditional criteria. To this end, it should describe a *stable regularity*, but not necessarily a universal and necessary one.

The precise relation between mechanisms and stable regularities, or between mechanistic models and pragmatic laws is still an open question, which I will address in this chapter. In sections 2.2 and 2.3 I will briefly present mechanisms and mechanistic models and raise the question whether mechanisms really are an alternative to regularities. In section 2.4 I will briefly revisit pragmatic laws and their corresponding regularities. Together, sections 2.2, 2.3 and 2.4 set the stage for the arguments presented in the rest of this chapter, where I will make four related claims – two ontological ones (section 2.5) and two epistemological ones (section 2.6). Firstly, mechanisms are ontologically dependent on stable regularities. There are no mechanisms without both macro-level and micro-level stable regularities. Secondly, there may be stable regularities without any underlying mechanism. Thirdly, models of mechanisms are epistemologically dependent on pragmatic laws. To adequately model a mechanism, one has to incorporate pragmatic laws. Finally, pragmatic laws are not epistemologically dependent on mechanistic models. It is possible to gain evidence for the stability of some regularity, and hence for the pragmatic lawfulness of the corresponding generalization, without relying on mechanistic background knowledge. Pragmatic laws thus needn't incorporate a description of a mechanism underlying the regularity at hand. In section 2.7 I will conclude by showing what are the implications for

the status of mechanistic explanation. As valuable as the concepts of mechanism and mechanistic explanation are, they cannot replace regularities nor undermine their relevance for scientific explanation.

2.2 Mechanisms

From the end of the 1970's onwards, the concept of 'mechanism' has become popular in philosophy of science. Different families of concepts can be distinguished. In the Salmon/Dowe account, mechanisms are characterized in terms of causal processes and causal interactions. Here I will not consider this account. Rather, I will focus on the complex-systems approach defended by i.a. Glennan, Woodward, Machamer et al., and Bechtel and Abrahamsen. In this approach, mechanisms are treated as complex systems of interacting parts. Contrary to Salmon/Dowe-mechanisms, complex-systems mechanisms (henceforth *cs*-mechanisms) are robust or stable. They form stable configurations of robust objects and as a whole they have stable dispositions: the behaviour they display (see Glennan, 2002, S344–S346). This difference will prove relevant in the following sections.

In this section, I will focus on the theories of Machamer et al. (2000) and of Bechtel and Abrahamsen (2005). (As I will indicate in section 2.7, my findings can be extended to e.g. Glennan (1996, 2002) and Woodward (2002b).) Both theories are mainly concerned with the life sciences and they both present mechanisms and mechanistic explanation as an alternative to strict laws of nature and D-N explanation.

In their influential paper, "Thinking about mechanisms", Peter Machamer, Lindley Darden and Carl F. Craver define mechanisms as complex systems:

(M*) [*cs*-]Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions. (Machamer et al., 2000, 3)

Entities are the things that engage in activities. Activities are the producers of change. The authors defend a dualistic metaphysics that combines substantivalist notions with concepts of process philosophy. Entities and activities, they claim, are complementary, interdependent concepts. (Machamer et al., 2000, 4–8) If entities and activities are adequately organized, they behave regularly.

William Bechtel and Adele Abrahamsen's definition of mechanisms is somewhat different, but it portrays mechanisms as organized, complex systems too.

(M†) A [*cs*-]mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena. (Bechtel and Abrahamsen, 2005, 423)

This theory strongly resembles the previous one. The component parts are clearly entities. Operations are strongly related to activities. (Bechtel and Abrahamsen (2005, 423) use the label 'operation' instead of 'activity' because they wish to draw attention to the involvement of parts). And the mechanism's parts and activities must be organized or orchestrated. But Bechtel and Abrahamsen add to **(M*)** the notion of function (and thus some kind of design). (Bechtel and Abrahamsen, 2005, 423–424)

Both theories stress the role of mechanistic models in explanation. What is important is that they do so *at the expense of strict laws*. They signal several distinct but related problems with respect to strict laws and D-N explanation. Firstly, attention is drawn to the fact that strict regularities are rarely if ever discovered in the life sciences. But if strict biological laws are rare or nonexistent, D-N explanations would not be practicable in the life sciences. Secondly, Bechtel and Abrahamsen append to this that their account avoids some hard ontological problems. Staking on mechanisms as real systems in nature, they write, has the advantage that "one does not have to face questions comparable to those faced by nomological accounts of explanation about the ontological status of laws." (Bechtel and Abrahamsen, 2005, 425) Thirdly, it is stressed that, even if there would be strict biological laws, there is no denying that explanation in the life sciences usually takes the form of mechanistic explanation. Bechtel and Abrahamsen write that (see also Bechtel and Richardson, 1993, 231):

> Explanations in the life sciences *frequently* involve presenting a model of the mechanism taken to be responsible for a given phenomenon. Such explanations depart in numerous ways from nomological explanations commonly presented in philosophy of science. (Bechtel and Abrahamsen, 2005, 421; my emphasis)

Machamer, Darden and Craver make a somewhat stronger claim:

> In many fields of science what is taken to be a satisfactory explanation requires providing a description of a mechanism. (Machamer et al., 2000, 1)

Fourthly, both groups of authors argue that even if there would be strict biological laws, D-N explanations would not be sufficiently explanatory. Explanation, they say, involves more than subsumption under a law or regularity. Laws or regularities do not explain *why* some phenomenon occurs.

Machamer et al. have it that activities are essential for rendering phenomena intelligible. A mechanistic explanation makes a phenomenon intelligible by providing an elucidative relation between the explanans and the explanandum, i.e. by revealing the productive relation between the mechanism's set-up conditions, intermediate stages and termination conditions. This productive relation is completely accounted for by the mechanism's activities: "It is not the regularities that explain but the activities that sustain the regularities." (Machamer et al., 2000, 21–22) They append to this that "[...] regularities are non-accidental and support counterfactuals to the extent that they describe activities. [...] No philosophical work is done by positing some further thing, a law, that underwrites the productivity of activities. (Machamer et al., 2000, 7–8) (Terminological prudence is in order here. In my terms regularities are ontological and cannot *describe* activities. And I do not adhere to laws as metaphysical entities that *underwrite* the productivity of activities.) According to Bechtel and Abrahamsen, subsumption under a law does not show why the explanandum phenomenon occurred.

> Even if accorded the status of a law, this statement [a statement concerning the ratio of oxygen molecules consumed to ATP in metabolism] merely brings together a number of actual and potential cases as exemplars of the same phenomenon and provides a characterization of that phenomenon. However, it would not explain *why* the phenomenon occurred – either in general or in any specific case. (Bechtel and Abrahamsen, 2005, 422)

To explain *why*, scientists (biologists) explain *how*. They provide a model of the mechanism underlying the phenomenon in question.

In short, $(\mathbf{M}^*)$ and $(\mathbf{M}^\dagger)$ are motivated by the apparent shortcomings of the concepts of strict law/regularity and D-N explanation (in the context of the life sciences). Mechanisms and mechanistic explanation are then put forward as an alternative to these problematic concepts. I will side with the mechanicists in their critical assessment of both strict laws/regularities and D-N explanation. I will also endorse the view that 'mechanism' and 'mechanistic explanation' are very fruitful concepts. But I will doubt whether mechanisms and mechanistic models are an *alternative* to regularities and laws.

2.3 Are mechanisms an *alternative* to regularities?

In this section I will show that both **(M*)** and **(M†)** depend on the concept of 'regularity' – at least *prima facie*. In section 2.5 I will argue that this is no coincidence: *cs*-mechanisms are ontologically dependent on the existence of regularities.

In definition **(M*)**, regularities are mentioned explicitly: mechanisms, it says, are productive of *regular changes*. In definition **(M†)**, regularities are not referred to explicitly. However, it states that mechanisms perform a *function*. Functions are best conceived of as dispositions. Dispositions always involve regularities. Hence, **(M†)** implicitly refers to regularities.

Functions are dispositional in two ways. The first way is that proposed by the dispositional theory of functions (e.g. Bigelow and Pargetter, 1987).

(DTF) An effect e of a character c is a function of that character if it confers a survival-enhancing *propensity* on the creature having c.

Bigelow and Pargetter interpret propensities dispositionally. It is not required, however, that e enhances survival (and/or reproduction) in all individuals all of the time. The dispositional theory of functions is not unquestioned, however. The main alternative is the etiological theory of functions (cf. Mitchell, 2003a, 92).

(ETF) An effect e of a character or component c is a function of that character or component if it has played an essential role in the causal history issuing in the presence of that very component.

During this causal history, c must have been selected over alternatives on the basis of its doing e, and it must have been produced or reproduced as a direct result of that selection process (Mitchell, 2003a, 96). By its reference to natural selection, the etiological theory links functions to fitness, which is a dispositional characteristic. So either way, functions are dispositional. (See also Walsh (1996) who proposes a relational account of functions which should cover both the dispositional and the etiological account.)

The second way in which functions are dispositional is compatible with both accounts of functions. Even if the function of x is to do f, it is not required that x does f all the time. The (or a) function of my stomach is to digest food, even if I haven't eaten for two days. The (or a) function of my legs is to allow me to walk, even if I'm sitting in a chair. This is true regardless of the reasons we have for attributing them these functions (i.e.

whether we refer to survival-enhancing propensities or to the causal history of organisms having stomachs or legs).

Since functions are dispositions, they presuppose the existence of regularities. Even if there is no consensus about the correct analysis of dispositions, all attempts seem to have in common that dispositions involve regularities. (For an overview of the most prevalent definitions of 'disposition', see Fara, 2006.) Roughly, a disposition can be characterized as follows:

(DISP) An object is disposed to M when C iff, if it were the case that C and Ψ, then it would $\phi(M)$.

M refers to a manifestation, C to the conditions of manifestation. In the case of a fragile glass, M could be 'breaking', and C could be 'being struck'. Ψ stands for the extra conditions that should be included in the definition or analysis of dispositions. The simple conditional analysis, which leaves Ψ empty, is victim to several counterexamples and a large part of the literature about dispositions concerns the question what other conditions should be included in Ψ. (For example, David Lewis has suggested that an object is disposed to M when C if and only if it has an intrinsic property B such that, if it were that C, and if the object were to retain B, then the object would M because C and because it has B; see Fara, 2006, section 2.3.) The ϕ-operator stands for the modal or probabilistic strength that should be included in the definition of dispositions. According to the simple conditional analysis, the object should *always* M if it were that C. Again, this makes the simple conditional analysis victim to several counterexamples. Therefore, it has been proposed to interpret ϕ less strictly, viz. *habitually* (Fara, 2006, section 2.4) or *probabilistically* (Prior et al., 1982). What is relevant for the present discussion is the following: even if we would allow for dispositions that are seldom manifested when their manifestation conditions C obtain, ϕ cannot be replaced by 'never', since this would result in a *contradictio in terminis*. If the conditions in Ψ are satisfied, $P(M \mid C) > 0$ (where M and C denote events, not variables).[2]

[2]If there are dispositions that are seldom manifested when their manifestation conditions C hold, the following may serve as an example. A lottery is being held with 1.000.050 tickets. The tickets range from 1 to 1.000.000, but there are 51 tickets with the number 666. In this case, one might say, the lottery is disposed to select 666 as the winning number when a draw is made, even if the chance of selecting 666 is very low ($\frac{51}{1.000.000}$).
I do not claim that such dispositions exist, and hence I do not claim that this lottery should be ascribed the disposition to have 666 as its outcome. What I want to claim is the following: even if our definition of dispositions is so liberal that ϕ might refer to very low probabilities, dispositions still depend on the existence of regularities. Only, in this case the regularities involved would have very limited strength, to be compared with e.g.

So far we can safely conclude that both Machamer et al., and Bechtel and Abrahamsen define mechanisms in terms of regularities (either explicitly or implicitly). This raises a first question, viz. whether this use of 'regularity' is necessary or unavoidable. This question will be answered in section 2.5. It should also be noticed that neither gives a minute characterization of these regularities. They only describe them negatively: they are *not* strict. This raises a second question which I will answer very briefly in section 2.4, viz. how these regularities should be conceived of. Regularities are a blind spot in the mechanistic literature. This blind spot can be removed by means of a more adequate theory of regularities and lawfulness. It should be no surprise that the concepts of chapter 1 will come in useful.

2.4 $(c)P$-laws and -regularities shortly revisited

Instead of rejecting the concept of law of nature, Sandra Mitchell (1997, 2000) sets out to refine it and to approach it pragmatically (cf. section 1.2). To count as a pragmatic law, a generalization need not satisfy the traditional criteria for strict lawfulness. It need only be useful for prediction, explanation and/or manipulation. Pragmatic laws may differ from each other in many ways. They may differ qua stability or strength (both are important gradual ontological parameters), or qua cognitive manageability, simplicity, and other gradual representational parameters.

Jim Woodward (2003b) also proposes an alternative to strict laws of nature (cf. section 1.3). He advocates the concept of invariant generalizations, i.e. generalizations that remain stable or unchanged under interventions. Invariance, too, is a gradual concept. Woodward considers invariance under interventions the *conditio sine qua non* for causal and explanatory relations.

The concepts of both Mitchell and Woodward are intended to better fit scientific practice. On the basis of these concepts, I introduced the notions of P-regularity, cP-regularity, P-law, and cP-law (cf. section 1.4).

Up till now we have seen that whereas *cs*-mechanisms are put forward as an alternative to strict regularities (section 2.2), they are nevertheless defined in terms of regularities (section 2.3). Now the question regarding the precise relations between *cs*-mechanisms and $(c)P$-regularities, and between mechanistic models and $(c)P$-laws can be addressed.

---

the regularity relating syphilis to paresis. (For the concept of strength of regularities, see section 1.2.)

2.5 The ontological relations between mechanisms and (causal) P-regularities

In this section, I will first argue that complex-systems mechanisms are ontologically dependent on (causal) P-regularities. No x can count as a mechanism, unless it involves regularities. Then I will investigate the reverse relation, i.e. whether there can be (causal) P-regularities without any underlying mechanism.

Mechanisms are ontologically dependent on the existence of regularities both at the macro-level and at the micro-level. Firstly, no x can count as a *cs*-mechanism, unless it produces some macro-level regular behaviour. Secondly, to produce such macro-level regular behaviour, this x has to rely on micro-level regularities.

In the life sciences, reference to mechanisms cannot be detached from matters of projectibility. Thomas Hunt Morgan and his co-workers sought after the mechanism of Mendelian heredity to explain both Mendel's findings and their exceptions in a systematic way (Morgan et al., 1915; Morgan, 1919, 1928).[3] Mainly drawing from findings in fruit flies, they explained definite macro-level behaviours (definite phenotypic ratios in subsequent generations of organisms) by referring to the behaviours (independent assortment, crossing-over, interference, ...) of a complex set of parts or entities (gametes, chromosomes, factors or genes, ...). But Morgan et al. were not only interested in the fruit flies in their laboratories. They were interested in the mechanism of heredity in *Drosophila* and in other species as well. As evidence accumulated, both Mendelian inheritance and the underlying chromosomal mechanism were more and more considered a general phenomenon. In the end, Morgan formulated the theory of the gene (including Mendel's two laws) without reference to any specific species (Morgan, 1928, 25).[4] He likewise gave an abstract mechanistic explanation (Morgan, 1928, chapter III). The case of T.H. Morgan illustrates not only that talk in terms of laws is compatible with talk in terms of mechanisms, but also that reference to mechanisms in the life sciences cannot be detached from matters of projectibility. Due to this concern for projectibility, Glennan (2002, S345) stresses that the behaviour of complex-systems mechanisms as a whole should be stable.

At this point, the reader might worry that metaphysical issues (about what a mechanism is) get conflated with epistemological ones (about the use of mechanistic knowledge). Such worry would be baseless. It is not that our concern for projectibility implies that mechanisms should be stable or robust.

[3]The status of mechanisms in classical genetics is ambiguous, however (see section 5.3).

[4]Morgan's formulation of the theory of the gene is quoted in section 4.2.

Rather, it implies that life scientists should search for robust mechanisms (it is a matter of fact that, to phrase it naively, they succeed in this). And if the concept of 'complex-systems mechanism' is to fit scientific practice (as is argued by Machamer et al. (2000, 1–2) and Bechtel and Abrahamsen (2005, 422)), it must incorporate this notion of stability. But, *per definitionem*, this comes down to the following:

(H-REG) There can be no *cs*-mechanism without some higher level (causal) P-regularity (viz. the stable behaviour produced by that mechanism).

In section 2.3 I showed that **(M*)** and **(M†)** conform to **(H-REG)**. Following **(M†)**, mechanisms perform a function. They have a dispositional property which *ϕ-regularly* results in M if the conditions C and Ψ are satisfied (see p. 23). Even very weak dispositions (see footnote 2) can be accounted for by the concept of (causal) P-regularity. Following **(M*)**, mechanisms are *productive* of *regular changes from* start or set-up *to* finish or termination conditions: they exhibit causal P-regularities.

Bogen (2005, 398–400) has criticised Machamer et al. for unfoundedly providing a regularist account of causation and Machamer (2004, note 1) has sided him. According to regularism, there is no causation without regularity. By contrast, Bogen argues for an Anscombian account in which causality is one thing, and regularity another. From this he concludes that mechanists needn't invoke regularities or invariant generalizations. Some causal mechanisms, he states, are too unreliable to fit regularism.

> The mechanisms which initiate electrical activity in post-synaptic neurons by releasing neuro-transmitters are a case in point. They are numerous enough, and each of them has enough chances to release neurotransmitters to support the functions of the nervous system. *But each one fails more often than it succeeds, and so far, no one has found differences among background conditions which account for this* [...]. (Bogen, 2005, 400, my emphasis)

In this chapter I will not address the nature of causation, and hence I do not want to argue pro or contra regularism. But I dismiss the conclusion Bogen draws regarding mechanisms. His example fails to show that some *cs*-mechanisms go without regularities. It only illustrates that some go with regularities that are very weak and of (as yet) poorly specified stability.

Let us now turn to the *cs*-mechanism's micro-level dependence on (causal) P-regularities. A mechanism's behaviour is not groundless. It is produced by its component parts. Suppose now, that some part p_i behaves completely irregularly: it may do a_{i1}, or a_{i2}, or ..., or a_{in}, or ... but what it does

is the result of a completely random internal process. There is no relation whatsoever to the behaviour of the other parts p_j of the mechanism, nor to the previous behaviours of p_i itself. And there isn't even a stable probability distribution over the behaviours $a_{i...}$. Suppose moreover, that the same holds for all the other parts of the mechanism. Clearly, this would make it very unlikely for the mechanism to produce a macro-level P-regular behaviour, let alone a causal P-regularity. So unless the behaviour of its parts is sufficiently stable and sufficiently strong, i.e. unless it is P-regular, and unless these behaviours are organized sufficiently well, the mechanism's overall behaviour will fail to be P-regular. (I do not rule out that some of the mechanism's parts behave randomly. Only, then, sufficiently many other parts should behave P-regularly and their behaviour should be organized sufficiently well.)

(L-REG) There can be no *cs*-mechanism without some lower level (causal) P-regularities (viz. the regular behaviours, operations, or activities displayed or engaged in by the mechanism's parts).

Again, this is stressed by Glennan (2002, S344): a mechanism's parts must be objects – in the absence of interventions, their properties must remain relatively stable. Translating this to $(\mathbf{M}^*)$ and $(\mathbf{M}^\dagger)$, these parts' activities or operations must be causal P-regularities.

Up till now, I have shown that mechanisms always involve both macro- and micro-level regularities. But what about the reverse relation? Can there be a (causal) P-regularity without an underlying mechanism? In other words, can there be fundamental regularities whose stability and strength are somehow *sui generis*? Glennan (1996, 61–63) assumes or stipulates they exist. That is more than I need. Since the concept of *cs*-mechanism entails the concept of (causal) P-regularity, it suffices for me that their existence is logically possible. It might be the case that, as a matter of fact, all (causal) P-regularities rest on some (hitherto unknown) underlying mechanism – I see nothing metaphysically wrong with an infinite ontological regress of mechanisms and regularities. What matters is that the concept of fundamental P-regularity/law is coherent. If there are such fundamental P-regularities and P-laws, and if we want to explain everything, we should seek for (or hope for) non-mechanistic forms of explanation.

2.6 The epistemological relations between mechanistic models and (causal) P-laws

The mechanicist may endorse the conclusion of the former section, while still holding exclusively to mechanistic explanation in biology and the biomed-

ical sciences. After all, why should biologists care about non-mechanistic forms of explanation in physics? (Let us assume, for the sake of the argument, that the fundamental regularities, if any, are or will be described by physics.) Drawing on the findings of the previous section, I will show that mechanistic explanation cannot dispense with (causal) *P*-laws. To adequately describe *cs*-mechanisms, mechanistic models need incorporate – and thus are epistemologically dependent on – (causal) *P*-laws. By contrast, a generalization may count as a *P*-law without describing or referring to any underlying mechanism.

A large part of the complex-systems literature about mechanisms, especially the contributions by Machamer et al. and by Bechtel and Abrahamsen, is motivated by the failure of the D-N model to provide an adequate account of scientific explanation (see section 2.2). Explanation, especially in the life sciences, rarely if ever involves subsumption under strict laws. Far more often it takes the form of mechanistic explanation: one models or describes the mechanism underlying the explanandum phenomenon.

This raises the question what criteria a model should satisfy to count as a model of a *cs*-mechanism. The trivial answer is that it should adequately represent that mechanism. As a mechanism is a system producing some behaviour on the basis of the organized operations of its component parts, the less trivial answer is that it should adequately represent (i) the macro-level behaviour, (ii) the mechanism's parts and their properties, (iii) the operations they perform or the activities they engage in, and (iv) the organization of these parts and operations. Let us call this the *adequacy criterion* for mechanistic models (see also Craver, 2006, 367–373). So, by section 2.5, the model should adequately describe both the macro-level and the micro-level (causal) *P*-regularities. Hence, by definition, it should incorporate (causal) *P*-laws. Thus the adequacy criterion implies that all mechanistic models must incorporate (causal) *P*-laws.

But then the following question arises. Is it possible to gain evidence for a generalization's lawfulness without relying on mechanistic background knowledge? Can one be convinced that some generalization describes a regularity that is sufficiently stable and/or strong (for some particular application context) and can one assess this stability and/or strength without any evidence for some underlying mechanism? In short, can a generalization count as a (causal) *P*-law without referring to mechanisms? To be sure, this question isn't idle and moreover it has large epistemological import. It isn't idle since mechanistic background knowledge is useful in assessing the lawfulness of regularities (as I illustrated with the case of T.H. Morgan, see page 25), and is used so in many different scientific disciplines. It has large epistemological import since, given what we know from the first part of this section,

the epistemological dependence of (causal) P-laws on mechanistic models would imply an infinite (and perhaps vicious) epistemological regress. To be sure that some model M is a model of a *cs*-mechanism, I would need to know that the generalizations $G_1, \ldots, G_n$ figuring in it are (causal) P-laws. But then I would have to know the underlying mechanisms ...

It certainly doesn't do to rely on the existence of fundamental laws. Firstly, I have argued for the logical possibility of fundamental regularities, but it is still an open question whether they actually exist. Secondly, granted that there are fundamental regularities, little or no practising biologists would turn to fundamental laws in explaining biological phenomena.

Machamer, Darden and Craver have tried to solve the problem of infinite regress by introducing the notion of 'bottoming out'. In their conception, nested hierarchical descriptions of mechanisms bottom out in lowest level mechanisms. These are but fundamental relative to the purposes of a given scientist, or discipline, ... Their entities and activities are taken to be fundamental (and hence as not calling for further explanation) relative to the research question at hand, even if it is known that in other scientific fields they are considered as non-fundamental, macro-level phenomena. (Machamer et al., 2000, 13) Although this notion nicely fits scientific practice, it offers at best a pseudo-solution to our problem. It only assures us that, if there is an infinite regress, it will do no harm. In the rest of this section I will face the problem head on and show that (causal) P-laws are not epistemologically dependent on mechanistic models. Mechanistic knowledge is not indispensable for the assessment of a generalization's lawfulness. Other means do at least as well.

A most natural candidate is performing experiments. Experiments are often ascribed the power to reveal causal connections and to confirm or refute claims about stable regularities, even if the relation between experiments and laws or theories is fraught with several problems (see Franklin, 1995, 2003).[5]

[5]In Franklin (1995, 196–204) three problems are discussed. The first is known as the 'theory-ladenness of observation'. Observation statements and measurement reports use terms whose meanings are determined by a particular theory. (This problem may be generalized. Not only the interpretation of experimental results may be theory-laden. The realization of an experiment often also depends on theoretical insights about the experimental (object-apparatus) system and the possible interactions with its environment. Prior knowledge is needed about the object under study and about the instruments used. (Radder, 2003b, 165, 168–169)) The second problem is the 'Duhem-Quine problem'. If some hypothesis h generates a prediction e, it always does so together with some background knowledge b. Hence, if $\sim e$ is observed instead of e, either h is to be blamed, or b, or both. So one can always save h by blaming only b. The third problem is the fact that experiments are fallible and that different experimental results may discord. Franklin concludes that although these problems are important and impel us to treat experimental

Moreover, experiments are very frequently performed in biology and the biomedical sciences. The question now is to what extent stable regularities may be experimentally discovered or established, *without any knowledge of some underlying mechanism*. I will start by giving a very general characterization of experiments.

(EXP) In an experiment, an *object* is placed in some *controlled environment*. It is *manipulated*, often using some *apparatus*, such that it assumes some definite property $X = x$. Then, again using some apparatus, the outcome is *measured* in some (other) property Y. More specifically, it is verified whether there is some relation between $X = x$ and $Y = y$ (for some or all possible values x of X and y of Y), and if so, what is its strength and how it can be characterized.[6]

Let me briefly dwell on this description. The term 'object' has to be interpreted as broadly as possible. It may refer to one particular material object, or to some complex of objects, or to some sample of liquid or gas, etc. An environment is 'controlled' if the relation between X and Y is not influenced or disturbed by other factors. To eliminate all possible disturbing factors (and all possible sources of error in general) is a very delicate and difficult task, a large part of which depends on statistical analysis and data-reduction (cf. Galison, 1987; Franklin, 1990). I will return to this issue in a moment. Emphasis is laid on 'manipulation' since this, much more than passive observation, is considered a particularly reliable way to find out causal relationships.[7] (The role of manipulation in experiments will be found to be significant with respect to the status of genetic crosses in classical genetics – see chapter 7.) Finally, apparatus are often indispensable in experimental designs. Sometimes these are relatively simple, but often they are very complex. They play at least three different roles: as a device for manipulation, for measurement, or to control disturbing influences. (In Radder (2003a), the role of technology and instruments in experiments is discussed several times by many different authors.)

This characterization, the mechanicist may argue, clearly reveals the use of mechanistic background knowledge in experimentation. If you want to

results carefully, they are not insuperable. Experimental evidence may serve to test laws and theories.

[6]For the notion of variables and their values, see page 41 and definition 3.19.

[7]Woodward (2003a) heavily stresses the connections between experimentation and manipulation on the one hand, and causation on the other hand. In his view, experiments not only are an excellent tool for causal discovery and causal inference. To say that X causes Y also "means nothing more and nothing less than that if an appropriately designed experimental manipulation of X were to be carried out, Y (or the probability of Y) would change in value." (Woodward, 2003a, 90, I have slightly changed notation)

create a controlled environment and rule out all disturbing influences, much is gained to know what these influences are. Such knowledge, furthermore, is outstandingly provided by mechanistic models. I endorse this claim, but challenge that it is noxious for my argumentation. Mechanistic background knowledge is highly valuable for the experimenter. But it certainly isn't indispensable.

In many experiments, viz. in *randomized experimental designs*, disturbing influences are not screened off physically. Instead, experimenters endeavour to cancel out their influence by means of *randomization*. From the target population P a sample S is randomly selected. The random sampling procedure should guarantee that the subjects in S do not drastically differ from the rest of the subjects in P. In other words, for any variable Z, its distribution in S should not deviate drastically from its distribution in P.[8] Then the subjects in S are randomly divided into an experimental group S^X and a control group S^K. All subjects in S^X are manipulated such that they assume some definite property $X = x$, whereas those in S^K are not so manipulated ($X = \sim x$) – often they are given a placebo. This procedure should guarantee that the subjects in S^X and S^K most closely resemble each other, except with respect to the cause variable X and its effects. (Instead of having only $X = x$ and $X = \sim x$, one may also create several experimental groups, each with a different level of X.) Then the relation between X and the effect variable Y is measured.

Randomization is a technique that is highly context-independent. It allows to control for disturbing influences *without even knowing them*. This clearly shows that mechanistic background knowledge is no *conditio sine qua non* for experimentation and hence is not indispensable regarding the assessment of the lawfulness of generalizations. Fortunately, we escape the problem of infinite epistemic regress.

2.7 Conclusion: can mechanisms really replace laws of nature?

In this chapter, I have substantiated the following claims. Firstly, *cs*-mechanisms as defined in $(\mathbf{M}^*)$ or $(\mathbf{M}^\dagger)$ necessarily involve regularities. Their macro-level behaviour is regular and it is based on the micro-level regular behaviours (the activities or operations) of the component parts. These regularities needn't be deterministic, nor necessary; i.e. they needn't be strict.

[8] In practice this is not guaranteed. Randomization only works in the limit as sample size tends to infinity.

It suffices that they have some stability and strength, that they are (causal) *P*-regularities. Secondly, even if it cannot be ruled out that all (causal) *P*-regularities involve an underlying mechanism, it is logically possible that there are fundamental regularities. Thirdly, no model can count as a mechanistic model, unless it incorporates (causal) *P*-laws. To adequately model a mechanism, it should model the regular behaviour of the mechanism and its parts. Finally, a generalization can be considered a (causal) *P*-law even in the absence of any knowledge of some underlying mechanism. Hence we do not face the problem of infinite epistemological regress. Randomized experimental designs allow us to assess a generalization's stability and strength without mechanistic background knowledge.

My arguments primarily addressed the theories of Machamer et al. (2000) and of Bechtel and Abrahamsen (2005). This may raise doubts on the external validity of my conclusions. For sure, they cannot be simply extended to the mechanistic theories within the Salmon/Dowe approach. (And I will leave the question to what extent this approach is victim to my arguments untouched.) On the other hand, they are generalizable to mechanicists such as Glennan (1996, 2002) and Woodward (2002b) who openly endorse the intimate relation between mechanisms and regularities (see also Craver (2006, 372) and Craver (2007)).[9]

In section 2.2 I listed four objections to strict laws and D-N explanation that are raised by Machamer et al. and by Bechtel and Abrahamsen, and that motivate the mechanistic account of explanation. Let us see to what extent these objections still hold if laws are conceived of pragmatically, and if the concept of explanation is relaxed so that it settles for less than strict laws, i.e. for (causal) *P*-laws. (Apart from mechanistic explanation, a detailed account of explanation without strict laws can be found in Woodward (2003b). It is absent, however, in Mitchell's work. I will discuss explanation with non-strict laws in chapter 5.)

Firstly, it is argued that in the life sciences strict laws of nature are rarely if ever discovered and that D-N explanation is hardly practicable. This is true, but it does not follow that we should reject the concept of law. Although I am willing to give up the word 'law', it think it is more fruitful to revise the concept and to do justice to its role in explanation (cf. infra).

Secondly, Bechtel and Abrahamsen argue that staking on mechanisms as real systems allows to avoid some hard ontological problems regarding laws. If by laws they mean some metaphysical entity producing regularities, they might be right. But if, as is most probable (cf. Bechtel and Abrahamsen, 2005, 422), they mean generalizations describing regularities, this is at

[9] I will discuss Woodward's concept of *cs*-mechanism in section 5.3.

best a pseudo-advantage. By section 2.6, even when staking on mechanisms and mechanistic explanation one cannot escape from questions regarding the status of (causal) P-laws and their relations to (causal) P-regularities.

Thirdly, it is argued that even if there would be strict biological laws, explanation in the life sciences still takes the form of mechanistic explanation in many cases. This is also true, but raises the question why mechanistic explanation is so successful. Mechanistic explanation is typically directed at regularities: the mechanism's macro-level behaviour (or derivatively at events that are the product of a reliable mechanism) (Glennan, 2002, S346–S349). To explain such regularities, a mechanistic model should incorporate (causal) P-laws, as we have seen in section 2.6. I endorse the view that mechanistic explanation is a valuable concept (this will emerge in section 5.3). It provides a very solid account of scientific explanation without strict laws. But it would fail to account for explanation in the life sciences if it would abandon the idea of projectibility. A large part of the success of mechanistic explanation derives from the fact that it involves (causal) P-laws.

Of course, some caution is in order here. The fourth (and final) objection states that explanation involves more than subsumption under some law or regularity. It is argued that laws or regularities do not explain *why* the explanandum phenomenon occurs (Bechtel and Abrahamsen), or that the explanatory power wholly resides in the productivity of the activities (Machamer et al.). Again, I endorse that subsumption under a strict law is not sufficient for explanation. But two points may be raised in defence of (causal) P-laws. Firstly, some notion of productivity is ingrained in Woodward's causal theory of explanation (an intervention on X with respect to Y *produces* a change in Y (Woodward, 2003b, 98)) and hence also in the notion of causal P-law. Secondly, even if this is not enough, i.e. even if (causal) P-laws are not deemed sufficient for explanation, it does not follow that they are not necessary.

This should suffice to show that the concepts of '(causal) pragmatic law' and '(causal) pragmatic regularity' are valuable alternatives to the decried strict laws and regularities, and that they are compatible with (but not superseded by) the concepts of '*cs*-mechanism' and 'mechanistic model'. Now it remains to be shown how they can be represented, how they can be used in prediction, explanation or manipulation, and how they are discovered.

Chapter 3

The Language of Causal Modelling

In the following chapters I will use the language of causal modelling to express the pragmatic laws of classical genetics, and to represent its causal structure. I will also use it to handle several central concepts from philosophy of science, such as 'exemplars', 'anomalies', 'explanation', 'policy', 'experiments', and 'causal discovery', all of which are closely tied to laws of nature. In this chapter, I will present this language. In section 3.2 I will discuss the concept of 'causal net' or 'causal model' based on the works of Pearl (2000), Spirtes et al. (2000), Neapolitan (2004), and Williamson (2005). As all the definitions in this section are standard, I will not present quotations as quotations. Then, in section 3.3, I will introduce several interesting relations between causal nets. In section 3.4 I will introduce the concepts of causal scheme, credal set and credal net, and again discuss several interesting relations that may obtain between such objects. Finally, in section 3.5 I will discuss a useful kind of relations between sets of variables, viz. where one set is the 'compound' of two or more other sets.

Credal sets and credal nets are representations of imprecise probabilities (Cozman, 2005). As often as not, they go hand in hand with a Bayesian interpretation of probabilities. In that case, they represent imprecise degrees of belief. Here however, their function will be slightly different. I will use them (along with causal schemes) to economically describe the set of causal models of a theory, in casu classical genetics. In each model, all probabilistic relations are precise and probabilities are interpreted physically; together, these models provide imprecise probabilities.

The word "model" has many different interpretations in philosophy of science. As I will repeatedly use the word, this may lead to confusion. To avoid any such confusion, I will start by specifying what precisely I will mean

by the term. To that extent, some few distinctions from the literature on models and modelling will be taken into account. (I do not intend to give an enumeration, let alone an elaborate overview of all possible types of models. I refer the reader to e.g. Lloyd, 1998 and Frigg, 2006).

3.1 Models, models, and models

The word "model" has many different interpretations in philosophy of science. Balzer et al. (1987) distinguish between two main interpretations:

> In ordinary language and in informal contexts within empirical science, the term "model" is used in an ambiguous way. If we consider the term "model" with respect to the relationship between a "picture" of something and the thing "depicted", it appears that "model" is sometimes used in the sense of picture and sometimes in the sense of the thing depicted. The two meanings are logically converse, so that one can be defined through the other by just transforming the phrase "x is a model of y" into "y is a model of x". Empirical scientists tend to use "model" in the sense of "picture", as when they say that a certain set of equations "is a model" of some subatomic phenomena or certain real-life market situations. Logicians and mathematicians consistently use "model" in the sense of the thing depicted by a picture (= by a *theory*). (Balzer et al., 1987, 2)

The mathematician's and logician's use of "model" is derived from Tarski. Suppes phrases it as follows:

> Roughly speaking, a model of a theory may be defined as a possible realization in which all valid sentences[1] of the theory are satisfied, and a possible realization of the theory is an entity of the appropriate set-theoretical structure. (Suppes, 1969b, 24)

Let me call these *logical models*, or *set-theoretical models*.[2] In the previous chapter, we already met a different kind of model. There the phrase "mechanistic model" was frequently used. These mechanistic models are

[1] By the valid sentences of a theory, Suppes means the logical consequences of its axioms. 'Valid sentence' is thus a syntactic notion, not a semantic one as in e.g. Boolos and Jeffrey (2002, 120).

[2] The term *mathematical models* may also be used in the sense of 'logical model' or 'set-theoretical model'. But this may lead to confusion since it is also applied to sets of mathematical equations.

what I will call *iconic models*, or *representational models*, i.e. "those that scientists construct as more or less abstract descriptions of a real system" (Craver, 2006, 356). They describe mechanisms and thus are 'pictures', not things depicted. Iconic or representational models may be conceived of as mini-theories.

This distinction between models as pictures (representational models) and models as the things depicted (logical models) raises two questions which should be answered before we may enter upon the conceptual framework of causal modelling and its use in the following chapters. Firstly, which of both uses of the word "model" is most fundamental or basic? Secondly, is the above distinction sharp or exclusive in the sense that there is an unbridgeable gap between the two interpretations?

Which use of the word "model" is most fundamental? The logical notion of "model" is taken to be fundamental by many philosophers, not only with respect to mathematics, but also regarding the empirical sciences. Suppes writes:[3]

> It is my opinion that this [logical] notion of model is the fundamental one for the empirical sciences as well as mathematics. To assert this is not to deny a place for variant uses of the word 'model' by empirical scientists, as, for example, when a physicist talks about a physical model, or a psychologist refers to a quantitative theory of behavior as a mathematical model. (Suppes, 1969b, 24)

The position of Balzer et al. (1987) is similar in this respect. They add the following to their previous quote:

> Since this second use of "model" is well-established and clearly defined in the formal sciences, it is the one we are going to adopt here. Therefore, instead of saying that certain equations are a model of subatomic or economic phenomena, we propose to say that the subatomic or economic phenomena are models of the theory represented by those equations. (Balzer et al., 1987, 2)

Depending on the formal system at hand, logical models can take several forms. I quote Pearl, since I will return to his position in a minute. He writes:

[3]Suppes (1969b) analyses the relations between theories, experiments and data via a hierarchy of logical (or set-theoretical) models. Although this is a very fruitful idea, I will not incorporate it explicitly in the following chapters.

> A 'model,' in the common use of the word, is an idealized representation of reality that highlights some aspects and ignores others. In logical systems, however, a model is a mathematical object that assigns truth values to sentences in a given language, where each sentence represents some aspect of reality. Truth tables, for example, are models in propositional logic; they assign a truth value to any Boolean expression, which may represent an event or a set of conditions in the domain of interest. A joint probability function, as another example, is a model in probability logic; it assigns a truth value to any sentence of the form $P(a \mid b) < p$, where a and b are Boolean expressions representing events. (Pearl, 2000, 202, I slightly change notation)

Let us turn now to the second question, viz. whether the distinction between representational models (models as pictures) and logical models (models as the things depicted) is a sharp one. I will tackle this question from two sides. Firstly, there is no unbridgeable gap between models of the first kind and models of the second kind. Models of both kinds may complement each other. Secondly, it will be seen that the distinction is not exclusive. Models may be both pictures and things depicted at the same time.

I endorse the view of Suppes that there is no unbridgeable gap between representational or 'physical' models and logical or set-theoretical models. By 'physical model' Suppes means a concrete physical thing. For example, "many physicists want to think of a model of the orbital theory of the atom as being more than a certain kind of set-theoretical entity. They envisage it as a very concrete physical thing built on the analogy of the solar system" (Suppes, 1969a, 13). This physical model is intended to represent the atom and thus should be considered a representational model. According to Suppes, there is no real incompatibility between set-theoretical models and such physical models (see also Suppe, 1974, 96–97):

> To define a model as a set-theoretical entity which is a certain kind of ordered tuple consisting of a set of objects and relations and operations on these objects is not to rule out the physical model of the kind which is appealing to physicists, for the physical model may simply be taken to define the set of objects in the set-theoretical model. (Suppes, 1969a, 13)

Another thing is that the distinction between logical models and representational models is not exclusive. Models may be both pictures and things depicted at the same time. Quite recently, the idea was put forward by Morgan and Morrison (1999) that models are relatively (but not completely!)

autonomous mediators between theory and reality. They mostly focus on models in the first sense, i.e. as some kind of mini-theories:

> In some cases the distinction between models and theories is relatively straightforward; theories consist of general principles that govern the behaviour or large groups of phenomena; models are usually more circumscribed and very often several models will be required to apply these general principles to a number of different cases. (Morrison and Morgan, 1999b, 12)

Although I will use the word "model" rather in the sense of the thing depicted, interesting lessons can be drawn from their account. In their view, there is a significant lacuna in our understanding of how models function to give us information about the world.

> [The] literature on scientific practice still characterises the model as a subsidiary to some background theory that is explicated or applied via the model. [...] The result is that we have very little sense of what a model is in itself and how it is able to function in an autonomous way. (Morrison and Morgan, 1999a, 7–8)

> Yet clearly, autonomy is an important feature of models [...]. Viewing models strictly in terms of their relationship to theory draws our attention away from the processes of constructing models and manipulating them, both of which are crucial in gaining information about the world, theories and the model itself. However, in addition to emphasising the autonomy of models as entities distinct from theory we must also be mindful of the ways that models and theory do interact. [...] Our goal is to clarify at least some of the ways in which models can act as autonomous mediators in the sciences and to uncover the means by which they function as a source of knowledge. (Morrison and Morgan, 1999a, 8)

The metaphor of "models as mediators" is highly valuable. I will not include all the subtleties of Morgan and Morrison's account (see Morrison and Morgan, 1999b). I will only retain the following central point. Even if models are interpreted in the logical sense, i.e. as models of a theory, as models that are described or depicted by a theory, they retain some kind of autonomy. These models are not identical to 'the world'. All models involve some degree of abstraction and idealization. To my opinion, Cartwright's label 'simulacra' nicely fits logical models too.

> A model is a work of fiction. Some properties ascribed to objects in the model will be genuine properties of the objects modelled, but others will be merely properties of convenience. (Cartwright, 1983, 153)

> In order to stress this 'anti-realistic' aspect of models, I call my view of explanation a 'simulacrum' account. (Cartwright, 1983, 152)

In this chapter, I will introduce several well-known concepts from the causal modelling literature, as well some new ones. In chapter 4 I will use these concepts to analyse the causal structure of classical genetics in a way that to some extent resembles the structuralist analyses of Balzer et al. (1987) and of Balzer and Lorenzano (2000). But where they use Bourbakian set-theoretical models, I will use causal models instead. My models play the same *role* as their models (both are things depicted), but they are constructed from a different formal language. This is *grosso modo* in line with the approach of Pearl (2000, chapter 7) – *grosso modo*, as there are also differences between our approaches (cf. section 4.4). To the above quote, Pearl adds: "A *causal model*, naturally, should encode the truth values of sentences that deal with causal relationships. [...] Such sentences cannot be interpreted in standard propositional logic or probability calculus because they deal with changes that occur in the external world rather than with changes in our beliefs about a static world." (Pearl, 2000, 202–203, original emphasis) In my analysis, the theory of classical genetics consists of, or describes, or depicts (interrelated sets of) causal nets or causal models (I will make no distinction between both terms). These causal models in turn resemble actual things in the world. Consequently, they are set-theoretic models (things depicted) and representational models (pictures or descriptions of the world) at the same time (see figure 3.1).

Before we start, a notational issue has to be addressed. In the following section, I will use the expression '$\mathcal{B} = \langle G, P \rangle$' to denote a causal net or causal model and I will continue to do so throughout all the following chapters, except one. In chapter 8, where I will present **ALIC**, the adaptive logic for causal discovery, I will use the expression '$M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$' to denote causal models. In that expression, $\mathbb{R}^+$ is just the set of nonnegative real numbers. The functions c and p can be equated with G and P in '$\mathcal{B} = \langle G, P \rangle$', respectively. This notational change will be helpful in order to make a clear distinction between the syntax and the semantics of the logic **ALIC** (and the related logic **LIC**).

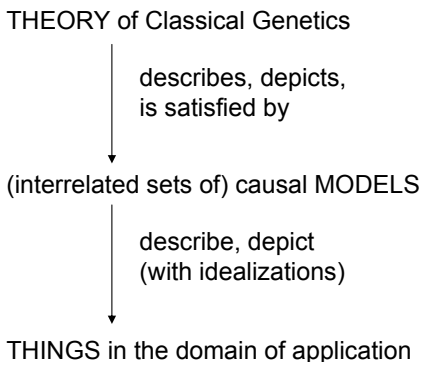

Figure 3.1: *Theories, models and reality*

3.2 Causal nets

A *causal net* is a causally interpreted Bayesian network. A *Bayesian network* $\mathcal{B} = \langle G, P \rangle$ is defined on a finite set of random variables V. It consists of a directed acyclic graph $G = \langle V, E \rangle$ and a probability distribution $P(V)$, where P and G satisfy the (causal) Markov condition.[4]

In statistics, a *random variable* represents some feature of an entity or set of entities. (In section 3.5 I will present the precise mathematical definition of 'random variable'.) Associated with each random variable $A \in V$ is the *space* of A, i.e. the set of mutually exclusive and jointly exhaustive values it may assume. Let $[A]$ denote the space of A. If A is a discrete variable, $[A]$ has finitely or countably many elements. Each value $a \in [A]$ represents a different state that the feature can take. Where $U = \{A_1, \ldots, A_n\}$ is a set of variables, let $[U] = [A_1] \times \ldots \times [A_n]$. The members of $[U]$ thus consist of the possible configurations of members of $[A_1], \ldots, [A_n]$, respectively. I immediately need to relax this definition. I will not attach significance to the order of the values in the elements of $[U]$ (note that the variables in U are not ordered either). For example, suppose that $[A_1] = \{a_{11}, a_{12}\}$, $[A_2] = \{a_{21}, a_{22}\}$ and $U = \{A_1, A_2\}$. Then I will not discriminate between the couple $\langle a_{21}, a_{11} \rangle$ and the couple $\langle a_{11}, a_{21} \rangle$, even if strictly speaking only one of them can be a member of $[U]$. I will also relax the above definition in a second way. In the interest of readability, each n-tuple $u = \langle a_1, a_2, \ldots, a_n \rangle$ may be written as $a_1 a_2 \ldots a_n$ or as $a_1, a_2, \ldots, a_n$. All this being said, let

[4] As I will not touch upon the problem of causal discovery from merely observational data in chapter 4, I will not require here that P and G are faithful. (Moreover, I will treat the credal nets of classical genetics as convex sets (cf. infra). It may be feared that convexity and faithfulness may conflict with each other.)

me also define $[[U]] = \bigcup [A_i]$ (for all $A_i \in U$). Obviously, $[U]$ and $[[U]]$ are different sets.

From a set of discrete random variables, (joint) probability distributions can be defined as follows.

Definition 3.1 (joint probability distribution) *Let a set of n random variables $U = \{A_1, A_2, \ldots, A_n\}$ be specified such that each A_i has a finite or countably infinite space $[A_i]$. A function, that assigns a real number $P(U = u_i)$ to every $u_i \in [U]$ (i.e. to every combination of the values of the variables A_i) is called a joint probability distribution of the random variables in U if it satisfies the following conditions:*

1. *For every $u_i \in [U]$,*

$$0 \leq P(U = u_i) \leq 1$$

2. *We have*

$$\sum P(U = u_i) = 1$$

Here is an intuitive example. Consider a pack of playing cards (with no joker). Each card has a definite colour, suit and face. Colour, suit and face are features of playing cards, so we may define random variables $Colour$, $Suit$ and $Face$ representing these features. In a pack of cards, some cards are red, others are black. Red and black are the different states that $Colour$ can take. Hence, $[Colour] = \{red, black\}$. Analogously, $[Suit] = \{hearts, diamonds, clubs, spades\}$ and $[Face] = \{ace, two, three, \ldots, ten, jack, queen, king\}$. Where $U = \{Colour, Face\}$, $[U] = \{\langle red, hearts\rangle, \langle red, diamonds\rangle, \langle red, clubs\rangle, \ldots, \langle black, spades\rangle\}$ and $[[U]] = \{red, black, clubs, spades, diamonds, hearts\}$.

Finally, the standard probability distribution associated with a pack of playing cards is such that

$P(red, hearts) = 0.25$	$P(red, diamonds) = 0.25$
$P(red, clubs) = 0.00$	$P(red, spades) = 0.00$
$P(black, hearts) = 0.00$	$P(black, diamonds) = 0.00$
$P(black, clubs) = 0.25$	$P(black, spades) = 0.25$

A *directed acyclic graph* $G = \langle V, E\rangle$ consists of a set V of vertices or nodes (this is the same set of random variables on which $\mathcal{B}$ is defined), and a set E of directed edges ($A \rightarrow B$, where $A, B \in V$). Since G is acyclic, there is no directed path from any variable to itself. If a Bayesian network $\mathcal{B} = \langle\langle V, E\rangle, P\rangle$ is *interpreted causally*, i.e. if $\mathcal{B}$ is a causal net, then for all

vertices $A, B \in V$, $A \rightarrow B \in E$ if and only if A denotes a direct cause of B, relative to V. The notion of a direct cause may be interpreted in Woodward's interventionist sense here:

(DC) A necessary and sufficient condition of X to be a direct cause of Y with respect to some variable set V is that there be a possible intervention on X that will change Y (or the probability distribution of Y) when all other variables in V besides X and Y are held fixed at some value by interventions. (Woodward, 2003b, 55)

As I said, for $\mathcal{B}$ to count as a causal net, its elements $G = \langle V, E \rangle$ and $P(V)$ must satisfy the *causal Markov condition* (cf. Williamson, 2005, 50).

Definition 3.2 (Causal Markov condition) *$G = \langle V, E \rangle$ and $P(V)$ satisfy the causal Markov condition if and only if each variable is probabilistically independent (according to P) of its non-effects, conditional on its direct causes (where causes and effects are relative to G).*

In case it is clear over which set of variables a distribution is defined, I will write P instead of $P(V)$.

3.3 Relations among causal nets

As is evident from the above definitions, causal nets may differ from each other along different lines. They may differ with respect to V and/or E and/or P. Therefore, I will define several possible relations between sets of variables, between graphs, and between causal nets that are helpful to analyse the causal structure of classical genetics. (I apologize for using the symbol '$\rightarrow$' in two different ways in the following definitions, viz. to refer to arrows in graphs and to define bijections.)

Firstly, sets of variables may be isomorphic or value-isomorphic.

Definition 3.3 ((value-)isomorphism for sets of variables) *V and V' are* isomorphic *if and only if there is a bijection $b : V \rightarrow V'$. V and V' are* value-isomorphic *if and only if there are bijections $b : V \rightarrow V'$ and $b' : [[V]] \rightarrow [[V']]$ such that for any $A \in V$ and $a \in [[V]]$, $a \in [A]$ if and only if $b'(a) \in [b(A)]$.*[5]

Sets of variables that are value-isomorphic are isomorphic, but not vice versa. Isomorphism and value-isomorphism for sets of variables are illustrated in figure 3.2. Graphs and causal nets may also be isomorphic or value-isomorphic.

[5]Recall that $[[V]] = \bigcup [A_i]$ (for all $A_i \in V$).

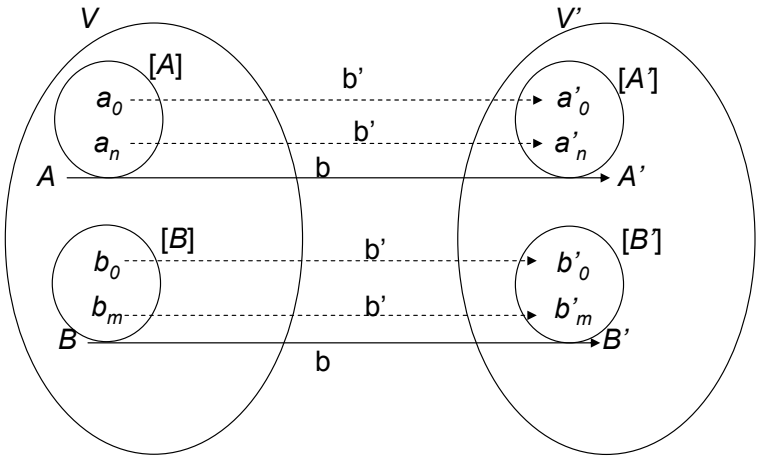

Figure 3.2: *Schematic representation of isomorphism and value-isomorphism between sets of variables*

Definition 3.4 ((value-)isomorphism for graphs) *$G = \langle V, E\rangle$ and $G' = \langle V', E'\rangle$ are* isomorphic *if and only if there is a bijection $b : V \to V'$ such that for any $A, B \in V$: $A \to B \in E$ if and only if $b(A) \to b(B) \in E'$.*[6] *G and G' are* value-isomorphic *if and only if there are bijections $b : V \to V'$ and $b' : [[V]] \to [[V']]$ such that for any $A, B \in V$: $A \to B \in E$ if and only if $b(A) \to b(B) \in E'$, and such that for any $A \in V$ and $a \in [[V]]$, $a \in [A]$ if and only if $b'(a) \in [b(A)]$.*

If $G = \langle V, E\rangle$ and $G' = \langle V', E'\rangle$ are isomorphic (respectively, value-isomorphic), then V and V' are isomorphic (respectively, value-isomorphic), but not vice versa.

Definition 3.5 ((value-)isomorphism for causal nets) *$\mathcal{B} = \langle G, P\rangle$ and $\mathcal{B}' = \langle G', P'\rangle$ are* isomorphic *if and only if G and G' are isomorphic. $\mathcal{B}$ and $\mathcal{B}'$ are* value-isomorphic *if and only if G and G' are value-isomorphic.*

With respect to isomorphic graphs and causal nets, let me introduce the following convention:

Convention 3.6 *If $G = \langle V, E\rangle$ and $G' = \langle V', E'\rangle$ are isomorphic, I will write $G' = \langle V', E\rangle$ instead of $G' = \langle V', E'\rangle$ – even if E and E' are specified over different sets of variables. Analogously, I will write $\mathcal{B}' = \langle\langle V', E\rangle, P'\rangle$ instead of $\mathcal{B}' = \langle\langle V', E'\rangle, P'\rangle$ in case $\mathcal{B}$ and $\mathcal{B}'$ are isomorphic.*

[6] This corresponds to the definition of isomorphic causal structures in Spirtes et al. (2000, 22). See also `http://mathworld.wolfram.com/IsomorphicGraphs.html`.

The relations of isomorphism and value-isomorphism of definition 3.5 concern *structural* similarities between causal nets. Let us now turn to *probabilistic* similarities.

Definition 3.7 (distribution-identity for causal nets) *Let* $\mathcal{B} = \langle\langle V, E\rangle, P\rangle$ *and* $\mathcal{B}' = \langle\langle V', E\rangle, P'\rangle$ *be value-isomorphic and let* b *and* b' *be bijections as in definition 3.3. Moreover, where* $u = \langle a_1, a_2, \ldots, a_n\rangle \in [V]$, *let* $b'(u) = \langle b'(a_1), b'(a_2), \ldots, b'(a_n)\rangle$. *Then* $\mathcal{B}$ *and* $\mathcal{B}'$ *are* distribution-identical *if and only if* $P(u) = P'(b'(u))$ *for each* $u \in [V]$. *(In other words,* $\mathcal{B}$ *and* $\mathcal{B}'$ *have like joint distributions.)*

3.4 Causal schemes and credal nets

Now that the relations of isomorphism, value-isomorphism and distribution identity are defined, let us turn to the concept of a causal scheme. A *causal scheme* is a set of isomorphic (or value-isomorphic) causal nets.[7]

Definition 3.8 (causal scheme (1)) *For any causal net* $\mathcal{B}$, *let the causal scheme* $\mathfrak{C}(\mathcal{B})$ *be the set of causal nets that are isomorphic to* $\mathcal{B}$

$$\mathfrak{C}(\mathcal{B}) = \{\mathcal{B}' \mid \mathcal{B}' \text{ is isomorphic to } \mathcal{B}\}$$

and let the causal scheme $\mathfrak{C}([\mathcal{B}])$ *be the set of causal nets that are value-isomorphic to* $\mathcal{B}$

$$\mathfrak{C}([\mathcal{B}]) = \{\mathcal{B}' \mid \mathcal{B}' \text{ is value-isomorphic to } \mathcal{B}\}$$

Note that $\mathfrak{C}([\mathcal{B}]) \subset \mathfrak{C}(\mathcal{B})$.

Causally interpreted credal nets are an interesting special case of causal schemes. A credal net may be thought of as a representation for a set of Bayesian networks over a fixed set of variables (Cozman, 2005, 171), so a causally interpreted credal net represents a set of causal nets. These causal nets all have the same graph, but differ with respect to P. A *credal net* $\mathfrak{B} = \langle G, \mathbb{P}\rangle$ consists of a directed acyclic graph $G = \langle V, E\rangle$ and a set $\mathbb{P}$ of probability distributions over V.

A set of probability distributions is called a *credal set*. In Bayesian contexts, these are used to model imprecise degrees of belief. Consider again

[7] Causal schemes (and generic credal nets, cf. infra), it will turn out, are similar in spirit to Kitcher's *general argument patterns* (Kitcher, 1989, 432–433) or to Darden's abstract explanatory patterns (Darden, 1991, 19). They also resemble the structuralist notion of theory-element (Balzer et al., 1987, 39).

our pack of playing cards and suppose that a croupier randomly draws one card. If I am convinced that this pack of cards is normal and contains no joker, my degree of belief that the card drawn is red should be one half: $P(Colour = red) = \frac{26}{52}$ (more briefly, $P(red) = \frac{26}{52}$). This is a precise degree of belief. Suppose, however, that I am convinced that the croupier has falsified the game by turning a red card into a black card at least one time. Then my degree of belief should be between $P(red) = \frac{25}{52}$ and $P(red) = \frac{0}{52}$, and is no longer precise. The credal set $\mathbb{P} = \{P \mid P(red) = \frac{26-n}{52}, 1 \leq n \in \mathbb{N} \leq 26\}$ represents these possible degrees of belief.

Definition 3.9 (credal set) *A* credal set $\mathbb{P}(V)$ *is a set of probability distributions* $P(V)$*.*

In case no confusion may arise, I will write $\mathbb{P}$ or P instead of $\mathbb{P}(V)$ or $P(V)$.

Credal sets may be convex or non-convex, and they may be open or closed (cf. Cozman, 2005, 175).

Definition 3.10 (convexity for credal sets) *A credal set* $\mathbb{P}$ *is convex if, for any measures* $P, P' \in \mathbb{P}$*, the measure* $\alpha P + (1-\alpha)P' \in \mathbb{P}$ *for any* $\alpha \in [0, 1]$*.*

Openness and closeness of a credal set can be compared with openness and closeness of intervals: $[a, b]$ is a closed interval, $]c, d[$ is open.

The credal set specified in the above example is non-convex. It could be argued, however, that my degrees of belief range over the complete convex hull of this set. (Where the convex hull of $\mathbb{P}$ is the intersection of all convex sets containing $\mathbb{P}$; Walley, 1991, 611) In that case, I better could have specified $\mathbb{P}$ as a closed convex set: $\mathbb{P} = \{P \mid 0 \leq P(red) \leq \frac{25}{52}\}$.

Credal nets can be defined in terms of credal sets.

Definition 3.11 (credal net) *A credal net is a set of causal nets with a common graph:* $\mathfrak{B} = \langle G, \mathbb{P}\rangle = \{\mathcal{B} = \langle G, P\rangle \mid P \in \mathbb{P}\}$

I stated above that causally interpreted credal nets are interesting special cases of causal schemes. The reason is that if $\mathcal{B}, \mathcal{B}' \in \mathfrak{B}$, their graphs are identical and hence (value-)isomorphic. Therefore, if $\mathcal{B}, \mathcal{B}' \in \mathfrak{B}$ then $\mathcal{B}, \mathcal{B}' \in \mathfrak{C}(\mathcal{B})$ and $\mathcal{B}, \mathcal{B}' \in \mathfrak{C}([\mathcal{B}])$ (note that the converse does not hold). It follows that we can also define causal schemes on the basis of credal nets.

Definition 3.12 (causal scheme (2)) *For any credal net* $\mathfrak{B}$*, let the causal scheme* $\mathfrak{C}(\mathfrak{B})$ *be the set of causal nets that are isomorphic to some member of* $\mathfrak{B}$

$$\mathfrak{C}(\mathfrak{B}) = \{\mathcal{B}' \mid \mathcal{B}' \textit{ is isomorphic to some } \mathcal{B} \in \mathfrak{B}\}$$

and let the causal scheme $\mathfrak{C}([\mathfrak{B}])$ *be the set of causal nets that are value-isomorphic to some member of* $\mathfrak{B}$

$$\mathfrak{C}([\mathfrak{B}]) = \{\mathcal{B}' \mid \mathcal{B}' \textit{ is value-isomorphic to some } \mathcal{B} \in \mathfrak{B}\}$$

Note that $\mathfrak{B} \subset \mathfrak{C}([\mathfrak{B}]) \subset \mathfrak{C}(\mathfrak{B})$.

Structural and probabilistic similarities can also be defined for credal sets, credal nets and causal schemes.

Definition 3.13 ((value-)isomorphism for credal sets) *The credal sets* $\mathbb{P}(V)$ *and* $\mathbb{P}'(V')$ *are* isomorphic *if and only if* V *and* V' *are isomorphic.* $\mathbb{P}(V)$ *and* $\mathbb{P}'(V')$ *are* value-isomorphic *if and only if* V *and* V' *are value-isomorphic.*

Definition 3.14 (distribution-identity for credal sets) *Let* V *and* V' *be value-isomorphic, let* b *and* b' *be bijections as in definition 3.3, and let* u *and* $b'(u)$ *be defined as in definition 3.7. Then the credal sets* $\mathbb{P}(V)$ *and* $\mathbb{P}'(V')$ *are* distribution-identical *if and only if there is a bijection* $b'' : \mathbb{P} \to \mathbb{P}'$ *such that* $P(u) = b''(P)(b'(u))$ *for all* $P \in \mathbb{P}$ *and all* $u \in [V]$. *(In other words, all distributions in* $\mathbb{P}$ *are distribution-identical to their image in* $\mathbb{P}'$.*)*

If two credal sets are distribution-identical, they are value-isomorphic (and hence isomorphic), but not vice versa. Let us turn now to relations between credal nets.

Definition 3.15 ((value-)isomorphism for credal nets) $\mathfrak{B} = \langle G, \mathbb{P}\rangle$ *and* $\mathfrak{B}' = \langle G', \mathbb{P}'\rangle$ *are* isomorphic *if and only if* G *and* G' *are isomorphic.* $\mathfrak{B} = \langle G, \mathbb{P}\rangle$ *and* $\mathfrak{B}' = \langle G', \mathbb{P}'\rangle$ *are* value-isomorphic *if and only if* G *and* G' *are value-isomorphic.*

Definition 3.16 (distribution-identity for credal nets) $\mathfrak{B} = \langle\langle V, E\rangle, \mathbb{P}\rangle$ *and* $\mathfrak{B}' = \langle\langle V', E\rangle, \mathbb{P}'\rangle$ *are* distribution-identical *if and only if* $\mathbb{P}(V)$ *and* $\mathbb{P}'(V')$ *are distribution-identical.*

Equivalently, $\mathfrak{B}$ *and* $\mathfrak{B}'$ *are* distribution-identical *if and only if there is a bijection* $b^* : \mathfrak{B} \to \mathfrak{B}'$ *such that for all* $\mathcal{B} \in \mathfrak{B}$, $\mathcal{B}$ *is distribution-identical to* $b^*(\mathcal{B})$.

If two credal nets are distribution-identical, they are value-isomorphic (and hence isomorphic), but not vice versa. The two formulations in definition 3.16 are equivalent. The second formulation can be used to define distribution-identity for causal schemes.

Theorem 3.17 *Both formulations in definition 3.16 are equivalent.*

Proof. $\Rightarrow$ Assume that $\mathfrak{B} = \langle\langle V, E\rangle, \mathbb{P}\rangle$ and $\mathfrak{B}' = \langle\langle V', E\rangle, \mathbb{P}'\rangle$ are distribution-identical in accordance with the first formulation. Then $\mathbb{P}$ and $\mathbb{P}'$ are distribution-identical (by definition 3.16). It follows that V and V' are value-isomorphic (by definition 3.14), so $G = \langle V, E\rangle$ and $G' = \langle V', E\rangle$ are value-isomorphic too (by definition 3.4). Since $\mathbb{P}$ and $\mathbb{P}'$ are distribution-identical, there is a bijection $b'' : \mathbb{P} \to \mathbb{P}'$ such that for all $P \in \mathbb{P}$, P is distribution-identical to $b''(P)$ (definition 3.14). Now define $b^* : \mathfrak{B} \to \mathfrak{B}'$ as follows: for any $\mathcal{B} = \langle G, P\rangle \in \mathfrak{B} : b^*(\mathcal{B}) = \langle G', b''(P)\rangle$. It follows that for any such $\mathcal{B}$, $\mathcal{B}$ is distribution-identical to $b^*(\mathcal{B})$ (by definition 3.7). This shows that $\mathfrak{B} = \langle\langle V, E\rangle, \mathbb{P}\rangle$ and $\mathfrak{B}' = \langle\langle V', E\rangle, \mathbb{P}'\rangle$ are distribution-identical in accordance with the second formulation.

$\Leftarrow$ Assume that $\mathfrak{B} = \langle\langle V, E\rangle, \mathbb{P}\rangle$ and $\mathfrak{B}' = \langle\langle V', E\rangle, \mathbb{P}'\rangle$ are distribution-identical in accordance with the second formulation. Then there is a $b^* : \mathfrak{B} \to \mathfrak{B}'$ such that for any $\mathcal{B} = \langle\langle V, E\rangle, P\rangle \in \mathfrak{B}$ it holds that $\mathcal{B}$ is distribution-identical to $b^*(\mathcal{B})$. As $b^*(\mathcal{B}) \in \mathfrak{B}'$, its graph is $\langle V', E\rangle$. Now define $b'' : \mathbb{P} \to \mathbb{P}'$ as follows: for any $P \in \mathbb{P}$, $b''(P) = P'$ if and only if $b^*(\langle\langle V, E\rangle, P\rangle) = \langle\langle V', E\rangle, P'\rangle$. This shows that $\mathbb{P}$ and $\mathbb{P}'$ are distribution-identical (definition 3.14). But then $\mathfrak{B} = \langle\langle V, E\rangle, \mathbb{P}\rangle$ and $\mathfrak{B}' = \langle\langle V', E\rangle, \mathbb{P}'\rangle$ are distribution-identical in accordance with the first formulation. ■

Let us now define distribution-identity for causal schemes.

Definition 3.18 (distribution-identity for causal schemes) *Two causal schemes $\mathfrak{C}$ and $\mathfrak{C}'$ are distribution-identical if and only if there is a bijection $b^{**} : \mathfrak{C} \to \mathfrak{C}'$ such that for all $\mathcal{B} \in \mathfrak{C}$, $\mathcal{B}$ is distribution-identical to $b^{**}(\mathcal{B})$.*

Summary of the above definitions:

	(value-)isomorphism	distribution-identity
sets of variables	definition 3.3	/
graphs	definition 3.4	/
causal nets	definition 3.5	definition 3.7
credal sets	definition 3.13	definition 3.14
credal nets	definition 3.15	definition 3.16
causal schemes	/	definition 3.18

3.5 Compound variables and sets of variables

Let us turn now to one more kind of useful relations, viz. relations of 'compoundedness' between variables and between sets of variables.

For the definition of a *compound variable*, let me briefly dwell on the precise definition of random variables. Higher I stated that a random variable represents some feature of an entity or set of entities, and that its values represent the different states this feature can take. This corresponds to the meaning of random variables in statistics (cf. Neapolitan, 2004, 28–34). The precise mathematical definition, however, is as follows (cf. Neapolitan, 2004, 3–5, 10–17):

Definition 3.19 (Random variable) *Let Ω be a sample space (i.e. the set of different outcomes of an experiment[8]) and let P be a probability function on the power set of Ω. Then the pair $\langle \Omega, P \rangle$ is called a probability space.*

Given a probability space $\langle \Omega, P \rangle$, a random variable X is a function on Ω that assigns a value $x \in [X]$ to each outcome in the sample space: $X : \Omega \rightarrow [X]$.

Now we can define compound variables and compound variable sets as follows:[9]

Definition 3.20 (Compound variable) *Random variable Z is the compound of variables X and Y ($Z = compound\langle X, Y \rangle$) if and only if*

1. $[Z] = [X] \times [Y]$, *and*
2. *for all $e \in \Omega$ and all $x_i \in [X]$, if $X(e) = x_i$, then $Z(e) \in [Z]^{x_i}$, where $[Z]^{x_i} = \{\langle x_i, y_j \rangle \mid y_j \in [Y]\}$, and*
3. *for all $e \in \Omega$ and all $y_j \in [Y]$, if $Y(e) = y_j$, then $Z(e) \in [Z]^{y_j}$, where $[Z]^{y_j} = \{\langle x_i, y_j \rangle \mid x_i \in [X]\}$*

By extension, we can define variables composed of more than two other variables recursively: $Z = compound\langle X_1, \ldots, X_n \rangle$ if and only if there is some Z' such that $Z = compound\langle Z', X_n \rangle$ and $Z' = compound\langle X_1, \ldots, X_{n-1} \rangle$.

[8] In probability theory, the following situations may be called *experiments*: tossing a coin once, tossing a coin 100 times, throwing three dice, selecting a random sample of 200 people and observing the number of left-handers in it, crossing two species of plants and observing the phenotypes of the offspring, ... Experiments are defined in terms of their outcomes or results. These are called *events*. For example, the experiment of randomly selecting 200 people and observing the number of left-handers may result in 201 different events. *Simple* (or indecomposable) events (e.g. observing 149 left-handers) are distinguished from *compound* (or decomposable) events (e.g. observing at least 149 left-handers). (Compound events should not be confused with compound variables.) Every compound event can be decomposed into simple events. Simple events are also called *sample points* (or points). The aggregate of all sample points is called the *sample space* (Ω). (Feller, 1950, 7–9)

[9] I could also have used the term 'joint variable' (cf. Antonucci and Zaffalon, LSCN, section 2).

Definition 3.21 (Compound variable set) *A set of variables V is the compound of sets V' and V'' ($V = compound\langle V', V''\rangle$) if and only if*

1. *V, V', and V'' are isomorphic (so that there are bijections $b : V \rightarrow V'$ and $b^* : V \rightarrow V''$), and*

2. *for all $A \in V$, $A = compound\langle b(A), b^*(A)\rangle$.*

By extension, we can define sets of variables composed of more than two other sets of variables recursively (cf. definition 3.20).

In the following chapters I will use all the concepts presented here to analyse the causal structure of classical genetics and to handle several central concepts from philosophy of science: exemplars, anomalies, explanation, experimentation, policy, etc.

3.6 Causal models and the structure of scientific theories

In section 3.1 I distinguished between two general kinds of models: set-theoretic models and representational models. Although both concepts of 'model' should not be confused, there is no unbridgeable gap between them. Theories describe, or are satisfied by, their set-theoretic models. But in turn, these models comprise abstract and idealized representations of 'reality', so they also function as representational models. In short, they are mediators between theory and reality.

In the following chapter, I will use one such kind of set-theoretic models: causal nets or causal models. I will use these models (and the other concepts defined in this chapter) to analyse the structure of classical genetics. To some extent my analysis will resemble the old structuralist approach of Balzer et al. (1987) and it will incorporate interesting insights from both Kitcher (1989) and Darden (1991). But it will supersede each of these frameworks in an important respect: it will be apt to represent the *causal* structure of classical genetics.

Chapter 4

The Causal Structure of Classical Genetics

4.1 Introduction

In this chapter I will tackle the following questions: "How can the structure of scientific theories, more specifically of classical genetics, be represented?" And: "How can the pragmatic laws that figure therein be expressed?" The question how to represent the structure of scientific theories has been asked many times in the history of philosophy of science, and many authors have used the theory of classical genetics as a case study.

The logical empiricists treated scientific theories as axiomatic theories formulated in first-order logic with identity. In their view, a scientific theory TC consists of a set of axioms or theoretical laws T and a set of correspondence rules C (Nagel, 1961, chapters 4–5). For example, Woodger (1952, part II) has formalized the theory of classical genetics by means of set theory, probability theory and first-order logic (see also Woodger, 1959 and Lindenmayer and Simon, 1980).

The logical empiricist approach became discredited (Suppe, 1974) and several alternatives emerged. Within the structuralist approach, theories are considered as classes of set-theoretic structures or models (Balzer et al., 1987). This framework has been applied to classical genetics by Lorenzano (1995, 2007) and by Balzer and Lorenzano (2000).[1] Although he did not directly tackle the question how to represent scientific theories, the works of Kitcher (1989) need also be mentioned. In his view, scientific theories consist of general argument patterns that can be used time and again for

[1] See also Balzer and Dawe (1986a), Balzer and Dawe (1986b) and Balzer and Dawe (1997).

explanation. He considers four examples of such patterns that have been employed in genetics in the twentieth century (Kitcher, 1989, 438–442). Finally, Darden (1991) analyses the theory of classical genetics in terms of changing theoretical components. She also attaches great importance to the role of diagrammatic representations. These representations, she argues, can be abstracted so as to obtain (some analogues of) Kitcherian argument patterns.

These alternatives all have their virtues, but they also have a joint shortcoming. None of them adequately handles the role of causation and causal relations in classical genetics. This is regrettable since causation is strongly tied to several other important topics in philosophy of science, such as explanation, experimentation, policy, etc. Balzer et al. (1987) pay absolutely no attention to the topics of causation, experiment or policy, and their concept of explanation has little to do with causation. (In their view, to explain a set of data is tantamount to fitting it in a model of a theory. But in these models no distinction is made between causal and non-causal relations.) In Balzer and Lorenzano (2000) several genetic relations are represented by functions. They acknowledge that some of these functions are in fact causal, but they lack the formal resources to distinguish them from non-causal ones. Moreover, they seem timorous with respect to causation. Regarding to genotypes, they write:

> Again, we distinguish between parental genotypes, those that 'cause' the parental phenotype, and a finite number of genotypes for the progeny of each parental pair, one for each progeny. (Balzer and Lorenzano, 2000, 247, original quotation marks)

In Kitcher's theory causation occupies a queer place. Causal relevance is dependent on explanatory relevance. The notion of the latter is tied to the systematization of belief in the limit of scientific inquiry, as guided by the search for derivational unification (Kitcher, 1989, 499). As we will see in chapter 5, there is little guarantee that causal relevance thus defined will fit 'the causal order of things in the world'.[2]

Finally, although Darden acknowledges that causal considerations played a role in genetics and cytology (Darden, 1991, 253–254), she pays little attention to causation.

I will not present the accounts of Kitcher (1989), Darden (1991) or Balzer and Lorenzano (2000) in full detail. Rather, I will occasionally touch upon the relevant aspects or the central concepts of their theories.

[2]This expression is adopted from Uskali Mäki (2001).

I will proceed as follows. In section 4.2 I will present the 'Theory of the Gene' as it was stated by Thomas Hunt Morgan. Then I will give a very short overview of the present-day terminology of genetics (section 4.3). In the subsequent sections I will represent the theory of classical genetics by means of interrelated causal nets. First I will list their causal relata and their joint graph (section 4.4). Then I will discuss Morgan's treatment of Mendel's crosses on tall and short pea plants (section 4.5), show how the results of these crosses were generalized to other reciprocal monohybrid crosses with complete dominance (section 4.6) and provide an explication of Darden's concept of exemplar in terms of generic credal nets (section 4.7). In sections 4.8 and 4.9 I will present reciprocal monohybrid crosses with incomplete dominance, discuss the role of (model) anomalies in classical genetics and show how the structuralist notion of theory-elements is useful to deal with such anomalies. In sections 4.10–4.13 I will treat multihybrid crosses with independent assortment, linkage and crossing-over. In section 4.14, I will quickly discuss other patterns of inheritance, as well as the role of environmental influences in (my representation of) classical genetics.

4.2 The theory of the gene

In 1926, Thomas Hunt Morgan published a very concise formulation of the 'theory of the gene'. It provides a succinct formulation of the theory of classical genetics which had developed since 1900 (or since 1865). A revised and enlarged edition was published in 1928. This formulation can be considered an end point in the development of classical genetics (Darden, 1991, 3, 38) and serves as a good starting point for exploring the causal structure of classical genetics.

According to Morgan, the modern theory of heredity "is primarily concerned with the distribution of units between successive generations of individuals." (Morgan, 1928, 1) These units are invisible and are called genes. To these genes, properties are assigned in a non-arbitrary way, based on "numerical data obtained by crossing two individuals that differ in one or more characters." (Morgan, 1928, 1) After presenting several examples of possible relations between genes and characters, Morgan formulates the *theory of the gene* as follows:

> We are now in a position to formulate the theory of the gene. The theory states that the characters of the individual are referable to paired elements (genes) in the germinal material that are held together in a definite number of linkage groups; it states

> that the members of each pair of genes separate when the germ-cells mature in accordance with Mendel's first law, and in consequence each germ-cell comes to contain one set only; it states that the members belonging to different linkage groups assort independently in accordance with Mendel's second law; it states that an orderly interchange – crossing-over – also takes place, at times, between the elements in corresponding linkage groups; and it states that the frequency of crossing-over furnishes evidence of the linear order of the elements in each linkage group and of the relative position of the elements with respect to each other.
> These principles, which, taken together, I have ventured to call the theory of the gene, enable us to handle problems of genetics on a strictly numerical basis, and allow us to predict, with a great deal of precision, what will occur in any given situation. In these respects the theory fulfills the requirements of a scientific theory in the fullest sense. (Morgan, 1928, 25; emphases omitted)

The theory of the gene, as it is stated here, is not one single theory. Instead, it is composed of different theoretical components (Darden, 1991) or theory-elements (Balzer et al., 1987; Balzer and Lorenzano, 2000).[3] These theory-elements have to some extent the same causal structure. This shared causal structure ties them together, and distinguishes the theory of the gene from alternative theories of inheritance. In chapter 9 I will discuss Francis Galton's 'theory of ancestral inheritance' which has a strongly different causal structure.

[3]Theoretical components are not the same as theory-elements. In Darden's framework, theoretical components evolve in the course of time.

> [...] I will be using the term 'theoretical component' to discuss parts of the theory that change over time. [...] The principal theory that I will be discussing is the theory of the gene; the early ideas that developed into it will be termed 'theoretical components of Mendelism.' (Darden, 1991, 18)

By contrast, theory-elements may be considered as the end-products of such theory change. They consist of a theory-core and a set of intended applications. The theory-cores of different theory-elements incorporate different 'laws' or axioms and may be used to account for different (but related) kinds of phenomena. See footnote 8 for a short presentation of some of the main components of theory-cores: potential models, actual models, partial potential models, ... (Balzer et al., 1987, 37–40; Balzer and Lorenzano, 2000, 245–246)

4.3 Classical genetics and present-day terminology

The history of classical genetics is replete with conceptual changes (cf. Darden, 1991). Many concepts were originally used ambiguously. Even in Morgan's monograph, several contemporary concepts are often tangled up. For example, Morgan often writes 'genes' where we would now use 'alleles'. I will often make use of present-day terminology (both from genetics and from cytology) to explicate Morgan's reasoning. This will add clarity and speed up my analysis.

In this section, I will briefly present this terminology, based on Klug et al. (2006, 1–136). I will confine myself to the terminology relating to classical genetics in eukaryotes. This means that I will not discuss molecular genetics or inheritance in prokaryotes. Since Klug et al. (2006) is a standard textbook, I will not present quotations as quotations.

In *eukaryotic organisms* such as plants and animals, each species has a characteristic number of chromosomes (the *diploid number* $2n$). Chromosomes in diploid cells exist in n pairs consisting of 2 *homologous chromosomes*. During *mitosis* (somatic cell division), chromosomes are copied and distributed so that the two resulting daughter cells each receive a diploid set of chromosomes. *Meiosis* is a form of cell division associated with gamete formation. *Gametes* are reproductive cells, i.e. sperm or pollen and eggs. Cells produced by meiosis receive only one member of each pair of homologous chromosomes, thus having n chromosomes in total (the *haploid number*). When two gametes fertilize, the resulting zygote has $n + n = 2n$ chromosomes. (Klug et al., 2006, 3–4)

Chromosomes are the bearers of the *genes*. Put simply, genes code for (or cause) observable *characters* (e.g. eye colour in humans). *Alleles* are alternative forms of genes, producing different *traits* of the same character (e.g. brown eyes or blue eyes). The organism's observable features are called its *phenotype*; the set of its alleles for a given character (or set of characters) is called its *genotype*. (Klug et al., 2006, 4, with slight modifications) In any diploid organism, two alleles of the same gene occur (each of which is part of a distinct member of a pair of homologous chromosomes). If both alleles are identical, the individual is *homozygous* (a *homozygote*), otherwise it is *heterozygous* (a *heterozygote*). (Klug et al., 2006, 42)

There are many more genes than chromosomes. Alleles that are part of the same chromosome are said to be linked, or to show *linkage* in genetic crosses. During meiosis, gametes receive one member of each pair of homologous chromosomes. Alleles that are part of the same chromosome are,

ceteris paribus, transmitted as a unit. However, during the first phase of meiosis (prophase I), homologous chromosomes may reciprocally exchange chromosome segments. This is called *crossing-over*. As the alleles on the chromosome segments in question are exchanged, linkage is violated. The degree of crossing-over provides a means to determine the *locus* (place) of genes on the chromosome. The distance between two genes on the same chromosome is called the *interlocus distance*. The degree of crossing-over between any two genes is proportional to their interlocus distance. By comparing the interlocus distances of many pairs of genes, *chromosome maps* describing the linear order of the genes can be constructed. (Klug et al., 2006, 28, 101–136)

For some genes, two alleles exist. Of these alleles, one may be dominant, the other recessive (e.g., the allele B for brown eyes is dominant to the allele b for blue eyes – so heterozygotes have brown eyes). This is called *dominance/recessiveness* or *complete dominance*. In *incomplete dominance*, neither allele is dominant. For example, red-flowered snapdragons are homozygous (R^1R^1), having two copies of the red-producing allele (R^1). White-flowered snapdragons are homozygous (R^2R^2) for the white-producing allele (R^2). Heterozygous snapdragons (R^1R^2) are neither red, nor white. They are pink.[4] In *codominance*, both alleles have an influence in the heterozygote (i.e. both can be considered as dominant). Here, both alleles produce distinct and detectable gene products (in the case of complete dominance or incomplete dominance, the gene products of one of the alleles are not detectable). The MN blood group in humans is a case of codominance. In people that are homozygous for M (i.e. have the M blood group) a glycoprotein molecule is found on the surface of red blood cells. In people that are homozygous for N (having the N blood group) a slightly different glycoprotein molecule is found. In heterozygotes (with the MN blood group), both variants of the molecule are found. (Klug et al., 2006, 68–69)

For other genes, more than two alleles exist, sometimes even up to a hundred or more (*multiple alleles*). For example, the ABO blood groups result from one gene having three alleles. I^A and I^B are each dominant to I^O, but they are codominant to each other. (Klug et al., 2006, 70–72)

The relation between genes and characters is not always one-to-one. On the one hand, phenotypes may be affected by more than one gene (*gene interaction*, also called *multiple factors*) (Klug et al., 2006, 75–79). On the other hand, expression of a single gene may have multiple effects (*pleiotropy*)

[4]This can be interpreted as follows. R^1 produces an enzyme that participates in a reaction leading to the synthesis of a red pigment. R^2 produces an enzyme that cannot catalyze the reaction leading to pigment. In the heterozygote (R^1R^2) only about half the pigment of the red-flowered plant is produced and the phenotype is pink. (Klug et al., 2006, 69)

(Klug et al., 2006, 80–81). Finally, phenotypic expression is not always a direct reflection of the genotype. Gene expression and the resulting phenotype are co-determined by the cell's internal milieu, the external environment, and the interaction between cells.

Sometimes a mutant genotype is only expressed in part of the mutants. The percentage of individuals that show at least some degree of expression of a mutant genotype defines the *penetrance* of the mutation (cf. section 5.5). By contrast, *expressivity* reflects the range of expression of the mutant genotype. For example, *Drosophila* flies that are homozygous for the recessive allele 'eyeless' may have normal eyes, or somewhat smaller eyes, or even lack one or two eyes. Expressivity ranges from complete loss of both eyes to completely normal eyes. This variation is due both to genetic background, and to environmental influences. Different kinds of genetic background effects have been distinguished. In the case of *genetic suppression* or *epistasis*, a suppressor gene causes the (complete) reversal of the expected phenotypic expression of some mutation. Also, the position of a gene on the chromosome may influence its expression (*position effect*). Some parts of the chromosome are more condensed and genetically inert. Genes may change position on the chromosome due to translocation or inversion. Analogously, different kinds of environmental influences have been distinguished. For example, some mutations are temperature sensitive (*temperature effect*). (Klug et al., 2006, 85–89)

4.4 The causal relata of classical genetics

In the first chapter of his monograph, "The fundamental principles of genetics", Morgan introduces the theory of the gene by means of several paradigmatic crosses (Morgan, 1928, 1–25). He cites Mendelian pea crosses (a tall variety versus a short variety) and crosses of humans (blue eye colour versus brown eye colour) to illustrate monohybrid crosses with complete dominance, and crosses of four-o'clocks (red flower colour versus white flower colour) to illustrate monohybrid crosses with what is now called incomplete dominance. In a monohybrid cross, only one set of opposing traits is observed. Mendelian pea crosses (yellow and round seeds versus green and wrinkled seeds) also serve to exemplify dihybrid crosses with independent assortment, whereas linkage and crossing-over are exemplified by crosses on fruit flies. These paradigmatic crosses will play an important role in the following sections.

Whereas these exemplars are *prima facie* very different, at a general level of description they are all similar, viz. at the level of their causal structure. More precisely, each of them relates to the inheritance of some particular set

of traits in some particular kind of organism. Each exemplar can be described by means of a credal net. The graph of this net describes the traits and factors at hand. Its credal set results from selecting those probability distributions that satisfy a certain set of principles. These are the principles by which the results of these crosses are described, predicted or explained. In other words, these principles are pragmatic laws. Where credal nets correspond to exemplars, their elements (causal nets) each describe some possible genetic cross on the organism and traits in question. The set-theoretical relations between different credal nets show the logical structure of classical genetics (cf. Balzer and Lorenzano, 2000). At the most general level, all causal nets belong to the same causal scheme, since their graphs are all isomorphic. This causal scheme represents the causal structure of classical genetics. (Causal schemes are similar to sets of potential models – cf. footnote 8.)

In each paradigmatic cross, two subsequent generations of individuals (or groups of individuals) are observed. In the parental generation, two (groups of) individuals are crossed. Their offspring is called the filial generation. So basically, three groups of individuals are observed: a first parental group, a second parental group, and a filial group. The members of the first parental group produce pollen or sperm (so let us call them the fathers); those of the second parental group produce seeds[5] or egg cells (so let us call them the mothers). Although I distinguish between two parental groups, cases of selfing (or self-fertilization, where the parents are identical) and cases of parthenogenesis (where one of the parental groups is absent) can be accounted for in my framework. In many studies, more than two generations are observed (e.g. a parental generation and two or three filial generations). These can be considered concatenations of the causal scheme to be outlined.

For each of the three groups, the following features are discussed (see figure 4.1): their observable characters (phenotype), their genetic make-up (genotype), and (mostly) the genetic make-up of their gametes.[6] These features may be represented by random variables. The phenotype of each individual or group of individuals i ($i = 1, 2, 3$) is represented by PT_i, their genotype by GT_i and the make-up of their gametes or germ-cells by GC_i. It will be seen that these variables are causally related.

Which are the causal relations that Morgan asserts between these random variables? From a present-day point of view, it is natural to consider genes as the causes of characters; i.e. to assume that for each i, $GT_i \rightarrow PT_i$. It

[5]Where Mendel wrote 'Samen' (Mendel, 1933, 18) or 'seeds' (Mendel, 1865, 18), Morgan (1928, 2–3) uses 'eggs' for the gametes of pea plants.

[6]The concepts of genotype and phenotype were introduced by Wilhelm L. Johannsen in 1909. Although they were not explicitly used in (Morgan, 1928), I will use them throughout this chapter to abbreviate Morgan's wording (cf. supra, section 4.3).

is not so natural, however, to attribute this assumption to Morgan. Causal language was used surprisingly infrequently in the genetics literature; and in his exposition of the theory of the gene, Morgan does not explicitly state that genes *caused* characters, only that the latter are *referable to* the former (Darden, 1991, 182). On the other hand, he implicitly uses causal language more than once. I will give three examples. Firstly, he hypothesizes that the tall variety of pea plants "contains in its germ-cells something that *makes* the plants tall" (Morgan, 1928, 2, my emphasis). Secondly, he distinguishes between red-*producing* genes and white-*producing* genes in the crosses of four-o'clock mentioned above (see fig. 6 in Morgan, 1928, 8). Finally, he writes that

> [i]t must not be supposed [...] that mutant changes *produce* only a single striking or even a single small change in one particular part of the body. On the contrary, [...] even in those cases where one part is especially modified, other *effects* are commonly present in several or in all parts of the body." (Morgan, 1928, 315, my emphasis)

'Making' and 'producing' clearly are causal verbs. 'Effect' clearly is a causal noun. These examples should suffice to show that, according to Morgan, genes cause characters (see also Darden, 1991, 182–183).

Let us turn now to the relations between the paired genes and the germ-cells. In "Chapter III. The mechanism of heredity" and "Chapter IV. Chromosomes and genes", Morgan (1928) lists overwhelming cytological evidence leading to the conclusion that $GT_i \rightarrow GC_i$ $(i = 1, 2, 3)$ and that $GC_1 \rightarrow GT_3 \leftarrow GC_2$, i.e. that the parental gametes together determine the filial genotype. Firstly, the chromosomes are the bearers of the genes.

> [...] evidence has accumulated supporting the view that the chromosomes are the bearers of the hereditary elements or genes, and this evidence has steadily grown stronger each year. (Morgan, 1928, 45)

Secondly, somatic cells have a diploid number of chromosomes, half of which come from the father, the other half from the mother.

> [...] there is a double set of chromosomes in each cell of the body and in the early stages of the germ-cells. The evidence of this duality came from observations on differences in the sizes of the chromosomes. Whenever recognizable differences exist there are two chromosomes of each kind in the somatic cells and one of each in the germ cells after maturation. One member of each

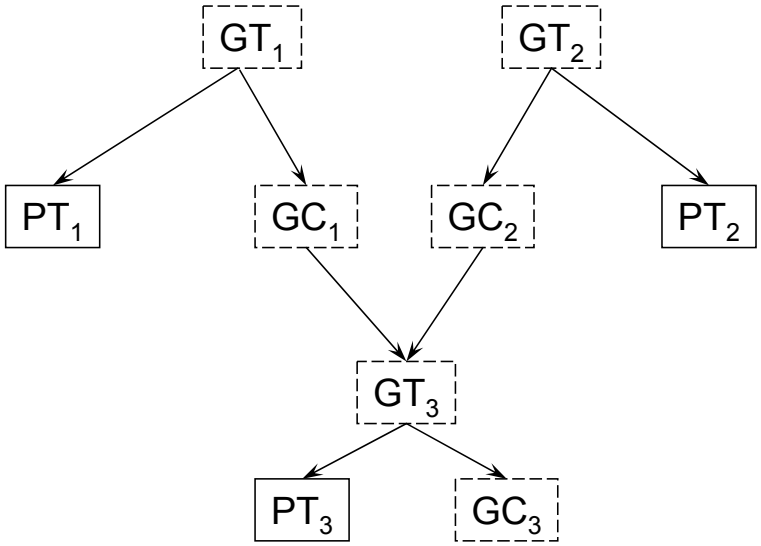

Figure 4.1: *The causal structure of classical genetics*

> kind has been shown to come from the father and the other from the mother. (Morgan, 1928, 32)

Finally, the germ cells receive a haploid number of chromosomes.

> Toward the end of the ripening period of the germ-cells, chromosomes of the same size come together in pairs. This is followed by a division of the cell, when the members of each pair go into opposite cells. Each mature germ-cell comes to contain only one set of chromosomes [...]. (Morgan, 1928, 33)

Together, these claims give rise to the structure presented in figure 4.1. The GC_i and the GT_i-variables are dotted to show that they are **CG**-theoretical terms in the structuralist sense (where **CG** is the theory of classical genetics). The variables PT_i are **CG**-non-theoretical and hence are not dotted. I will first present the structuralist criterion for theoreticity. Then I will argue that *phenotype* is **CG**-non-theoretical, whereas *genotype* and *gametic make-up* are **CG**-theoretical. I will also show that this distinction is relevant in at least three respects.

Definition 4.1 (T-theoreticity) *Informal criterion: a concept whose determination involves some kind of measurement will be called theoretical with respect to theory* $\boldsymbol{T}$ *if all methods of measurement involved in its determination have to be conceived as models of* $\boldsymbol{T}$ *[i.e. invoke the laws of* $\boldsymbol{T}$ *– cf. footnote 8] or as presupposing some models of* $\boldsymbol{T}$*. (Balzer et al., 1987, 50)*

Semi-formal criterion: a concept t is called theoretical relative to theory ***T*** *(or just* ***T****-theoretical) iff every determination of (a relation belonging to) t in any application of* ***T*** *presupposes the existence of at least one actual model of* ***T****. (Balzer et al., 1987, 55)*

Phenotype is a **CG**-non-theoretical concept. One can tell whether some pea plant is tall or short without invoking any of the principles of classical genetics. (Note that this does not commit us to some naive version of direct observability. The MN blood groups in humans provide a nice example of a phenotype the observation of which is relatively complicated. The MN blood groups are characterized by antigens on the surface of red blood cells (Klug et al., 2006, 69–70). Detecting the antigens is less straightforward than measuring the height of pea plants, *but nevertheless it can be done without relying on classical genetics.*) By contrast, *genotype* and *gametic make-up* are **CG**-theoretical concepts. Until the advent of molecular genetics and genomics, there was no way to observe the genotype without relying on phenotypes and the principles of classical genetics.

In the context of classical genetics, the structuralist distinction between **T**-theoretical and **T**-non-theoretical concepts is relevant in at least three respects. Firstly, it will emerge in the following sections that genetic explanations are to a certain extent abductive. Their abductive character results from the **CG**-theoretical nature of the genotype. (This point is neither stressed by Balzer et al. (1987), nor by Balzer and Lorenzano (2000).) Secondly, the history of classical genetics is replete with theory changes to account for recalcitrant anomalies. In Darden's framework, these are called *model anomalies*, i.e. anomalies that require a change in the theory. Such changes either involved the alteration of a typical explanatory pattern or the addition of one or more new patterns and were most often related to changes in the concept of the genotype.[7] (Darden, 1991, 199–201) Given the **CG**-theoretical character of the genotype, such changes run the risk of being *ad hoc*. Their *ad hoc*ness was reduced in several ways: by looking for inter-theoretic relations with cytology, by systematically applying the resulting explanatory pattern in different contexts, etc. Thirdly, within the structuralist framework there is a straightforward relation between **T**-non-theoretical concepts and data. Data report on the theory's **T**-non-theoretical parts.[8] This is clearly revealed in the case of classical genetics: the "numeri-

[7]Model anomalies are contrasted with *monster anomalies*, which do not require such a change (see section 4.14.1 for an example of monster anomalies).

[8]The relations between theories, theoreticity, and data are as follows within the structuralist approach. The set of *potential models* $\mathbf{M_p}(\mathbf{T})$ of a theory **T** consists of those set-theoretic structures that can be subsumed under that theory's conceptual framework

cal data" on which T.H. Morgan lays so much stress (cf. section 4.2) always concern phenotypic distributions.

In the following sections I will show how each of the exemplars discussed by Morgan can be described by means of a credal net. I will also show that these credal nets have graphs that are isomorphic to the graph in figure 4.1. This figure represents the causal scheme containing all models of classical genetics. (Note that not all members of this causal scheme are models of classical genetics. Causal schemes are similar to sets of potential models in the language of Balzer et al. (1987). Not all potential models of a theory are actual models of that theory.) From figure 4.1, causal nets $\mathcal{B} = \langle\langle V, E\rangle, P\rangle$ can be obtained by specifying V and P such that these $\mathcal{B}$'s explain the data at hand. This picture strongly resembles Kitcher's views on unification and explanation (Kitcher, 1989). But, as I will argue, it surpasses these views on several interesting points. It is also *grosso modo* in line with (Pearl, 2000, 202–207). Pearl defines a *causal theory* as a set of causal worlds. A *causal world* consists of a causal model and a particular realization of the exogenous variables. A *causal model*, in his view, consists of a set of variables and a set

(Balzer et al., 1987, 15–17). The set of its *models* (or actual models) $\mathbf{M}(\mathbf{T})$ contains those potential models that moreover satisfy its laws (Balzer et al., 1987, 3, 15–17). Finally, the set of its *partial potential models* $\mathbf{M_{pp}}(\mathbf{T})$ consists of those set-theoretic structures that contain no $\mathbf{T}$-theoretical concepts as in definition 4.1 (where such concepts take the form of relations defined on the domain or base sets) but that can be 'extended' to potential models by adding suitable ($\mathbf{T}$-theoretical) relations (Balzer et al., 1987, 56–57). Finally, the authors take it that a theory's intended applications have the structure of partial potential models: $\mathbf{I} \subseteq \mathbf{M_{pp}}$ (where $\mathbf{I}$ is the set of intended applications) (Balzer et al., 1987, 86–89).

Data describe intended applications and are 'explained' by the theory.

> When confronted with some given "data" or "phenomena" [the authors make no distinction between both concepts] we might want to use a theory $\mathbf{T}$ to "understand" them, to "explain" them, to "predict" them, – in short, we might want to apply $\mathbf{T}$ to those data. To do this, the first thing we try is to conceptualize the domain $\mathbf{I}$ of data in terms of $\mathbf{T}$, i.e. to use the concepts appearing in potential models of $\mathbf{T}$ to refer to $\mathbf{I}$. We create a potential model of $\mathbf{T}$ for $\mathbf{I}$ [by extending the partial potential model at hand in a suitable way]. This is the more "conceptual" aspect of the application of a theory. The next step is to make an assertion about $\mathbf{I}$ in terms of $\mathbf{T}$ – an assertion with empirically testable consequences. We then assert that $\mathbf{I}$ satisfies the fundamental laws of $\mathbf{T}$, which, of course, only makes sense if $\mathbf{I}$ has already been conceptualized in terms of $\mathbf{T}$. In other words, we make the empirical assertion that the potential model considered is also an *actual* model of $\mathbf{T}$. This empirical assertion can be either true or false. If it turns out to be true, we can say that we have applied $\mathbf{T}$ to $\mathbf{I}$ successfully. (Balzer et al., 1987, 23, original emphasis)

of functional relations between these variables. But whereas Pearl requires that for all endogenous variables U^i in a causal world there is a $u^j \in [U^i]$ such that $P(u^j) = 1$ (or equivalently, that there is a $u \in [U]$ such that $P(u) = 1$, where U is the set of exogenous variables), I will not. (A variable is *exogenous* in a graph G if it has no parents in that graph. It is *endogenous* otherwise.)

Before we turn to the first of Morgan's exemplars, let me briefly dwell on the representational capabilities of my framework in relation to the concept of the gene. Whereas it will allow us to elucidate the possible structural and probabilistic relations between the genes (or the genotype) on the one hand, and the phenotype and the gametic makeup on the other hand, it is inapt to fully reveal the nature of the genes. In other words, in my framework it can be seen how genes are related to e.g. phenotypic traits. But it cannot be seen what genes *are*. In the history of genetics, many different (explicit or implicit) views on what genes are have relieved each other. Although it has extensively been questioned whether Mendel indeed invoked invisible elements to explain his phenotypic distributions (cf. Darden, 1991, 40–42), there is good evidence that he did (see footnote 11). According to Meijer (1983, section 5.3), there are some indirect indications that Mendel did not conceive of genes as 'particles' but as discrete, uncountable 'fluids'. By discrete it is meant that in the hybrid the different fluids (responsible for opposing traits) do not mix together (cf. oil and water), so that the respective traits do not blend either. By uncountable it is meant that when two identical fluids come together, they only count as one. Later, Bateson conceived of genes as 'vibrations', 'vortices' or 'forces' (Darden, 1991, 47), or as 'enzymes' (where it was not known that enzymes were proteins (Darden, 1991, 210)). After he opposed Mendelism and the chromosome theory for a long time (see Morgan, 1909, 1910), T.H. Morgan became one of the most important proponents of both, and the Morgan group introduced the view that genes were like 'beads on a string' (see also Darden, 1991, *passim*).

My framework has two important strengths in this respect. Firstly, all these views on what genes are, are compatible with it. So it is flexible enough to model different accounts of classical genetics. Secondly, notwithstanding this flexibility, my framework does allow to distinguish between these different views. To the extent they generate divergent predictions that may be tested empirically, their internal differences can be represented in my framework.

4.5 Of tall and short pea plants

Morgan's first exemplar consists of monohybrid crosses with complete dominance performed by Mendel. Recall that in a monohybrid cross, only one pair of opposing characteristics is observed. Mendel crossed a tall variety of edible pea (*Pisum*) with a short variety.[9] The tall plants he used were true-breeding. A plant is true-breeding for some trait if, when self-fertilized, it only produces offspring with this trait. (This definition is not watertight. All the offspring of non-true-breeding plants may, by accident, have the parental phenotype.) True-breeding plants may also be called pure (cf. Morgan, 1928, 4).[10]

Mendel observed that when true-breeding tall plants are crossed with short plants, all the offspring or hybrids (F_1) are tall (cross 1). It did not matter whether the tall plants produced pollen and the short plants produced eggs or vice versa. Reciprocal crosses gave identical results (cf. Mendel, 1865, 8–9; see also page 68). In a second cross (cross 2), self-fertilization of the F_1-generation resulted in offspring (F_2) 75% of which was tall, the other 25% being short. Finally, when the F_1 hybrids (pollen plants) were back-crossed to the short plants (egg plants), 50% of the resulting off-spring $F_{2'}$ was tall, the other 50% was short (cross 3). Cross 3 served as a test cross for the validation of his explanatory principles.[11] How can these phenotypic

[9] *Pisum* plants are one kind of model organisms that were frequently used in classical genetics. Other kinds of model organisms were i.a. *Drosophila* (fruit flies), mice, ... Model organisms can be regarded as (representational) models (see Leonelli, 2007).

[10] True-breeding Pisum plants were sold by seed dealers (the procedure by which these true-breeding plants were obtained is described in section 6.2):

> From several seed dealers a total of 34 more or less distinct varieties of peas were procured and subjected to two years of testing. In one variety a few markedly deviating forms were noticed among a fairly large number of like plants. These, however, did not vary in the following year and were exactly like another variety obtained from the same seed dealer; no doubt the seeds had been accidentally mixed. All other varieties yielded quite similar and constant offspring; at least during the two test years no essential change could be noticed. Twenty-two of these varieties were selected for fertilization and planted annually throughout the entire experimental period. They remained stable without exception. (Mendel, 1865, 4)

Three decades later, Correns writes in a footnote in his paper "G. Mendel's Law Concerning the Behavior of Progeny of Varietal Hybrids," that

> The names of the varieties given in this chapter are those which I received from [the seed firm] Haage and Schmidt in Erfurt. (Correns, 1900, 120)

[11] There is much debate regarding the question whether Mendel indeed invoked invisible elements to explain his phenotypic distributions (cf. Darden, 1991, 40–42). In contempo-

distributions be accounted for? It will turn out that the results of each can be accounted for by means of a causal net whose graph is isomorphic to the graph in figure 4.1.

Let us first give an appropriate interpretation to each of the nodes in figure 4.1 by specifying their respective sets of possible values. The plants in each of the three groups in these crosses have one of the following observable traits: either they are tall ($PT_i = tall$) or short ($PT_i = short$). These traits are referable to paired alleles in the germinal material. Let t and s denote a 'tall-producing' allele and a 'short-producing' allele, respectively.[12] There are three possible configurations of paired genes, i.e. three possible values for the variables GT_i, viz. $tt, ts,$ and ss. Finally, the gametes or germ-cells contain one gene of each pair. Hence, the variables GC_i may assume the values t or s.

These specifications can be summarized as follows. Let $V^1 = \{PT_1, \ldots, GC_3\}$ be a set of variables corresponding to the nodes in figure 4.1. Let $[PT_i] = \{tall, short\}, [GT_i] = \{tt, ts, ss\}$, and $[GC_i] = \{t, s\}$, $(i = 1, 2, 3)$. Let $G^1 = \langle V^1, E\rangle$ be isomorphic to the graph in figure 4.1 (where $PT_1 \in V^1$ corresponds to the node PT_1 in the graph, ...) This completes the description of G^1. By adding probability distributions to this graph, we obtain causal nets that may represent the phenotypic distributions in crosses 1 to 3. How should the probability distributions look like?

---

rary genetics textbooks, Mendel's views are often cast as if he did. For example, according to Klug et al. (2006), Mendel derived the following postulate:

> *Genetic characters are controlled by unit factors existing in pairs in individual organisms.* (Klug et al., 2006, 41, original emphasis)

Such way of presenting his views has been criticized by historians of science. Nevertheless, I am inclined to think that Mendel indeed invoked some kind of elements *responsible for* the phenotype. Firstly, it is very natural to read his *Versuche* this way (Mendel, 1933). For example, when discussing the experimental results of Kölreuter and Gärtner, he writes the following (see also section 4.9.1):

> Both concur in the opinion that, *in external appearance*, hybrids either maintain a form intermediate between the parental strains or they approach the type of one or the other, sometimes being barely distinguishable from them. (Mendel, 1865, 39, my emphasis)

The subordinate clause in italics would make little sense if Mendel did not distinguish between external appearance and internal make-up. Secondly, Meijer (1983) provides convincing arguments supporting this view. Fortunately, this debate need not concern us here. What is relevant for the present discussion is the fact that Morgan (1928, 3) *attributes to* Mendel the use of cross 3 as a test cross.

[12]Morgan does not use the words "tall-" or "short-producing alleles". But his presentation of inheritance of flower-colour in four-o'clocks is phrased in terms of white- and of red-producing genes (Morgan, 1928, 5–7).

table 1

	PT_i	
GT_i	*tall*	*short*
tt	1.00	0.00
ts	1.00	0.00
ss	0.00	1.00

table 2

	GC_i	
GT_i	*t*	*s*
tt	1.00	0.00
ts	0.50	0.50
ss	0.00	1.00

table 3

		GT_3		
GC_1	GC_2	*tt*	*ts*	*ss*
t	*t*	1.00	0.00	0.00
t	*s*	0.00	1.00	0.00
s	*t*	0.00	1.00	0.00
s	*s*	0.00	0.00	1.00

Conditional probability tables 1–3: these provide constraints for $P \in \mathbb{P}^1$

To explain the phenotypic distributions of crosses 1 to 3, Morgan appeals to the following principles. Firstly, he implicitly assumes that the following relations hold between the genotype and the phenotype of each group: plants that have two tall-producing alleles are tall; hybrids, which have both an allele for tall and one for short, are tall; and plants that have two short-producing alleles are short. These relations correspond to the principle of '*complete dominance*'. They are summarized in conditional probability table 1. This table should be read as follows: in the right-hand column, the probability distribution is given for the variables PT_i $(i = 1, 2, 3)$, conditional on the possible values of their respective graphic parents (i.e. the variables GT_i, $i = 1, 2, 3$), which are listed in the left-hand column. For example, the first row states that $P(PT_i = tall \mid GT_i = tt) = 1.00$ and that $P(PT_i = short \mid GT_i = tt) = 0.00$. Analogously, the second row states that $P(PT_i = tall \mid GT_i = ts) = 1.00$. It expresses the fact that tall is dominant to short and corresponds to the second part of the following quote:

> If the tall variety contains in its germ-cells something that makes the plants tall and if the short variety carries something in its germ-cells that makes the plants short, the hybrid contains both [cf. table 3]; *and since the hybrid is tall it is evident that when both are brought together the tall dominates the short, or, conversely, short is recessive to tall.* (Morgan, 1928, 2–3, my emphasis)

As a second principle, Morgan assumes that

> [i]f the element for tall and the one for short (that are both present in the hybrid) separate in the hybrid when the eggs and pollen grains come to maturity, half the eggs will contain the tall and half the short element [...]. Similarly for the pollen grains. (Morgan, 1928, 3)

This hypothesis he attributes to Gregor Mendel, calling it *Mendel's first law* (Morgan, 1928, 3, 5). With respect to true-breeding plants (in contrast with

the hybrids), he implicitly assumes that all the eggs or pollen grains of plants that have only tall-producing alleles (respectively only short-producing alleles) will contain the tall element (respectively the short element). From Morgan's discussion of the underlying chromosomal mechanism it is evident that Mendel's first law, which is now called the *law of segregation*, not only applies to the hybrids or heterozygotes, but also to the homozygotes.[13] Mendel's first law is summarized in table 2.

As a third principle, Morgan assumes that when an egg and a pollen grain fertilize, their respective elements together make up the genotype of the resulting offspring. I will call this the *principle of combination*.[14] It is summarized in table 3. The relations in tables 1 and 3 are joined in the following quote:

> [...] when tall meets tall a tall plant is produced; when tall meets short a tall plant results; when short meets tall, a tall plant is produced; and when short meets short, a short plant arises. (Morgan, 1928, 3)

Finally, it should be noted that Morgan assumes that *chance fertilization* occurs, i.e. that GC_1 is probabilistically independent from GC_2. This assumption is incorporated in G^1, as GC_1 and GC_2 are d-separated by the empty set in the graph in figure 4.1 (to which G^1 is isomorphic).

[13] In the following quote, part of which I already used on page 60, Morgan does not restrict segregation to chromosomes carrying different alleles or Mendelian units.

> Toward the end of the ripening period of the germ-cells, chromosomes of the same size come together in pairs. This is followed by a division of the cell, when the members of each pair go into opposite cells. Each mature germ-cell comes to contain only one set of chromosomes, [...]. This behavior of the chromosomes in the maturation stages parallels Mendel's first law. A chromosome derived from the father separates from a chromosome derived from the mother for each pair of chromosomes. The germ-cells that result contain one chromosome of each kind. Taking the chromosomes in pairs we may say, half of the germ-cells, when mature, contain one member of each pair, the other half the mates of those chromosomes, pair for pair. *If one substitutes Mendelian units for chromosomes, the statement is the same.* (Morgan, 1928, 33–34, my emphasis)

[14] The principle of combination should not be confused with the function $COMBINATOR$ in Balzer and Lorenzano (2000). This function represents the transition from the parental genotypes to the genotypes of the progeny. My principle of combination only deals with the relation between the parental gametes and the filial genotypes. In Balzer and Lorenzano (2000), the role of the gametes and their genetic make-up is not made explicit.

Together, the graph G^1 and tables 1–3 give rise to a credal set:

$$\mathbb{P}^1(V^1) = \{P(V^1) \mid P(V^1) \text{ satisfies tables 1–3}\}$$

Joined with G^1, this defines a credal net :

$$\mathfrak{B}^1 = \langle G^1, \mathbb{P}^1 \rangle$$

Recall that a credal net is a set of causal nets defined over the same graph. I will now show that each of the crosses cited above can be described or explained by some $\mathcal{B} \in \mathfrak{B}^1$.

Let us first turn to cross 1. In one variant of cross 1, the pollen plants are true-breeding tall and the egg plants are short: $PT_1 = tall, PT_2 = short$. (Recall that $GT_1, PT_1,$ and GC_1 correspond to the paternal plants and that $GT_2, PT_2,$ and GC_2 correspond to the maternal plants.) If it is assumed that a true-breeding tall plant contains the element for tall twice present (Morgan, 1928, 6), it can be abduced that $GT_1 = tt$ and $GT_2 = ss$ (cf. the **CG**-theoretical nature of the variables GT_i). If we denote the probability distribution that corresponds to this cross by $P^1(V^1)$, or briefly P^1, we can express this as follows: $P^1(PT_1 = tall) = P^1(PT_2 = short) = 1.00,$ and $P^1(GT_1 = tt) = P^1(GT_2 = ss) = 1.00$. Assume now that $P^1 \in \mathbb{P}^1$, so that it satisfies the constraints in tables 1–3.[15] Then it follows that $P^1(GC_1 = t) = P^1(GC_2 = s) = 1.00$ (from table 2), and that $P^1(GT_3 = ts) = 1.00$ (from table 3). But then $P^1(PT_3 = tall) = 1.00$ (from table 1). Hence it is 'explained' why all the F_1 hybrids are tall. (I will provide a more elaborate account of explanation by means of causal nets in chapter 5.)

Mendel also performed another variant of cross 1, in which the pollen plants were short and the egg plants were pure tall. Reciprocal crosses gave identical results.

> All experiments proved further that it is entirely immaterial whether the dominating trait belongs to the seed or pollen plant; the form of the hybrid is identical in both cases. (Mendel, 1865, 9)

This result nicely fits my analysis: there is a $P^2 \in \mathbb{P}^1$ such that $P^2(PT_1 = short) = P^2(PT_2 = tall) = 1.00,$ and $P^2(GT_1 = ss) = P^2(GT_2 = tt) = 1.00$. Given that $P^2 \in \mathbb{P}^1$, $P^2(PT_3 = tall) = 1.00$. This corresponds to the data.

In short, both variants of cross 1 can be described by a causal net. The first variant by $\mathcal{B}^1 = \langle G^1, P^1 \rangle$. The second variant by $\mathcal{B}^2 = \langle G^1, P^2 \rangle$. Since $P^1, P^2 \in \mathbb{P}^1$, it follows that $\mathcal{B}^1, \mathcal{B}^2 \in \mathfrak{B}^1$.

[15] I use 'constraints' in the sense of Williamson (2005, 84), not in the sense of Balzer et al. (1987, 46–47).

Analogously, the phenotypic distributions in cross 2 can be explained by $\mathcal{B}^3 = \langle G^1, P^3 \rangle$, where $P^3 \in \mathbb{P}^1$. In cross 2, the F_1 hybrids are selfed and they produce F_2 offspring 75% of which is tall and 25% of which is short. Obviously, if $P^3(GT_1 = ts) = P^3(GT_2 = ts) = 1.00$ and if $P^3 \in \mathbb{P}^1$, then $P^3(GT_3 = tt) = 0.25, P^3(GT_3 = ts) = 0.50$, and $P^3(GT_3 = ss) = 0.25$, so that $P^3(PT_3 = tall) = 0.75$ and $P^3(PT_3 = short) = 0.25$.[16] This corresponds to the data.

Finally, in cross 3 the F_1 hybrids are back-crossed with the recessive parental plants. Let $P^4(GT_1 = ts) = P^4(GT_2 = ss) = 1.00$. If $P^4 \in \mathbb{P}^1$, then we would expect that $P^4(PT_3 = tall) = 0.50$. "The results confirm the expectation." (Morgan, 1928, 4) Hence, cross 3 can be represented by $\mathcal{B}^4 = \langle G^1, P^4 \rangle \in \mathfrak{B}^1$ (where $P^4 \in \mathbb{P}^1$).

So far we have seen that a number of crosses with tall and short pea plants can be represented by distinct causal nets that belong to one common credal net $\mathfrak{B}^1$. Each of these causal nets is a model of classical genetics, more specifically of a reciprocal monohybrid cross with complete dominance. We have also seen that these models can be used to explain data or phenomena (phenotypic distributions) obtained in experimental crosses. (I use the word 'experimental' in a loose sense here. Also, I will distinguish between data and phenomena hereafter.)

This account of explanation is similar to that in Balzer et al. (1987), but surpasses it in an interesting way in that it explicitly accounts for the use of causal knowledge in explanation. According to Balzer et al., to explain data or phenomena[17] one (i) extends the corresponding partial potential model to a potential model by adding suitable **T**-theoretic relations, and (ii) shows that this potential model is an actual model of **T** (cf. footnote 8). In a like manner, I have shown that the phenotypic distributions in crosses 1 to 3 can be extended to a causal net $\mathcal{B} = \langle G, P \rangle$, where G (more specifically, G^1) 'contains' the conceptual framework of classical genetics and where P satisfies the laws of classical genetics (tables 1 to 3). (For a more elaborate discussion of explanation in genetics, see chapter 5.)

In the following sections, I will show how crosses with other genera, and crosses on other characters in *Pisum* can be represented by causal nets that fit in analogous credal nets, and I will explicate the set-theoretical relations between these respective credal nets.

First, however, we should have a closer look at the extensions of $\mathfrak{B}^1$ and $\mathbb{P}^1$. The extension of the former is completely tied to the extension of the

[16] Note that, at least in the artificial cases described by Mendel and Morgan, selfing or self-fertilization does not contradict the assumption of random fertilization.

[17] Recall that they do not distinguish between these concepts.

latter, as $\mathfrak{B}^1 = \{\mathcal{B} = \langle G^1, P\rangle \mid P \in \mathbb{P}^1\}$. $\mathfrak{B}^1$ is a separately specified, finitely generated, closed and convex credal net. Since it is convex, it has uncountably many elements.

A credal net $\mathfrak{B} = \langle\langle V, E\rangle, \mathbb{P}\rangle$ is finitely generated if $\mathbb{P}$ is finitely generated, i.e. if it is obtained as the convex hull of a finite number of (conditional) probability distributions (Antonucci and Zaffalon, LSCN, section 3). The convex hull of a set of probability distributions $\mathbb{P}$ is the intersection of all convex sets containing $\mathbb{P}$ (Walley, 1991, 611).[18] $\mathfrak{B}$ is separately specified if $\mathbb{P}$ is obtained by local or separate specifications, i.e. if for each variable $X_i \in V$, $P(x_i \mid pa_{ij}) = [r, s]$ is given for each value $x_i \in [X_i]$ and for each value $pa_{ij} \in [PA_i]$, where PA_i is the set of graphic parents of X_i. (Note that I also could have used an open interval, e.g. $]r, s[$. $\mathbb{P}$ would then be an open credal set.) This probability distribution may be imprecise or, as a limit case, it may be precise for some x_i and pa_{ij}. In separately specified credal nets, all conditional probabilities are determined locally: $P(x_i \mid pa_{ij})$ is allowed to take any value in $[r, s]$, independently of what happens elsewhere in the credal net. (Antonucci and Zaffalon, LSCN, section 4)

Strictly speaking, $\mathfrak{B}^1$ is a degenerate credal net. All probability specifications in tables 1 to 3 are precise. All imprecision is relegated to the root variables GT_1 and GT_2. (In later sections I will introduce genuine credal nets.) In experimental contexts only a few typical distributions over these variables were studied. These correspond to the vertices, in the sense of extreme points, of $\mathbb{P}^1$ (cf. Antonucci and Zaffalon, LSCN, section 3). In experimental contexts, the organisms in each parental group mostly had the same genotype (or they were deemed to have the same genotype). But in principle any mixed group can serve as a parental group in a genetic cross. For example, the first parental group may consist of $f\%$ true-breeding tall plants, $g\%$ hybrids and $(100 - f - g)\%$ short plants, and the second parental group may consist of $f'\%$ true-breeding tall plants, $g'\%$ hybrids and $(100 - f' - g')\%$ short plants. The phenotypic distribution of the resulting off-spring can then be determined (on the assumption of random mating) by means of the principles discussed, or in other words, by selecting an appropriate $\mathcal{B} \in \mathfrak{B}^1$.

[18] Recall that a credal set $\mathbb{P}$ is convex if, for any measures $P, P' \in \mathbb{P}$, the measure $\alpha P + (1 - \alpha)P' \in \mathbb{P}$ for any $\alpha \in [0, 1]$. Recall also that I did not require that the members of $\mathfrak{B}^1$ are faithful.

4.6 Monohybrid crosses with complete dominance

The members of $\mathfrak{B}^1$ can be used to describe monohybrid crosses with tall and short *Pisum* plants, where tall is completely dominant with respect to short. In the history of classical genetics, many other monohybrid crosses with complete dominance have been reported. In his *Versuche über Pflanzen-Hybriden*, Mendel discussed six more such crosses with *Pisum* (Mendel, 1865, 5–17). He selected traits relating to the shape of the ripe seeds (round versus angular or wrinkled), the colouration of the seed albumen (yellow versus green), ... Likewise, Morgan (1928, 4–5) discusses the inheritance of eye colour in man (blue eyes versus brown eyes) as an example of complete dominance.

These crosses cannot be described by the members of $\mathfrak{B}^1$, since the set of variables V^1 on which they are defined is tied to the case of stem length in *Pisum*. However, by slightly changing the members in V^1, other monohybrid crosses with complete dominance can be accounted for.

Consider the case of eye colour in man. Two phenotypic traits are studied, where *brown* is dominant to *blue*. These traits are referable to paired genes in the germinal material, which Morgan denotes by *br* and *bl*. (In the works of Morgan and his co-workers there is a large variety qua symbols used to denote alleles.) The results of crosses with blue and brown eyes in man can be accounted for with the help of the principles discussed in section 4.5, viz. Mendel's first law, complete dominance, etc. Thus, they are very similar to the case of stem length in *Pisum*. This similarity can be explicated in a precise way. There is a credal net, say $\mathfrak{B}^2$, that is both *value-isomorphic* and *distribution-identical* to $\mathfrak{B}^1$. The members of $\mathfrak{B}^2$ can be used to describe or explain crosses on eye colour in man.

Let us see how $\mathfrak{B}^2$ can be characterized. $V^2 = \{PT'_1, \ldots, \ldots, GC'_3\}$ is a set of variables corresponding to the nodes in figure 4.1, where $[PT'_i] = \{brown, blue\}, [GT'_i] = \{brbr, brbl, blbl\}$, and $[GC'_i] = \{br, bl\}, (i = 1, 2, 3)$. Let $b : V^1 \rightarrow V^2$ be a bijection such that $b(PT_i) = PT'_i, b(GT_i) = GT'_i$, $b(GC_i) = GC'_i$. Let $b' : [[V^1]] \rightarrow [[V^2]]$ be a bijection such that for any $A \in V^1$ and $a \in [[V^1]]$, $a \in [A]$ if and only if $b'(a) \in [b(A)]$. More specifically, let $b'(tall) = brown, b'(short) = blue, b'(tt) = brbr, \ldots, b'(s) = bl$. Obviously, V^2 is value-isomorphic with V^1 (definition 3.3). It follows that $G^2 = \langle V^2, E\rangle$ is value-isomorphic to $G^1 = \langle V^1, E\rangle$ (definition 3.4).

Let $\mathbb{P}^2$ be the set of probability distributions over V^2 that satisfy constraints analogous to tables 1–3 and that are causally Markov to G^2. More specifically, where the members of $\mathbb{P}^1$ satisfy $P(a \mid pa(A)) = r$, let the mem-

bers of $\mathbb{P}^2$ satisfy $P(b'(a) \mid b'(pa(A))) = r$ for any $A \in V^1, a \in [A]$, and $pa(A) \in [PA(A)]$.[19] ($PA(A) \subset V^1$ is the set of A's graphic parents in G^1.) It follows that $\mathbb{P}^2$ is distribution-identical to $\mathbb{P}^1$ (definition 3.14).

Let $\mathfrak{B}^2 = \langle G^2, \mathbb{P}^2 \rangle$. $\mathfrak{B}^2$ is *value-isomorphic* to $\mathfrak{B}^1$ (definition 3.15). Hence $\mathfrak{B}^2 \subset \mathfrak{C}([\mathfrak{B}^1])$, or equivalently, $\mathfrak{C}([\mathfrak{B}^1]) = \mathfrak{C}([\mathfrak{B}^2])$. $\mathfrak{B}^2$ is also *distribution-identical* to $\mathfrak{B}^1$ (definition 3.16). Thus for any possible cross of tall and short pea plants, there is an analogous cross on eye colour in man.

4.7 Exemplars, general argument patterns, and generic credal nets

The relations between crosses on stem length in pea plants and crosses on eye colour in man can be generalized to all monohybrid crosses with complete dominance. Therefore, let $\mathfrak{B}^\alpha = \langle G^\alpha, \mathbb{P}^\alpha \rangle$ (with $G^\alpha = \langle V^\alpha, E \rangle$) be a generic credal net that is distribution-identical to $\mathfrak{B}^1$. Any monohybrid cross with complete dominance can be described by means of a causal net that is distribution-identical to some member of $\mathfrak{B}^\alpha$, by filling in V^α and $\mathbb{P}^\alpha$. $\mathfrak{B}^\alpha$ is a general representation of these crosses.

By defining $\mathfrak{B}^\alpha = \langle G^\alpha, \mathbb{P}^\alpha \rangle$ in this way, I refer indirectly to $\mathcal{B}^1, \ldots, \mathcal{B}^4$. This is in line with the way Morgan presents the theory of classical genetics. He expounds this theory by providing concrete examples of each type of process (in casu monohybrid crosses with complete dominance) (Darden, 1991, 195). More specifically, he uses $\mathcal{B}^1, \ldots, \mathcal{B}^4$ to introduce Mendel's first law, the principle of complete dominance, etc.

The concepts of "credal net" (such as $\mathfrak{B}^1$ and $\mathfrak{B}^2$) and "generic credal net" (such as $\mathfrak{B}^\alpha$) can be looked upon as precise *and fruitful* explications of the notions of *exemplar* and *general argument pattern* in the works of Lindley Darden. These notions incorporate aspects from both Kuhnian exemplars and Kitcherian general argument patterns.

Kuhn (1996) characterizes *exemplars* as "[...] concrete problem solutions in which a formalism (such as a mathematical equation) is applied and given empirical grounding." (Darden, 1991, 18) A Kitcherian *general argument pattern* is defined as follows:

> A *schematic sentence* is an expression obtained by replacing some, but not necessarily all, the nonlogical expressions occurring in a sentence with dummy letters. [...] A set of *filling instructions* for a schematic sentence is a set of directions for replacing the dummy

[19] Where $U = \{A_1, \ldots, A_n\} \subseteq V$ and $b : V \to V'$, let $b(U) = \{b(A) \mid A \in U\}$ and $b'([U]) = [b(A_1)] \times \ldots \times [b(A_n)]$.

> letters of the schematic sentence, such that, for each dummy letter, there is a direction that tells us how it should be replaced. [...] A *schematic argument* is a sequence of schematic sentences. A *classification* for a schematic argument is a set of statements describing the inferential characteristics of the schematic argument: it tells us which terms of the sequence are to be regarded as premises, which are inferred from which, what rules of inference are used, and so forth. Finally, a *general argument pattern* is a triple consisting of a schematic argument, a set of sets of filling instructions and a classification for the schematic argument. (Kitcher, 1989, 432)

Kitcher (1989, 438–442) characterizes classical genetics as a set of problem-solving patterns that serve to answer specific types of questions and that each exemplify one and the same general idea. The questions to be answered are:

> What is the expected distribution of phenotypes in a particular generation? Why should we expect to get that distribution? What is the probability that a particular phenotype will result from a particular mating?, and so forth. (Kitcher, 1989, 438)

The general idea common to all patterns is that these questions may be answered

> [...] by making hypotheses about the relevant genes, their phenotypic effects, and their distribution among the individuals in the pedigree. (Kitcher, 1989, 438–439)

He distinguishes between the following patterns (in chronological order): `Mendel`, `Refined Mendel`, `Morgan`, and `Watson-Crick`. The more recent patterns are refinements of the older patterns and they can accommodate previously recalcitrant cases (anomalies). (`Mendel` will be cited as an example on page 121.)

Kuhnian exemplars, Darden argues, may serve to generate Kitcherian argument patterns.

> [They] may serve in the construction of abstract explanatory patterns or schemas [...]. The patterns abstractly characterize mechanisms, which, when they are operating, produce observable data-points as output. Thus, fitting an observation into a pattern is a way of explaining it. A set of exemplary patterns constitutes the explanatory repertoire of Mendelian genetics [...]. (Darden, 1991, 19)

The crosses on stem length in *Pisum* play the role of Dardenian exemplars in Morgan (1928). Darden writes, "[t]he examples supplied model cases. Similar results of similar hybrid crosses could be explained by invoking similar steps and filling in the details about the characters in the specific cross." (Darden, 1991, 195)

Darden's view is precisely explicated in my framework. The causal nets $\mathcal{B}^1, \ldots, \mathcal{B}^4$ are used to introduce, to illustrate, and to test several explanatory principles. Together, these principles give rise to a credal net $\mathfrak{B}^1$ that fully expresses their meaning. $\mathfrak{B}^1$ serves as an exemplar. By abstracting from the precise details in question, a generic credal net $\mathfrak{B}^\alpha$ is obtained. Any reciprocal monohybrid cross with complete dominance (i.e. any similar result of a similar hybrid cross) can be described or explained by means of a causal net that is distribution-identical to some member of $\mathfrak{B}^\alpha$, viz. by filling in the details about the characters at hand. $\mathfrak{B}^\alpha$ is a general argument pattern, or the semantic counterpart of such a general argument pattern, in Darden's sense.

The fruitfulness of this explication will emerge in chapter 5. I will show that my framework provides an interesting account of explanation *cum* unification which improves on those of Kitcher and Darden.

As was to be expected, my framework cannot account for every single detail of Darden's analysis. Darden (1991, 195–196) strongly clings to the role of diagrammatic representations (e.g. pedigree diagrams) in the history of classical genetics. Admittedly, the concrete details of the diagrams used in theory presentation (but also in the creation and development of a theory) may produce interesting, unexpected, fruitful and/or restrictive constraints. Here, I cannot frame all these aspects of diagrammatic representations since I focus on their underlying causal contents and force or translate them into Bayesian nets.

4.8 Monohybrid crosses with incomplete dominance

After presenting Mendel's crosses and those regarding eye colour in man, Morgan turned to crosses that were interestingly different. Whereas in the crosses represented by $\mathfrak{B}^\alpha$, hybrids phenotypically resemble one of their parents (the parent with the dominant trait), in other crosses they don't.

> There are other crosses that give, perhaps, a more striking illustration of Mendel's first law. For instance, when a red and a white-flowered four-o'clock [*Mirabilis jalapa*] are crossed, the

> hybrid [F_1] has pink flowers [...]. If these pink-flowered hybrid plants self-fertilize, some of their offspring (F_2) are red like one grandparent, some of them pink like the hybrid, and others white like the other grandparent, in the ratio of 1:2:1. Here one original parental color is restored when red germ-cell meets red, the other color is restored when white meets white, and the hybrid combinations appear as often as red meets white, or white meets red. All the colored flowered plants in the second generation taken together are to the white-flowered plants as 3:1.
>
> In passing it is important to note two facts. The red and the white F_2 individuals are expected to breed true, because they contain the elements for red, or for white, twice present [...], but the pink F_2 individuals should not breed true, since they are like the first hybrid generation, and contain one red and one white element [...]. All this turns out to be true when these plants are tested. (Morgan, 1928, 5–6)

Morgan's explanation of the phenotypic distributions rests on the following principles. Firstly, as in the case of the previous crosses, he invokes Mendel's first law (or the law of segregation). Recall that he deems this cross even "a more striking example of Mendel's first law." Secondly, as in the case of the previous crosses, he assumes that when two germ-cells fertilize (again, he assumes random fertilization), their respective elements together make up the genotype of the resulting off-spring (the principle of combination). What is new, however, is the relation between genotype and phenotype. Let me jointly label the relations in the quote above the '*principle of incomplete dominance*'.[20]

How can we account for this difference in terms of causal nets? It is easily seen that crosses with incomplete dominance are structurally different from crosses with complete dominance in the following sense: crosses on flower colour in four-o'clocks cannot be described or explained by means of causal nets that are value-isomorphic (let alone distribution-identical) to members of $\mathfrak{B}^{\alpha}$, since the corresponding set of variables V^3 (cf. infra) is not value-isomorphic to V^{α}. The flower colour (character) of four-o'clocks has three possible states (traits): white, pink, and red. So let me define $V^3 = \{PT_1, \ldots, \ldots, GC_3\}$, where $[PT_i] = \{red, pink, white\}, [GT_i] = \{rr,$

[20]In the beginning of the 20th century, the concept of "incomplete dominance" was used ambiguously. Sometimes, it was used for actual cases of incomplete dominance (where the parental characters are absent in the F_1-generation). Other times it was used for cases of multiple factors. Morgan et al. (1915, 27–32) have sorted out these different cases (cf. Darden, 1991, 68–69).

table 4

	PT_i		
GT_i	*red*	*pink*	*white*
rr	1.00	0.00	0.00
rw	0.00	1.00	0.00
ww	0.00	0.00	1.00

table 5

	GC_i	
GT_i	*r*	*w*
rr	1.00	0.00
rw	0.50	0.50
ww	0.00	1.00

table 6

		GT_3		
GC_1	GC_2	*rr*	*rw*	*ww*
r	*r*	1.00	0.00	0.00
r	*w*	0.00	1.00	0.00
w	*r*	0.00	1.00	0.00
w	*w*	0.00	0.00	1.00

Conditional probability tables 4–6: these provide constraints for $P \in \mathbb{P}^3$

$rw, ww\}$, and $[GC_i] = \{r, w\}, (i = 1, 2, 3)$.[21] Let $G^3 = \langle V^3, E\rangle$ be isomorphic to the graph in figure 4.1. What about the probability distributions over G^3? The constraints generated by the principle of incomplete dominance, by Mendel's first law, and by the principle of combination are summarized in tables 4, 5 and 6, respectively. Together, they determine the credal set $\mathbb{P}^3(V^3)$ and the credal net $\mathfrak{B}^3 = \langle G^3, \mathbb{P}^3\rangle$.

In line with section 4.6, we can show how the crosses cited by Morgan can be described or explained by means of members of $\mathfrak{B}^3$. In a first cross, a red flowered four-o'clock is joined with a white-flowered four-o'clock. Morgan does not clarify which of the two is the pollen producing plant but, as reciprocal crosses again give the same results, we may assume that the pollen producing plant is red.[22] Now consider a distribution $P^1(V^3)$ such that $P^1(PT_1 = red) = P^1(PT_2 = white) = 1.00$.[23] If $P^1(V^3) \in \mathbb{P}^3$, it follows that $P^1(PT_3 = pink) = 1.00$. This corresponds to the data, so it may be assumed that $\mathcal{B}^1 = \langle G^3, P^1\rangle$ (with $P^1 \in \mathbb{P}^3$) represents this first cross. In the second cross, the F_1 hybrids are selfed, so let $P^2(V^3)$ be such that $P^2(PT_1 = pink) = P^2(PT_2 = pink) = 1.00$. If $P^2(V^3) \in \mathbb{P}^3$, it follows that $P^2(PT_3 = red) = 0.25, P^2(PT_3 = pink) = 0.50$, and $P^2(PT_3 = white) = 0.25$. Hence, the 1:2:1 ratio can be explained by assuming that this cross is rightly described by $\mathcal{B}^2 = \langle G^3, P^2\rangle$ (where $P^2 \in \mathbb{P}^3$).

As we saw in section 4.5, Morgan cited *test crosses* to confirm the law of segregation. With respect to incomplete dominance, he considers three such crosses: selfing of the red F_2 plants, selfing of the white F_2 plants, and selfing of the pink F_2 hybrids. All predictions turned out to be true. In other words, for all three crosses there is a $P^i \in \mathbb{P}^3$ such that $\mathcal{B}^i = \langle G^3, P^i\rangle$ explains the

[21] Morgan did not use r and w for the red- and white-producing genes (alleles). Instead he used small black circles and small white circles in his diagrammatic representation of these crosses (Morgan, 1928, 7).

[22] As in section 4.5, the reverse assumption would result in a different causal net in $\mathfrak{B}^3$.

[23] Note that $P^1(V^1)$ and $P^1(V^3)$ are different distributions, given that V^1 and V^3 are different sets of variables (they are not even value-isomorphic). Likewise, $\mathcal{B}^1 = \langle G^3, P^1\rangle$ with $P^1 \in \mathbb{P}^3$ should not be confused with $\mathcal{B}^1 = \langle G^1, P^1\rangle$ with $P^1 \in \mathbb{P}^1$.

data. As a result, the adequacy of $\mathbb{P}^3$ and of $\mathfrak{B}^3$ was corroborated.[24]

Flower colour in four-o'clocks is not the only character that shows incomplete dominance. Like results have been obtained with snapdragons (*Antirrhinum*), where crosses of red- and white-flowered plants give rise to pink-flowered offspring (Klug et al., 2006, 68–69).

These relations can be generalized to all monohybrid crosses with incomplete dominance. Therefore, let $\mathfrak{B}^\beta = \langle G^\beta, \mathbb{P}^\beta \rangle$ (with $G^\beta = \langle V^\beta, E \rangle$) be a generic credal net such that V^β is value-isomorphic to V^3, G^β is value-isomorphic to G^3, and $\mathfrak{B}^\beta$ is distribution-identical to $\mathfrak{B}^3$. Any monohybrid cross with incomplete dominance can be described by means of a causal net that is distribution-identical to some member of $\mathfrak{B}^\beta$. $\mathfrak{B}^\beta$ is a generic net representing these crosses.[25] Clearly, complete dominance and incomplete dominance are completely different, as $\mathfrak{B}^\alpha \cap \mathfrak{B}^\beta = \emptyset$.

4.9 Anomalies, theory-elements, and the logical subordination of the Law of Dominance

The history of classical genetics is replete with theory changes to account for recalcitrant anomalies. In section 4.9.1, I will show how the status of complete dominance was heavily debated in the beginning of the 20th century, how cases in which complete dominance failed were accounted for, and how Bateson drew attention to what he considered the core of Mendel's theory,

[24] I use corroboration in a loose sense here.

[25] It should be noted that monohybrid crosses with codominance, instead of incomplete dominance, can also be represented by causal nets that are distribution-identical to some member of $\mathfrak{B}^\beta$. As we saw in section 4.3, the MN blood group, discovered by Karl Landsteiner and Philip Levin in 1927, provides an example of codominance. In 1930, F. Schiff expressed his hope that the inheritance of MN blood groups would fit $\mathfrak{B}^\beta$ as follows:

> Eigene Beobachtungen an Familien sowie Müttern und Kindern sprechen ebenso wie Massenuntersuchungen (1420 Berliner, 180 Wolgadeutsche) dafür, daß die Vererbung der Faktoren M und N durch ein einziges Genpaar bedingt ist. Keines der beiden Gene dominiert derartig, daß das andere Gen in der Heterozygote phänotypisch unterdrückt würde. *Sollte sich die Vererbung von M und N nach dem monohybriden Schema als ausnahmslos gültig erweisen, so wäre damit für die gerichtliche Abstammungsuntersuchung ein wesentlicher Fortschritt erreicht* [my emphasis]. Es würde dann möglich sein, durch Untersuchung auf M-N*und* Blutgruppe rund jeden dritten zu Unrecht als Vater angegebenen Mann auszuschließen (gegen bisher jeden sechsten bis siebenten). (Schiff, 1930, Zusammenfassung)

In Morgan (1928), no exemplar of codominance is mentioned.

viz. the "purity of the germ-cells". In this part, I will mainly rely on the works of Mendel (1865), de Vries (1900a), Correns (1900), Bateson (1900, 1902), and Weldon (1902). In section 4.9.2, I will show how the role of anomalies can be explicated in my framework, and how the status of complete dominance is adequately represented. I will also show how this fits with Darden's concepts of model anomalies and monster anomalies, and how these issues provide an extra justification for partly basing my framework on the structuralist approach of Balzer et al. (1987).

Together, sections 4.9.1 and 4.9.2 will illustrate that *pragmatic laws*, such as the principle of complete dominance, may have exceptions and that they have limited domains of application. In other words, they are not universal and have limited stability.

4.9.1 The status of complete dominance in the early days of classical genetics

In Mendel's original paper, *Versuche über Pflanzen-Hybriden*, only crosses with complete dominance are mentioned but there is no clear indication that Mendel deemed complete dominance a universal phenomenon. Moreover, he stated that

> [...] Kölreuter and Gärtner, *the two authorities in this field* [...] concur in the opinion that, in external appearance, *hybrids either maintain a form intermediate between the parental strains* or they approach the type of one or the other, sometimes being barely distinguishable from them. (Mendel, 1865, 39, my emphasis)

Mendel not only assentingly mentioned the findings of Kölreuter and Gärtner, he also stated that his choice for dominant and recessive traits was methodological (cf. Mendel, 1865, 3–5). Nevertheless, in 1900 the principle of complete dominance was taken to be universal (or nearly universal) by Hugo de Vries:

> My experiments have led me to make the two following statements: [...] 1. *Of the two antagonistic characteristics, the hybrid carries only one*, and that in complete development. Thus in this respect the hybrid is indistinguishable from one of the two parents. *There are no transitional forms* [my emphasis]. 2. *In the formation of pollen and ovules the two antagonistic characteristics separate*, following for the most part simple laws of probability.

> These two statements, in their most essential points, were drawn up long ago by Mendel for a special case (peas). These formulations have been forgotten and their significance misunderstood. [...] As my experiments show, they possess generalized validity for true hybrids.
>
> The lack of transitional forms between *any* [my emphasis] two simple antagonistic characters in the hybrid is perhaps the best proof that such characters are well delimited units.[26] (de Vries, 1900a, 110)

Whereas de Vries did not really question the principle of complete dominance, Carl Correns did.[27] In the same year, 1900, he wrote:

> I can not understand why de Vries assumes that in *all* pairs of traits which differentiate two strains, one member must always

[26]It should be noted, however, that de Vries also mentions exceptions to the principle of complete dominance.

> Exceptions occur seldom; an example is presented by some sectional segregations. Thus *Veronica longifolia* (blue) × *V. longifolia alba* in my experiments not infrequently forms racemes whose flowers are white on one side and blue on the other. (de Vries, 1900a, 112)

de Vries merely mentions these anomalies, without proposing any explanation, and without considering them problematic.

[27]Correns and de Vries are often cited as independent rediscoverers of the laws of Mendel. (They are often cited so along with Tschermak, but there are good reasons not to grant Tschermak this honour – see Stern and Sherwood (1966, x–xii) and the text of Tschermak (1950).) Correns and de Vries each put very much effort in stressing their originality, both with respect to Mendel as to each other.

> This important treatise [viz. Mendel (1933, 1865)] is so seldom cited that I first learned of its existence after I had completed the majority of my experiments and had deduced from them the statements communicated in the text. (de Vries, 1900a, 110)

> When I discovered the regularity of the phenomena, and the explanation thereof [...] the same thing happened to me which now seems to be happening to de Vries: I thought that I had found *something new.* [...] *But then I convinced myself that the Abbot Gregor Mendel in Brünn, had, during the sixties, not only obtained the same result through extensive experiments with peas, which lasted for many years, as did de Vries and I, but had also given exactly the same explanation, as far as that was possible in 1866.* (Correns, 1900, 119–120, original emphasis)

For more information on questions relating to the allegedly independent rediscoveries of Mendel's works, see inter alia Meijer (1983, sections 1 and 2), Darden (1991, 42–46) and Orel (1996, 284–288).

> dominate. Even in peas, where some traits completely conform to this rule, other trait pairs are also known in which neither trait dominates [...]. (Correns, 1900, 121, original emphasis)

After mentioning examples of such failures of complete dominance, Correns provisionally restricted the domain of application of Mendel's principles to cases of complete dominance:

> The following holds only for pairs of traits which have a dominating and a recessive member; there is no reason to believe that it may not hold for other types of pairs of traits as well, but at present we know of no example. [In a footnote added later, he writes: "In the meantime I have found an example."] (Correns, 1900, 122)

In a postscript to his paper, Correns reemphasized these points:

> I must emphasize again: 1. that in many pairs of traits there is no dominating member [...], 2. that Mendel's Law of segregation cannot be applied universally [...]. (Correns, 1900, 132)

Later in 1900, William Bateson treated Mendel's work on a par with Galton's, in spite of the fact that the two seem incompatible. According to Bateson, Galton was the first to search systematically for and enunciate a "law of heredity" (i.e. a 'general expression capable of sufficiently wide application to be justly called a "law" of heredity') (Bateson, 1900, 3).[28] However, Galton's law has many exceptions.

> These large classes of exception – to go no further – indicate that, as we might in any case expect, the principle is not of universal application,[29] and will need various modifications if it is to be extended to more complex cases of inheritance of varietal characters. (Bateson, 1900, 4–5)

Fortunately, Bateson wrote, Galton's theory had recently been supplemented with Mendel's, which seems fairly general:

[28] In chapter 9 I will discuss Galton's theory in more detail and I will show how it was biased by his knowledge of statistical tools.

[29] Bateson makes a distinction between 'general' principles, and 'universal' ones. The latter are stronger than the former (see footnote 37).

> Professor de Vries has worked at the same problem in some dozen species belonging to several genera, using pairs of varieties characterized by a great number of characters: for instance, colour of flowers, stems, or fruits, hairiness, length of style, and so forth. *He states that in all these cases Mendel's law is followed.*[30] [...] When we consider, besides, that Tschermak and Correns announce definite confirmation in the case of *Pisum*, and de Vries adds the evidence of his long series of observations on other species and orders, there can be no doubt that Mendel's law is a substantial reality; though whether some of the cases that depart most widely from it can be brought within the terms of the same principle or not, can only be decided by further experiments. (Bateson, 1900, 7–8, my emphasis)

Bateson concluded in a pluralist fashion:

> That different species should follow different laws, and that the same law should not apply to all characters alike, is exactly what we have every right to expect. It will also be remembered that the principle is only declared to apply to discontinuous characters. (Bateson, 1900, 8)

Despite Bateson's pluralism, W.F.R. Weldon wrote a harsh article in the newly founded journal *Biometrika* in which he set out to undercut "Mendel's laws of alternative inheritance in peas" and to argue for the supremacy of Galton's theory (Weldon, 1902). This would provoke a heated and longstanding discussion among proponents and opponents of Mendelism.[31]

Weldon's arguments against Mendelism are as follows. Firstly, he attributes the "Law of Dominance" (his capitals) a central place in Mendel's theory, along with the "Law of Segregation" (Weldon, 1902, 229).[32] Secondly, using Pearson's χ^2-test, he shows that Mendel's data fit his theory too nicely

[30] Note that Bateson writes 'Mendel's law' (singular).

[31] According to Sloan, Weldon's paper cannot be characterized as the broadside against Mendel's paper that Bateson (1902) depicted in his *Defence*. (Sloan, 2000, 1074) It is nevertheless true that Weldon's introduction is fairly explicit: whereas the theory of Galton and Pearson provides an account for blended inheritance "which seems likely to prove generally applicable," our knowledge of (particulate and) alternative inheritance, which is deemed the intended domain of application of Mendel's theory, is still rudimentary. (Weldon, 1902, 228)

[32] Weldon defines the Law of Dominance as follows:

> *If peas of two races be crossed, the hybrid offspring will exhibit only the dominant characters of the parents; and it will exhibit these without (or almost without) alteration, the recessive characters being altogether absent, or*

(Weldon, 1902, 232–235) – in a way that foreshadows Fisher's famous paper "Has Mendel's work been rediscovered?" (Fisher, 1936).[33] Thirdly, he endeavours to show that the Law of Dominance is plagued by exceptions, *even in its paradigmatic cases*, and that Mendel's failure is due to his neglect of the influence of ancestry (Weldon, 1902, 236–252).

Weldon's paper occasioned Bateson's lengthy monograph, *Mendel's Principles of Heredity. A Defence* (Bateson, 1902). For Bateson the stakes were high.[34] His monograph is interesting for several reasons. Firstly, it contains a reprint (with additions) of his 1900 paper. Secondly, what is more interesting, it also contains a reprint (with slight modifications) of the first English translation of Mendel's *Versuche über Pflanzen-Hybriden*, which was published by the Royal Horticultural Society in 1901. It is titled "Mendel's Experiments in Plant Hybridisation". And it also contains the first English translation of Mendel's second paper, *Über einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde* (cf. Mendel, 1869). These translations all contained errors and mistakes which fundamentally changed the meaning

---

> *present in so slight a degree that they escape notice.* (Weldon, 1902, 229, original emphasis)

His definition of the "Law of Segregation" also deserves our attention (albeit as a side issue). Just like the Law of Dominance, it is phrased in a Pearsonian way, i.e. without any reference to unobservable entities.

> *If the hybrids of the first generation produced by crossing two races of peas which differ in certain characters, be allowed to fertilize themselves, all possible combinations of the ancestral race-characters will appear in the second generation with equal frequency, and these combinations will obey the Law of Dominance, so that characters intermediate between those of the ancestral races will not occur.* (Weldon, 1902, 229, original emphasis)

In the rest of his paper, however, Weldon at times uses non-Pearsonian language.

[33] It is interesting to note that in the bibliography of Fisher (1936), no reference is made to Weldon (1902). Between 1902 and 1936, the theory of χ^2-tests had changed, however. In 1922, Fisher had shown that Karl Pearson used χ^2 distributions with too high numbers of degrees of freedom (Yates and Mather, 1963, 101).

[34]

> In the world of knowledge we are accustomed to look for some strenuous effort to understand a new truth even in those who are indisposed to believe. It was therefore with a regret approaching to indignation that I read Professor Weldon's criticism. Were such a piece from the hand of a junior it might safely be neglected; but coming from Professor Weldon *there was the danger – almost the certainty – that the small band of younger men who are thinking of research in this field would take it they had learnt the gist of Mendel, would imagine his teaching exposed by Professor Weldon, and look elsewhere for lines of work.* (Bateson, 1902, vi, my emphasis)

of Mendel's sentences (Stern and Sherwood, 1966, vi-vii). Thirdly, what is most interesting in the context of this section, Bateson's monograph contains a lengthy attempt to defend Mendel against Weldon.

Bateson's defence shows one way in which supposed anomalies or counterexamples may be cleared away. He attributes a central place to the "purity of germ-cells" in Mendel's theory, at the expense of the alleged Law of Dominance.[35] The "purity of the germ-cells" is "a proposition at variance with all the laws of ancestral heredity, however formulated"[36] (Bateson, 1902, 114), and it contradicts Weldon's suggestion that Mendel's failure is due to his neglect of the influence of ancestry.

> In those cases to which it applies strictly, this principle declares that the cross-breeding of parents need not diminish the purity of their germ-cells or consequently the purity of their off-spring. When in such cases individuals bearing opposite characters, A and B, are crossed, the germ-cells of the resulting cross-bred, AB, are each to be bearers either of character A or of character B, but not both. (Bateson, 1902, 114)

He even leaves open the possibility that the principle of the purity of the germ-cells also applies to cases of *blended inheritance*. (In the case of blended inheritance, there are no pairs of strictly distinguishable, opposing or *alternative* phenotypic traits. What used to be called blended inheritance is now attributable to incomplete dominance, or to multiple factors.) Thereby, he denies the Law of Dominance a central place in Mendel's theory.

> Mendel's own cases were almost all alternative; also the fact of dominance is very dazzling at first. But that was two years ago, and when one begins to see clearly again, it does not look so certain that the real essence of Mendel's discovery, the purity of the germ-cells in respect of certain characters, may not apply also to some phenomena of blended inheritance. The analysis of this possibility would take us to too great length, but I commend to those who are more familiar with statistical method, the consideration of this question: whether dominance being absent, indefinite, or suppressed, the phenomena of heritages completely blended in the zygote, may not be produced by gametes presenting Mendelian purity of characters. (Bateson, 1902, 115)

[35] Curiously, Bateson (1902) nowhere mentions (let alone addresses) Weldon's statistical arguments on goodness of fit.

[36] For the law of ancestral heredity, see section 9.5.1.

Mendel's choice for antagonistic traits was methodological (cf. supra), and it made his discovery possible. But as such, complete dominance "is still an accident of particular cases." (Bateson, 1902, 117)

> The whole question whether one or other character of the antagonistic pair is dominant though of great importance *is logically a subordinate one*. It depends on the specific nature of the varieties and individuals used, sometimes probably on the influence of external conditions and on other factors we cannot now discuss. *There is as yet no universal law here perceived or declared.*[37] (Bateson, 1902, 117–118, my emphasis)

By denying the principle of complete dominance a central place in his framework, Bateson effected that exceptions to this principle should not count heavily against the theory of Mendelian genetics.

4.9.2 Anomalies and theory-elements

More and more exceptions to the principle of complete dominance were reported, both by Mendelians (e.g. Correns) and biometricians (e.g. Weldon). The status of these *anomalies* in my framework is straightforward. They give rise to data for which there is no $\mathcal{B} = \langle G, P \rangle$ such that (i) $\mathcal{B}$ is distribution-identical to some member of $\mathfrak{B}^{\alpha}$ and (ii) P is consistent with the data (cf. section 4.8).

Darden (1991, 199–201) distinguishes between two kinds of anomalies.

> A monster anomaly does not require a change in the set of patterns for normal, well-functioning cases. In contrast, a model anomaly does requires [sic] such a change, either the alteration of a typical pattern or the addition of one or more new patterns to the set. (Darden, 1991, 199)

> Monster anomalies are resolved by showing what went wrong in the normal process. The theory is saved. [...] Model anomalies,

[37] Bateson distinguishes between *general* truths and *universal* ones (Bateson, 1902, 119), and he is surprised that Weldon makes no such distinction (Bateson, 1902, 106).

> That the dominance of yellow cotyledon-colour over green, and the dominance of the smooth form over the wrinkled, is a *general* truth for *Pisum sativum* appears at once; that it is a universal truth I cannot believe any competent naturalist would imagine, still less assert. Mendel certainly never did. (Bateson, 1902, 119, original emphasis)

It is reasonable to interpret 'universal' in the sense of chapter 1 here. By contrast, Bateson's use of 'general' should not be equated with that in the same chapter.

> however, require a change in the claims about what is normal (or general) in hereditary processes. (Darden, 1991, 203)

Darden nowhere discusses the status of incomplete dominance. However, it can be deduced from her criteria that it should be considered a 'small' model anomaly. It should be considered a model anomaly since it gives rise to a different exemplar in Morgan (1928), and hence to a different explanatory pattern. It should be considered small since Morgan attaches little importance to it. He does not explicitly refer to it in his abstract statement of the theory of the gene (Morgan, 1928, 25). As we saw in section 4.9.1, Bateson (1902, 117–118) granted complete dominance but a logically subordinate status. By 1915, Morgan and his coworkers attached even less importance to it (cf. Darden, 1991, 72):

> Whether a character is completely dominant or not appears to be a matter of no special significance. In fact, the failure of many characters to show complete dominance raises a doubt as to whether there is such a condition as complete dominance. (Morgan et al., 1915, 31)

In my framework, it is easily seen that incomplete dominance is a model anomaly. After all, the crosses $\mathcal{B}^1 = \langle G^3, P^1 \rangle, \mathcal{B}^2 = \langle G^3, P^2 \rangle, \ldots$ with $P^1, P^2, \ldots \in \mathbb{P}^3$ all belong to $\mathfrak{B}^3$, which in turn can be generalized to the generic credal net (i.e. the general argument pattern) $\mathfrak{B}^\beta$ which is substantially different from $\mathfrak{B}^\alpha$: $\mathfrak{B}^\beta$ is not value-isomorphic to $\mathfrak{B}^\alpha$, let alone distribution-identical.

In the structuralist approach, scientific theories are represented as sets of interrelated theory-elements. These elements and their interrelations are depicted in a theory-net – a tree-shaped graph in which the nodes correspond to theory-elements and the edges to their interrelations (in the structuralist approach, several such relations are discerned: specialization, refinement, etc.). Each theory-element, or the set of laws in each theory-element, has a limited domain of intended application. In Balzer and Lorenzano (2000, 256), models with complete dominance ($M_{compdom}$) and models with incomplete dominance ($M_{incompdom}$) are relegated to different theory-elements, each with a different domain of intended application. In my opinion, this account provides a nice way to explicate the way in which classical geneticists dealt with exceptions. Moreover, $M_{compdom}$ and $M_{incompdom}$ each are obtained by 'specializing' another theory-element, M_{one} (which encompasses all monohybrid crosses in which Mendel's first law holds). These specializations are removed relatively far from the root node in the theory-net of classical genetics. This indicates that, indeed, the issue of dominance is logically subordinate.

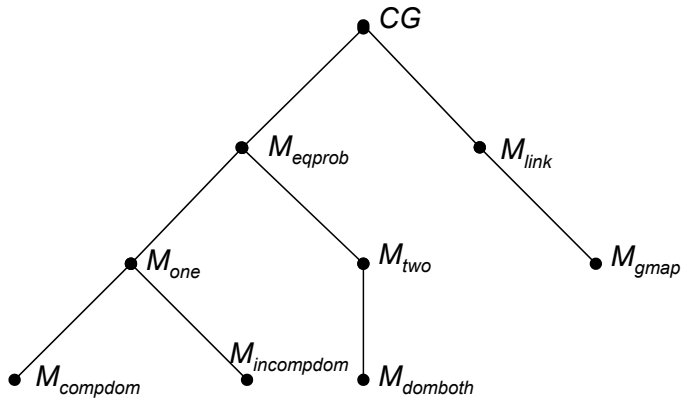

Figure 4.2: *The theory-net of classical genetics in Balzer and Lorenzano (2000, 261)*

As I have shown above (section 4.5), there are strong intuitive links between the concepts 'model,' 'potential model,' and 'partial potential model' on the one hand, and my framework on the other hand. I will not elaborate the details of these links here. It suffices to state that the generic credal nets $\mathfrak{B}^{\alpha}$ and $\mathfrak{B}^{\beta}$ (and those to follow) not only provide an explication for the Kitcher-Darden explanatory patterns, but also an indirect representation for (an adapted version of) the structuralist's theory-elements. Each generic credal net incorporates a set of implicit *ceteris paribus* conditions[38] and determines the structural and probabilistic properties of the causal models that may account for a limited domain of intended applications (where $\mathfrak{B}^{\beta}$'s domain of intended applications consists of phenomena that are anomalous with respect to $\mathfrak{B}^{\alpha}$). The set of *ceteris paribus* conditions determines this domain of intended applications. (Hence the domain of intended applications is independently specified. The range of possible exceptions is not explicitly incorporated in the laws of the theory-elements in question. For the distinction between 'independent specification' and 'exception-incorporation', see Woodward (2003b, 273ff).)

[38]I mean a more or less well delineated set of *ceteris paribus* conditions, based on existing scientific knowledge. So it should not be feared that these *ceteris paribus* conditions make the principles of classical genetics trivially true (cf. Pietroski and Rey, 1995, Earman et al., 2002, Woodward, 2002a, and Mitchell, 2002). For example, both in $\mathfrak{B}^{\alpha}$ and in $\mathfrak{B}^{\beta}$ it is assumed that mutation does not occur. This assumption is part of the set of implicit *ceteris paribus* conditions incorporated by these generic credal nets. The assumption that no angel interferes to change the relation between genotypes and phenotypes is not a part of this set.

4.10 Multihybrid crosses and independent assortment

After presenting two monohybrid crosses with complete dominance, and one with incomplete dominance, Morgan turns to dihybrid crosses. His most important mainspring is to better come to grips with the ontological status of the gene. With respect to the crosses in section 4.8, he writes:

> So far the results tell us no more than that something derived from one parent separates, in the germ-cells of the hybrid, from something brought in by the other parent. The results might be interpreted, on this evidence alone, to mean that red-flowered and white-flowered plants behave as wholes or entities in inheritance. (Morgan, 1928, 7)

To prove that plants do not behave as wholes in inheritance, he cites one of Mendel's dihybrid crosses.

> Mendel crossed peas whose seeds were yellow and round with peas whose seeds were green and wrinkled. Other crosses had shown that yellow and green constitute a pair of contrasted characters giving a 3 to 1 ratio in the second generation, and that round and wrinkled constitute another pair.
>
> The offspring [F_1] were yellow and round [...]. When selfed, they produced four kinds of individuals [F_2], yellow round, yellow wrinkled, green round, and green wrinkled in the ratio of 9:3:3:1.
>
> Mendel pointed out that the numerical results found here can be explained, if the separation of the elements for yellow and for green is independent of that for round and wrinkled. This would give four kinds of germ-cells in the hybrid, yellow round, yellow wrinkled, green round, and green wrinkled [...].
>
> If the fertilization of the four kinds of ovules by the four kinds of pollen grains is at random, there will be sixteen combinations possible. Remembering that yellow dominates green, and that round dominates wrinkled, these sixteen combinations will fall into four classes, that are in the ratios of 9:3:3:1. (Morgan, 1928, 7–9)

To explain the phenotypic distributions in the F_1 and the F_2 generation, Morgan relies on the following principles. He takes Mendel's first law (the *law*

of segregation) and adds the assumption that segregation (separation) for one pair of elements is independent of segregation for another pair. This he calls Mendel's second law, or the *law of independent assortment* (Morgan, 1928, 10). He assumes that '*complete dominance*' holds for the pairs yellow/green and round/wrinkled, as it did in the monohybrid crosses. He adopts some revised version of the *principle of combination* (cf. figure 8 in Morgan, 1928, 9). Finally, he assumes *random fertilization*.

Clearly, dihybrid crosses cannot be described by any set of variables that is value-isomorphic to V^α or V^β. Each phenotypic variable has four possible values: $[PT_i] = \{yellowround, yellowwrinkled, greenround, greenwrinkled\}$. Morgan refers these phenotypes to more complex genotypes than in the previous sections: $[GT_i] = \{GGWW, GGWw, \ldots, ggww\}$. Likewise, the gametes also have a more complex make-up: $[GC_i] = \{GW, Gw, gW, gw\}$.[39] Let V^4 be the set of these variables. (Note that the members of $[GT_i], [GC_i]$ and $[PT_i]$ are not labeled fully rigorously. In accordance with definitions 3.20 and 3.21, they should be tuples; e.g. $\langle yellow, round\rangle, \langle yellow, wrinkled\rangle, \ldots$ In the interest of readability, however, I have allowed for this relaxed notation.)

As Morgan indicates, these dihybrid crosses are strongly related to other, monohybrid crosses performed by Mendel. At the level of random variables, there is a strong relation between V^4 and V^α. More specifically, let $V^{1'}$ and $V^{1''}$ be two sets of variables such that there are credal nets $\mathfrak{B}^{1'} = \langle\langle V^{1'}, E\rangle, \mathbb{P}^{1'}(V^{1'})\rangle$ and $\mathfrak{B}^{1''} = \langle\langle V^{1''}, E\rangle, \mathbb{P}^{1''}(V^{1''})\rangle$ that are distribution-identical to $\mathfrak{B}^\alpha$ and such that the members of $\mathfrak{B}^{1'}$ and $\mathfrak{B}^{1''}$ can be used to describe or explain crosses on albumen colour and on seed shape in *Pisum*, respectively. (It follows by definitions 3.16 and 3.7 that $V^{1'}$ and $V^{1''}$ are value-isomorphic to V^α.) I will show that V^4 can be interpreted as a set of compound variables that can be defined in terms of members of $V^{1'}$ and $V^{1''}$.

Let $[PT'_1] = \{yellow, green\}$ be the space of $PT'_1 \in V^{1'}$ and let $[PT''_1] = \{round, wrinkled\}$ be the space of $PT''_1 \in V^{1''}$. Clearly, $PT_1 = compound\ \langle PT'_1, PT''_1\rangle$ (definition 3.20). Analogously, all other variables in V^4 can be defined as compounds of suitable variables in $V^{1'}$ and $V^{1''}$, respectively. In short, $V^4 = compound\langle V^{1'}, V^{1''}\rangle$ (definition 3.21).

Let us turn now to the credal set $\mathbb{P}^4$ over V^4 that is generated by Morgan's explanatory principles. Any $P \in \mathbb{P}^4$ is such that it satisfies tables 7 and 8, plus some revised version of the principle of combination. In table 7, the relations between the genotype and the phenotype are listed. It shows that yellow dominates green, and that round dominates wrinkled. Table 8 lists the relations between the genotype and the gametic make-up. It incorporates both Mendel's first law (the law of segregation) and Mendel's second law (the

[39] For the extensions of $[GT_i]$ and $[GC_i]$, see Morgan (1928, 9), figure 8.

table 7

	PT_i			
GT_i	*ye rnd*	*ye wrd*	*gr rnd*	*gr wrd*
$GGWW$	1.00	0.00	0.00	0.00
$GGWw$	1.00	0.00	0.00	0.00
$GGww$	0.00	1.00	0.00	0.00
$GgWW$	1.00	0.00	0.00	0.00
$GgWw$	1.00	0.00	0.00	0.00
$Ggww$	0.00	1.00	0.00	0.00
$ggWW$	0.00	0.00	1.00	0.00
$ggWw$	0.00	0.00	1.00	0.00
$ggww$	0.00	0.00	0.00	1.00

table 8

	GC_i			
GT_i	GW	Gw	gW	gw
$GGWW$	1.00	0.00	0.00	0.00
$GGWw$	0.50	0.50	0.00	0.00
$GGww$	0.00	1.00	0.00	0.00
$GgWW$	0.50	0.00	0.50	0.00
$GgWw$	0.25	0.25	0.25	0.25
$Ggww$	0.00	0.50	0.00	0.50
$ggWW$	0.00	0.00	1.00	0.00
$ggWw$	0.00	0.00	0.50	0.50
$ggww$	0.00	0.00	0.00	1.00

Conditional probability tables 7–8: together with the principle of combination, these provide constraints for $P \in \mathbb{P}^4$

law of independent assortment). I leave the specification of the conditional probability table explicating the principle of combination for this dihybrid cross to the reader.

Instead of specifying the tables 7 and 8 directly for $\mathbb{P}^4$, I could have tried to determine $\mathbb{P}^4$ on the basis of $\mathbb{P}^{1'}$ and $\mathbb{P}^{1''}$. This can be done by specifying, for each $Z \in V^4$, for each $z_k \in [Z]$ and for each $pa_l \in [PA(Z)]$, the probability $P(z_k \mid pa_l)$ in terms of $P'(x_i \mid pa_m)$ and $P''(y_j \mid pa_n)$, where $X \in V^{1'}$, $Y \in V^{1''}$, $Z = compound\langle X, Y\rangle$, $z_k = \langle x_i, y_j\rangle$, $pa_m \in [PA(X)]$, $pa_n \in [PA(Y)]$, and $pa_l = \langle pa_m, pa_n\rangle$.

I will do this only partly here. I will concentrate on the relation between the variables GT_i and GC_i (for each i), described in table 8. The reason is that here the quintessence of independent assortment is revealed. Let $b : V^4 \to V^{1'}$ and $b^* : V^4 \to V^{1''}$ be bijections as in definition 3.21. For each $GT_i, GC_i \in V^4$, $GT_i = compound\langle b(GT_i), b^*(GT_i)\rangle$ and $GC_i = compound\langle b(GC_i), b^*(GC_i)\rangle$. For each i, $PA(GC_i) = \{GT_i\}, PA(b(GC_i)) = \{b(GT_i)\}$, and $PA(b^*(GC_i)) = \{b^*(GT_i)\}$.

Recall that the credal nets $\mathfrak{B}^{1'}$ and $\mathfrak{B}^{1''}$ are distribution-identical to $\mathfrak{B}^{\alpha}$. From table 8 it is easily seen that the relation between GT_i and GC_i is such that, for each $z_k \in [GC_i]$ and for each $pa_l \in [GT_i]$,

$$P(z_k \mid pa_l) = P'(x_i \mid pa_m) \times P''(y_j \mid pa_n) \tag{4.1}$$

For example,

$$P(GW \mid GgWW) = 0.50 = P'(G \mid Gg) \times P''(W \mid WW) = 0.50 \times 1.00$$

What conditions need be in place for equation (4.1) to hold? Note that each $z_k = \langle x_i, y_j\rangle$ can be regarded as a conjunction $x_i \wedge y_j$, and each $pa_l =$

$\langle pa_m, pa_n\rangle$ as a conjunction $pa_m \wedge pa_n$. Hence let us redescribe $P(z_k \mid pa_l)$ as $P(x_i \wedge y_j \mid pa_m \wedge pa_n)$. (I will not make a notational distinction between $P(V^4)$ and $P(V^{1'} \cup V^{1''})$.) $P(x_i \wedge y_j \mid pa_m \wedge pa_n) = P(x_i \mid pa_m) \times P(y_j \mid pa_n)$ for all $x_i, y_j, pa_m,$ and pa_n if the following conditions hold:

$(b(GC_i) \amalg b^*(GC_i) \mid b(GT_i), b^*(GT_i))$, and

$(b(GC_i) \amalg b^*(GT_i) \mid b(GT_i))$, and

$(b^*(GC_i) \amalg b(GT_i) \mid b^*(GT_i))$.

This can be seen as follows (I substitute $b(GC_i)$ with A, $b^*(GC_i)$ with B, $b(GT_i)$ with C, and $b^*(GT_i)$ with D):

$$\begin{aligned}
P(a \wedge b \mid c \wedge d) &= P(a \mid c \wedge d) \times P(b \mid c \wedge d) \\
&\quad \text{for all } a \in [A], b \in [B], c \in [C], d \in [D] \\
&\quad \text{if } (A \amalg B \mid C, D), \\
P(a \mid c \wedge d) &= P(a \mid c) \\
&\quad \text{for all } a \in [A], c \in [C], d \in [D] \\
&\quad \text{if } (A \amalg D \mid C), \text{ and} \\
P(b \mid c \wedge d) &= P(b \mid d) \\
&\quad \text{for all } b \in [B], c \in [C], d \in [D] \\
&\quad \text{if } (B \amalg C \mid D). \text{ Hence,} \\
P(a \wedge b \mid c \wedge d) &= P(a \mid c) \times P(b \mid d) \\
&\quad \text{for all } a \in [A], b \in [B], c \in [C], d \in [D] \\
&\quad \text{if } (A \amalg B \mid C, D), (A \amalg D \mid C), \\
&\quad \text{and } (B \amalg C \mid D).
\end{aligned}$$

This shows why equation (4.1) holds when

$(b(GC_i) \amalg b^*(GC_i) \mid b(GT_i), b^*(GT_i))$, and

$(b(GC_i) \amalg b^*(GT_i) \mid b(GT_i))$, and

$(b^*(GC_i) \amalg b(GT_i) \mid b^*(GT_i))$.

These conditions are satisfied in the case of *independent assortment*. Then "the separation of the elements for yellow and for green is independent of that for round and wrinkled." (Morgan, 1928, 8)

As is to be expected, all crosses mentioned in Morgan's citation can be described or explained by means of a causal net in $\mathfrak{B}^4 = \langle\langle V^4, E\rangle, \mathbb{P}^4\rangle$. Crosses similar to this exemplar can be explained in a similar way. Hence, let

$\mathfrak{B}^\gamma = \langle G^\gamma, \mathbb{P}^\gamma \rangle$ (where $G^\gamma = \langle V^\gamma, E \rangle$ and where $\mathfrak{B}^\gamma$ is distribution-identical to $\mathfrak{B}^4$) be the generic representation of dihybrid crosses with complete dominance and independent assortment (and without gene interaction, ...).

4.11 Linkage and crossing-over

From the crosses discussed in section 4.10, Morgan concludes the following:

> [...] it can no longer be assumed that the whole parental germ-materials are separated in the hybrid; for yellow and round that went in together have, in some cases, come out separated. Similarly for green and wrinkled. (Morgan, 1928, 9–10)

Other crosses performed by Mendel, involving three or even four pairs of characters, had corroborated this conclusion.[40] Is independent assortment a universally applicable law, then? No.

> It might, then, have seemed justifiable to extend this conclusion to as many pairs of characters as enter any particular cross. This would mean that there are as many independent pairs of elements

[40] In Morgan (1928), multihybrid crosses with complete dominance and independent assortment are mentioned but left unspecified. I refer the reader to Mendel (1865, 17–23) for Mendel's own description of such crosses, and to the earlier works of Morgan:

> When three independent factor-pairs are present the numerical expectation can be directly derived from the 9:3:3:1 ratio in the same way that the latter was derived from the 3:1 ratio. Thus:
>
> | | | | | |
> |---|---|---|---|---|
> | 3 | | 1 | | One pair of characters |
> | 9 | 3 | 3 | 1 | Two pairs of characters |
> | 27:9 | 9:3 | 9:3 | 3:1 | Three pairs of characters |
>
> Each F_2 class of the two-factor case (9:3:3:1) will contain a three-to-one ratio for the third factor-pair. Thus in the 9 class there will be 3 dominants of the third factor to one recessive (27:9). So for each 3 class: each contains the third factor in the ration of 3:1. So also for the 1 class. The total result therefore is:
>
> $$27 : 9 : 9 : 9 : 3 : 3 : 3 : 1$$
>
> (Morgan, 1919, 71–72)

Crosses with three or more independent pairs of characters can be easily incorporated in my framework (by means of recursively specified compound variables and sets of variables), but I will not do it here.

> in the germinal material as there are possible characters. Subsequent work has shown, however, that Mendel's second law of independent assortment has a more *restricted application*, since many pairs of elements do not assort freely, but certain elements that enter together show a tendency to remain together in succeeding generations. This is called linkage. (Morgan, 1928, 10, my emphasis)

More precisely,

> By linkage we mean that when certain characters enter a cross together, they tend to remain together in later generations, or, stated in a negative way, certain pairs of characters do not assort at random. (Morgan, 1928, 10)

Like independent assortment, linkage is relevant for the ontological status of the gene:

> It would seem, then, so far as linkage holds, that there are limits to the subdivision of the germinal material. (Morgan, 1928, 10–11)

Genes that are linked belong to the same linkage group. For example, in *Drosophila melanogaster*, four linkage groups were discovered (Morgan, 1928, 11–12). It was soon discovered that genes in the same linkage group are not always completely linked. There may be some interchange between linkage groups.[41]

> This interchange is called crossing-over, which means that, between two corresponding linked series, there may take place an orderly interchange involving great numbers of genes. (Morgan, 1928, 14)

Linkage and crossing-over are illustrated by means of crosses performed by Bateson and Punnett.

> For instance, when a sweet pea having purple flower-color and long pollen grains is crossed to one with red flowers and round pollen grains, the two types that go in together come out together more frequently than expected for independent assortment of purple-red and round-long [...]. (Morgan, 1928, 10)

[41] The phenomena to be discussed in the present section were not always explained in terms of linkage and crossing-over (see footnote 47).

Flower colour and pollen shape in sweet peas (*Lathyrus odoratus*) were known to show normal Mendelian segregation and to satisfy dominance/ recessiveness (with purple being dominant to red, and long being dominant to round) (Darden, 1991, 122–123).[42] In other words, monohybrid crosses on flower colour and on pollen shape in *Lathyrus odoratus* can be described by means of credal nets $\mathfrak{B}^{1^*} = \langle\langle V^{1^*}, E\rangle, \mathbb{P}^{1^*}\rangle$ and $\mathfrak{B}^{1^{**}} = \langle\langle V^{1^{**}}, E\rangle, \mathbb{P}^{1^{**}}\rangle$ that are distribution-identical to $\mathfrak{B}^{\alpha}$. By contrast, there is no credal net $\mathfrak{B}$ such that $\mathfrak{B}$ is distribution-identical to $\mathfrak{B}^{\gamma}$ (i.e. the generic representation of dihybrid crosses with independent assortment) and such that the members of $\mathfrak{B}$ can be used to represent or explain dihybrid crosses on flower colour *and* pollen shape in sweet peas. The reason is that the independence conditions listed on page 89 are not satisfied. It follows that we should seek a new credal net $\mathfrak{B}^5 = \langle\langle V^5, E\rangle, \mathbb{P}^5\rangle$, which is not distribution-identical to $\mathfrak{B}^{\gamma}$, such that $V^5 = \mathit{compound}\langle V^{1^*}, V^{1^{**}}\rangle$ and such that $\mathbb{P}^5$ accounts for the failure of independent assortment.

Let us first have a look at the data reported by Bateson and Punnett (Morgan, 1928, 11, figure 9).[43] Cross 1: Sweet peas with purple flowers and long pollen grains were crossed with sweet peas with red flowers and round pollen grains. (It may be assumed that these parental plants were true-breeding.) The resulting hybrids (F_1) all had long pollen grains and purple flowers, as could be expected from the dominance of long to round, and of purple to red. Cross 2: Self-fertilization of the F_1 generation produced F_2 individuals in the following proportions (the absolute frequencies are Morgan's, I have added the percentages):

long, purple	round, purple	long, red	round, red
583	26	24	170
73%	3%	3%	21%

The results of cross 2 can be explained as follows. We know that the F_1 plants are hybrid, so $P(GT_1 = GgWw) = P(GT_2 = GgWw) = 1.00$, where G denotes the purple-producing allele, g the red-producing allele, W the long-producing allele and w the round-producing allele.[44] We may assume that the dominance/recessiveness relations between purple and red, and between long and round still hold in the dihybrid case, so that we may rely on some variant

[42] So purple flower colour is not an intermediate between red and white, as in four-o'clock. This is not a case of incomplete dominance.

[43] The text in Morgan (1928, 11, figure 9) gives the impression that the cross concerned purple and *white* flowers, instead of purple and *red* ones. This conflicts with Morgan's main text and with Darden (1991, 122).

[44] Morgan (1928, 11) does not use letters to denote the alleles in these crosses. He uses instead pictorial elements (coloured flowers and little circles and ovals for the shape of the pollen grains).

of table 7. We may also assume that the principle of combination holds. Then all we need to do is to find the appropriate conditional probability table for the relation between GT_i and GC_i. I will show that the phenotypic distribution for the F_2 individuals can be explained if

$$\begin{aligned} P(GC_i = GW \mid GT_i = GgWw) &= 0.46 \\ P(GC_i = Gw \mid GT_i = GgWw) &= 0.04 \\ P(GC_i = gW \mid GT_i = GgWw) &= 0.04 \\ P(GC_i = gw \mid GT_i = GgWw) &= 0.46 \end{aligned}$$

For example, by the principle of dominance F_2 individuals can only have round pollen and red flowers if their genotype is $ggww$. By the principle of composition, this genotype can only result when the make-up of both fertilizing gametes is gw. The following calculations retrodict that 21% of the F_2 individuals will have round pollen and red flowers. Firstly, we know that $P(GT_1 = GgWw) = P(GT_2 = GgWw) = 1.00$. By the conditional probabilities just given, we can compute that $P(GC_1 = gw) = 0.46$, and likewise that $P(GC_2 = gw) = 0.46$. Since $(GC_1 \amalg GC_2)$ (see figure 4.1), $P(GC_1 = gw \wedge GC_2 = gw) = 0.46 \times 0.46 = 0.21$. So by the principle of combination and by some analogue of table 7, $P(GT_3 = ggww) = 0.21$ and $P(PT_3 = roundred) = 0.21$. Like calculations allow to retrodict the probabilities of the other phenotypes in F_2.[45]

Can we specify the extension of $\mathbb{P}^5$ by demanding that for all $P \in \mathbb{P}^5$ the above conditional probabilities hold? No. These conditional probabilities are cross-dependent. Assume that, in cross 1, $GGww$ individuals had been crossed with $ggWW$ individuals (instead of $GGWW \times ggww$).[46] The resulting F_1 hybrids would all be $GgWw$, as in Morgan's example. But selfing of these hybrids would give F_2 plants in the following proportions:

long, purple	round, purple	long, red	round, red
50,16%	24,84%	24.84%	0.16%

The reason is that here G and w (and thus also g and W), tend to remain together, whereas in the original cross G and W (and thus also g and w) tended to remain together (cf. the conditional probabilities below). In modern terminology, it's the genes that are linked, not their respective

[45]The calculations for the other phenotypes are somewhat more elaborate, given the multiple realizability of dominant phenotypic traits.

[46]Morgan did not discuss this particular cross, but an analogous way of reasoning can be found in Morgan (1928, 16–17).

alleles.[47]

$$\begin{aligned} P(GC_i = GW \mid GT_i = GgWw) &= 0.04 \\ P(GC_i = Gw \mid GT_i = GgWw) &= 0.46 \\ P(GC_i = gW \mid GT_i = GgWw) &= 0.46 \\ P(GC_i = gw \mid GT_i = GgWw) &= 0.04 \end{aligned}$$

How, then, should we characterize $\mathbb{P}^5$? As we have seen, linkage and crossing-over influence the probabilistic relations between GT_i and GC_i when $GT_i = GgWw$ (i.e., when doubly heterozygous plants are involved). Both sets of conditional probabilities listed above can be summarized as follows:

$$\begin{aligned} P(GC_i = GW \mid GT_i = GgWw) &= \frac{a}{2} \\ P(GC_i = Gw \mid GT_i = GgWw) &= \frac{1-a}{2} \\ P(GC_i = gW \mid GT_i = GgWw) &= \frac{1-a}{2} \\ P(GC_i = gw \mid GT_i = GgWw) &= \frac{a}{2}, \end{aligned}$$

where $a \in \{0.08, 0.92\}$ and where $\min\{0.08, 0.92\} = 0.08$ is the *frequency of crossing-over* for dihybrid crosses on flower colour and pollen shape in *Lathyrus odoratus*.

$GgWw$ is the only possible value of GT_i for which independent assortment on the one hand, and linkage and crossing-over on the other hand, give rise to probabilistically different results. Firstly, suppose that $GT_i = GGWW$ or any other doubly homozygous value. Then (in the absence of mutation) only one kind of gametes can be produced: $GC_i = GW$. Secondly, suppose that $GT_i = GGWw$ or any other singly homozygous value. Then one chromosome will carry G and W. The other chromosome will carry G and w. Part of the gametes (say, $x\%$) will not be the result of crossing-over. Half

[47] Between 1905 and 1911, Bateson tried to explain linkage phenomena in terms of the concepts of coupling and repulsion, which were strongly related to his presence and absence theory. According to the *presence and absence theory*, a dominant trait results from the presence of a dominant factor, while a recessive trait results from its absence. For example, purple flowers result from the presence of G, while red flowers result from g, which denotes the absence of G. Accordingly, linkage of dominant traits (such as purple and long, see the original cross 1) should then be interpreted in terms of a *coupling* of the dominant factors G and W. If, on the other hand, a dominant trait is linked with a recessive trait (such as purple and round, see my simulated cross), this should be interpreted in terms of a *repulsion* of the dominant factors G and W. (Darden, 1991, 121–123)

table 9

	PT_i			
GT_i	*pu lng*	*pu rnd*	*red lng*	*red rnd*
$GGWW$	1.00	0.00	0.00	0.00
$GGWw$	1.00	0.00	0.00	0.00
$GGww$	0.00	1.00	0.00	0.00
$GgWW$	1.00	0.00	0.00	0.00
$GgWw$	1.00	0.00	0.00	0.00
$Ggww$	0.00	1.00	0.00	0.00
$ggWW$	0.00	0.00	1.00	0.00
$ggWw$	0.00	0.00	1.00	0.00
$ggww$	0.00	0.00	0.00	1.00

table 10

	GC_i			
GT_i	GW	Gw	gW	gw
$GGWW$	1.00	0.00	0.00	0.00
$GGWw$	0.50	0.50	0.00	0.00
$GGww$	0.00	1.00	0.00	0.00
$GgWW$	0.50	0.00	0.50	0.00
$GgWw$	$\frac{a}{2}$	$\frac{1-a}{2}$	$\frac{1-a}{2}$	$\frac{a}{2}$
	$a \in \{0.08, 0.92\}$			
$Ggww$	0.00	0.50	0.00	0.50
$ggWW$	0.00	0.00	1.00	0.00
$ggWw$	0.00	0.00	0.50	0.50
$ggww$	0.00	0.00	0.00	1.00

Conditional probability tables 9–10: together with the principle of combination, these provide constraints for $P \in \mathbb{P}^5$. Table 9 is analogous to table 7 (complete dominance holds, even though pollen shape is linked with flower colour). The difference between tables 8 and 10 reveals the difference between independent assortment on the one hand, and linkage and crossing-over on the other hand. $\min\{0.08, 0.92\}$ is the frequency of crossing-over.

of them ($\frac{x}{2}\%$) will be GW, the other half will be Gw. The rest of the gametes, $(100 - x)\%$, will be the result of crossing-over. Half of these, $\frac{100-x}{2}\%$, will be Gw; the other half will be GW. Consequently, $\frac{x}{2} + \frac{100-x}{2} = 50\%$ of the gametes will be GW, the other half will be Gw. Hence, the gametes of both doubly homozygous plants and singly heterozygous plants are as in the case of independent assortment (even if they result from strongly different mechanisms; i.e. even if linkage and crossing-over take place). Linkage and crossing-over only make a probabilistic difference in the case of doubly heterozygous plants. This is summarized in table 10.

$\mathbb{P}^5$ thus is the set of distributions over V^5 that satisfy tables 9 and 10 (plus the principle of combination). It is important to note that the physical probability a may take two possible values and that hence the distribution over GC_i conditional on GT_i is imprecise. But regarding crosses on flower colour and pollen shape in *Lathyrus odoratus*, the physical probability a may not take any value between 0.08 and 0.92. In crosses involving different kinds of hybrids, however, a may take any value between 0.08 and 0.92. Then it should not be deemed a physical probability, but just a relative frequency.

Other crosses, for example on *Drosophila*, also revealed linkage (and most often crossing-over), but showed different frequencies of crossing-over.[48] The

[48]The crosses discussed by Morgan show another interesting phenomenon, viz. sex linked inheritance (cf. section 4.14).

frequency of crossing-over for some particular pair of genes in some particular organism is most easily determined observationally as follows (provided complete dominance holds for both genes). Let A, B denote dominant alleles, and let a, b denote their recessive counterparts. Let two grandparental individuals (e.g. $AABB$ and $aabb$, or $AAbb$ and $aaBB$) together produce a double heterozygote ($AaBb$) (F_1). Perform the cross $AaBb \times aabb$ (the resulting offspring is F_2). Barring problems of statistical inference (parameter estimation) and barring cases of double crossing-over, the frequency of crossing-over is identical to the proportion of individuals in F_2 that do not phenotypically resemble any of the grandparents.

Crosses on wing colour (yellow versus gray) and eye colour (white versus red) in *Drosophila* revealed a frequency of crossing-over of 1% (99% of the F_2 individuals resembled one of their grandparents in crosses like those just described). Other crosses in *Drosophila* gave other frequencies of crossing-over: 33% cross-over types versus 67% grandparental types (white versus red eyes and miniature versus long wings), or 40% cross-over types versus 60% grandparental types (white versus red eyes and forked versus normal bristles). If there are no cross-over types, *linkage* is *complete*.

> A study of crossing-over has shown that all possible percentages of crossing-over occur, up to nearly 50 per cent. If exactly 50 per cent of crossing-over took place, the numerical result would be the same as when free assortment occurs. That is, no linkage would be observed even though the characters involved are in the same linkage group. Their relation as members of the same group could, nevertheless, be shown by their common linkage to some third member of the series. If more than 50 percent crossing-over should be found, a sort of inverted linkage would appear, since the cross-over combinations would then be more frequent than the grandparental types.
>
> The fact that crossing-over in the female of Drosophila is always less than 50 per cent, is due to another correlated phenomenon called double crossing-over. By double crossing-over is meant that interchange takes place twice between two pairs of genes involved in the cross. The result is to lower the *observed* cases of crossing-over, since a second crossing-over undoes the effect of a single crossing-over. This will be explained later. (Morgan, 1928, 19–20)

From this quote, it can be seen how the generic representation of dihybrid crosses with linkage and crossing-over (and with complete dominance) should

look like. Let $\mathfrak{B}^\delta = \langle\langle V^\delta, E\rangle, \mathbb{P}^\delta\rangle$ be such that (i) $G^\delta = \langle V^\delta, E\rangle$ is value-isomorphic to $G^5 = \langle V^5, E\rangle$, and such that (ii) any $P \in \mathbb{P}^\delta$ satisfies complete dominance, the principle of combination and some analogue of table 10, where $a \in \{b, 1-b\}$ for some $b \in [0.00, 0.50]$.

It should be noted that the characterization of $\mathbb{P}^\delta$ should invoke one more principle. This principle relates to the linear ordering of the genes, and will be discussed in section 4.13.

Notwithstanding the need for this extra principle, I can point already to the following interesting point. As Morgan writes, if exactly 50% of crossing-over takes place (i.e. if $b = 0.50$), then the numerical result would be the same as when free assortment occurs. Hence $\mathfrak{B}^\gamma \subset \mathfrak{B}^\delta$, although the underlying mechanisms are different.

4.12 The simultaneous interchange of many genes in crossing-over

In section 4.11 we saw that an interchange (crossing-over) may occur between corresponding linkage groups (or, cytologically, between homologous chromosomes). We also saw that double crossing-over might take place and that this hampers the correct interpretation of the data in dihybrid crosses. It leads to a systematic underrating of the frequencies of crossing-over. This problem is solved by taking into account more than two pairs of traits.

> In the examples of crossing-over just given, two pairs of characters were studied. The evidence involved only those cases of crossing-over that took place once between the two pairs of genes involved in the cross. In order to obtain information as to how frequently crossing-over takes place elsewhere, *i.e.*, in the rest of the linkage group it is necessary to include pairs of characters [traits] that cover the entire group. (Morgan, 1928, 20)

As an example, Morgan discusses a series of crosses along the lines of the method presented in section 4.11.

> For example, if a female with the following nine characters of Group I, scute, echinus, cross-veinless, cut, tan, vermilion, garnet, forked and bobbed, is crossed to a wild type male, and if the F_1 female [...] is back-crossed to the same multiple recessive type, the offspring produced will give a record of every crossing-over. (Morgan, 1928, 20)

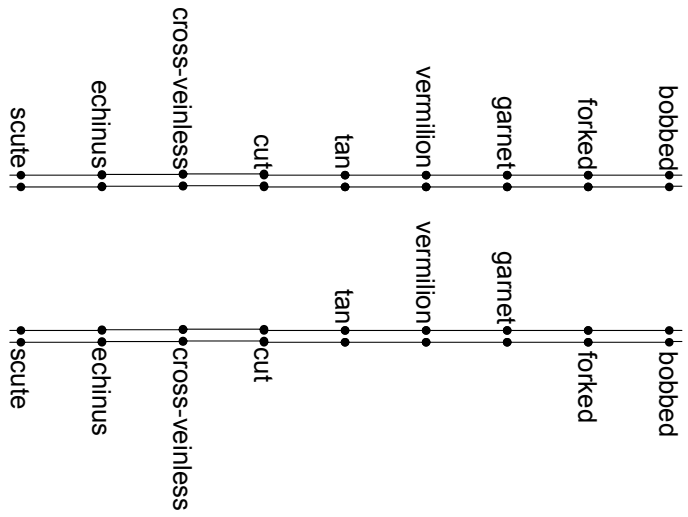

Figure 4.3: *The linear order of the genes in Group I of* Drosophila *(cf. Morgan, 1928, 21). At the top, no crossing-over has occurred. At the bottom, crossing-over took place twice (between tan and cut, and between garnet and forked). (In contrast with Morgan, I have equated all map distances.)*

He draws a diagram depicting the linear order of the corresponding genes in the linkage group (see figure 4.3) and discusses some possible scenarios (Morgan, 1928, 20–22). Crossing-over may take place in the middle of the series, such that two complete halves are interchanged. Or it may take place near one end of the series (e.g. between echinus and cross-veinless). Finally, simultaneous crossing-over may occur at two levels at the same time (cf. supra, double crossing-over). In that case,

> [a]ll the genes in the middle of the two series have been interchanged. This would pass unobserved were there no mutant genes in the region to indicate the fact that two crossings-over had taken place, since the two ends of both series remain the same as before.[49] (Morgan, 1928, 22)

How the Morgan group retrieved the linear order of the genes will be shown in the following section. The reasoning involved is directly relevant for the characterization of $\mathbb{P}^{\delta}$.

[49]If this would pass unobserved, the frequencies of crossing-over would be underrated (cf. supra).

4.13 Frequencies of crossing-over and the linear order of the genes

When genes are linked, they belong to the same linkage group. It can be asked how linkage groups should be represented. As we saw in section 4.12, the answer is: linearly.

> In the preceding illustrations of linkage and crossing-over, that have been given, the genes are represented as lying in a line – like beads on a string. (Morgan, 1928, 24)

On the one hand, the linear order of the genes is of course strongly suggested by the chromosome theory. Cytological evidence showed that chromosomes were threadlike entities (cf. Morgan, 1928, 38–44). If genes are located on the chromosomes, it is most natural to assume they are organized linearly. On the other hand, drawing on assumptions from the chromosome theory, genetic evidence pointed in the same direction in a way that is directly relevant for the characterization of $\mathbb{P}^{\delta}$.

Let us further consider the relations between the behaviour of the genes and the behaviour of the chromosomes. It can be assumed that crossing-over is the result of some interchange taking place at the level of the chromosomes:

> If, as other evidence clearly shows, the chromosomes are the bearers of genes, and if the genes may interchange between members of the same pair, it follows that sooner or later we may expect to find some kind of mechanism by which such interchange takes place. (Morgan, 1928, 39)

Though cytological evidence for such an interchange between like chromosomes was not conclusive, it still was quite convincing:

> [...] it has been shown in a number of cases that the chromosomes are brought into a position where such an interchange might readily be supposed to take place. (Morgan, 1928, 44)

Such considerations gave rise to the concept of *map distance*.[50]

> It is self-evident that if two pairs of genes should be near together, the chance that crossing-over occurs between them is smaller than if they are further apart. If the other genes are still further apart

[50]Sturtevant equated one map unit (mu) with 1 percent recombination. In honour of Morgan's work, map units are often referred to as centimorgans (cM). (Klug et al., 2006, 105)

> the chance of crossing-over is correspondingly increased. We may utilize these relations to obtain information as to the "distance" at which any two pairs of elements lie with respect to each other. (Morgan, 1928, 22)

The concept of map distance allowed to further corroborate the hypothesis of the linear order of the genes.[51]

> Suppose that crossing-over between yellow wings and white eyes occurs in 1.2 per cent of cases. If we then test white with a third member of the same series, such as bifid wings, we find 3.5 per cent of crossing-over [...]. If bifid is in line and on one side of white it is expected to give with yellow 4.7 per cent crossing-over, if on the other side of white it is expected to give 2.3 per cent of crossing-over with yellow. In fact, it gives one of these values, namely, 4.7. We place it, therefore, below white in the diagram. This sort of result is obtained whenever a new character is compared with two other members of the same linkage group. *The crossing-over of a new character is found to give, in relation to two other known factors, either the sum or the difference of*

[51] Morgan's presentation of the argument is somewhat different from mine here. He starts with discussing four drawings of the linear arrangement of nine characters in the first linkage group of *Drosophila* (Morgan, 1928, 20–22, cf. my figure 4.3). This is strange, since no argument has yet been given to support the choice for this linear representation. In each of the drawings, zero, one or two crossings-over are depicted. In these crossings-over, whole series of genes are interchanged simultaneously. Then he introduces the concept of "map distance" (Morgan, 1928, 22) and the 'proof' of the linear order of the genes which I will present below (Morgan, 1928, 24). The relations between linkage, crossing-over and the position of genes on the chromosomes are considered only in later chapters (Morgan, 1928, chapters III, IV). It is obvious that assumptions regarding the relations between chromosomes and genes, and observations on the shape and structure of chromosomes partly guided the choice for a linear representation of the genes as beads on a string, even if Morgan's discussion of the linear order precedes his chapters on the chromosome theory. This is also evident from the following quote:

> The cytologist, then, has given us an account of the chromosomes that fulfills to a degree the requirements of genetics. When we recall the fact that much of the evidence was obtained prior to the rediscovery of Mendel's paper, and that none of the work has been done with a genetic bias, but quite independently of what the students of heredity were doing, it does not seem probable that these relations are mere coincidences, but rather that students of the cell have discovered many of the essential parts of the mechanism by which the hereditary elements are sorted out according to Mendel's two laws and are interchanged in an orderly way between members of the same pair of chromosomes. (Morgan, 1928, 44)

> *their respective cross-over values. This is the known relation of points on a line, and is the proof of the linear order of the genes; for no other spatial relation has yet been found that fulfills these conditions.* (Morgan, 1928, 24, my emphasis)

Of course, map distances should be construed in a relative sense, given that not all parts of the chromosome are evenly prone to interchange.[52]

What is the relevance of the linear ordering of the genes for the characterization of $\mathbb{P}^{\delta}$? Morgan's last citation can be explicated as follows. Let the credal nets $\mathfrak{B}^i = \langle\langle V^i, E\rangle, \mathbb{P}^i\rangle, \mathfrak{B}^j = \langle\langle V^j, E\rangle, \mathbb{P}^j\rangle$, and $\mathfrak{B}^k = \langle\langle V^k, E\rangle, \mathbb{P}^k\rangle$ be distribution-identical to $\mathfrak{B}^{\alpha} = \langle\langle V^{\alpha}, E\rangle, \mathbb{P}^{\alpha}\rangle$. In other words, let the members of $\mathfrak{B}^i, \mathfrak{B}^j$, and $\mathfrak{B}^k$ be descriptions of monohybrid crosses with complete dominance. Assume, moreover, that these crosses concern the same kind of organisms, e.g. *Drosophila*. Then three kinds of dihybrid crosses are feasible. Let $\mathfrak{B}^l = \langle\langle V^l, E\rangle, \mathbb{P}^l\rangle$ be distribution-identical to $\mathfrak{B}^{\delta}$ and let $V^l = \mathit{compound}\langle V^i, V^j\rangle$. In other words, the members of $\mathfrak{B}^l$ describe dihybrid crosses involving the characters described by the members of $\mathfrak{B}^i$ and $\mathfrak{B}^j$, respectively. Let $\mathfrak{B}^m$ and $\mathfrak{B}^n$ be defined analogously: $\mathfrak{B}^m = \langle\langle V^m, E\rangle, \mathbb{P}^m\rangle$ and $\mathfrak{B}^n = \langle\langle V^n, E\rangle, \mathbb{P}^n\rangle$ are distribution-identical to $\mathfrak{B}^{\delta}$, $V^m = \mathit{compound}\langle V^j, V^k\rangle$, and $V^n = \mathit{compound}\langle V^i, V^k\rangle$. Finally, let a^l, a^m, and a^n be the frequencies of crossing-over for $\mathfrak{B}^l, \mathfrak{B}^m$ and $\mathfrak{B}^n$. In other words, let $a^l = \min\{b^l, 1 - b^l\}, a^m = \min\{b^m, 1 - b^m\}, a^n = \min\{b^n, 1 - b^n\}$ (with $b^l, b^m, b^n \in [0.00, 0.50]$). Then, according to Morgan's findings, if at least two of the three frequencies differ from 0.50 (i.e. if the genes in question all belong to the same linkage group),[53]

$$\text{either } a^l = a^m + a^n \text{ or } a^l = |a^m - a^n|$$

[52]Note also that a crossover event in one region of the chromosome may inhibit a second event in nearby regions. This is called *positive interference*. Positive interference increases as the genes in question are closer. This may be explained by physical constraints preventing the formation of closely aligned chiasmata. (Klug et al., 2006, 114)

[53]It is sufficient to demand that at least two frequencies differ from 0.50. Moreover, to demand that all three of them do so, would be too restrictive. Consider three genes, say A, B and C, and their respective frequencies of crossing-over, say a^{AB}, a^{AC}, and a^{BC}. If these three genes belong to three different linkage groups, they will all assort independently and $a^{AB} = a^{AC} = a^{BC} = 0.50$ (in the case of independent assortment, the label 'frequency of crossing-over' should be interpreted in a loose sense). If they belong to two different linkage groups, say A and B to a first group and C to a second group, then $a^{AC} = a^{BC} = 0.50$. Whether $a^{AB} = 0.50$ or $a^{AB} \neq 0.50$ depends on the distance between A and B on the chromosome (note that $a^{AB} = 0.50$ is a limit case). This shows that it is sufficient to demand that at least two frequencies differ from 0.50. Suppose, now, that they all belong to the same linkage group. In most cases, $a^{AB} \neq 0.50, a^{AC} \neq 0.50$, and $a^{BC} \neq 0.50$, but not necessarily so. This shows that it would be too restrictive to demand that all frequencies differ from 0.50. Even if one of them equals 0.50, the others won't.

In other words, the crossing-over of a new character is found to give, in relation to two other known factors, either the sum or the difference of their respective cross-over values.

4.14 Other patterns of inheritance and the role of environmental influences

Up till now I have shown that a variety of paradigmatic crosses discussed by Morgan can be represented as interrelated sets of causal nets. Before I turn to the next chapter, I want to do two more things. Firstly I want to briefly indicate that many other patterns of inheritance can be represented too. Some are seemingly different but can be represented by credal nets that are distribution-identical to $\mathfrak{B}^{\alpha}, \mathfrak{B}^{\beta}, \mathfrak{B}^{\gamma}$, or $\mathfrak{B}^{\delta}$, others require slight modifications to these generic nets. Secondly I want to elucidate the role of environmental influences in (my representation of) classical genetics. Together, this will show that my account is fairly general (at least regarding the case of classical genetics).

4.14.1 Other patterns of inheritance

Some patterns of inheritance once seemed to provide anomalies for the theory of classical genetics, but eventually turned out to fit it. I will give one example.

Cuénot (1902, 1904, 1905) studied the inheritance of coat colour in mice. In 1905, he discovered a striking exception to the well-known 3:1 ratio's that are covered by $\mathfrak{B}^{\alpha}$. Crosses of yellow mice with mice of other colours revealed that yellow is dominant. But when the resulting F_1 hybrids were bred, less than 75% of the F_2 mice were yellow. (In Cuénot's data the ratio was about 2.55:1. Later the ratio was shown to approximate 2:1.) Moreover, none of the F_2 mice turned out to be homozygous for yellow. These ratio's seemed to challenge the principle of segregation (Mendel's first law) and the initial explanations drastically diverged from $\mathfrak{B}^{\alpha}$. Morgan suggested that they provided evidence contra the 'purity of the germ cells'.[54] Cuénot thought they resulted from non-random or selective fertilization. Eventually, Castle and Little suggested that the allele for yellow coat colour is lethal such that zygotes having two of them are not viable. This last hypothesis turned out to be correct. As a result, no alteration was needed to the explanatory principles of classical genetics. Cuénot's anomalous 2:1 mice ratio's belonged

[54] In that time, Morgan was not yet an adherent of Mendelian genetics.

to $\mathfrak{B}^{\alpha}$'s domain of application. (Darden, 1991, 98–105) In other words, these ratio's were not model anomalies, but monster anomalies (Darden, 1991, 199–201).

Other patterns of inheritance cannot directly be accounted for by $\mathfrak{B}^{\alpha}, \mathfrak{B}^{\beta}$, $\mathfrak{B}^{\gamma}$, or $\mathfrak{B}^{\delta}$: cases of sex-linked inheritance, cases of multiple alleles, cases of multiple factors,[55] of gene interaction, of genetic suppression or epistasis, and of pleiotropy, cases of inheritance in non-diploid organisms, cases of parthenogenesis, cases in which mutations arise, etc. However, each of these cases can easily be accounted for by introducing new generic nets $\mathfrak{B}^{\epsilon}, \mathfrak{B}^{\zeta}, \ldots$ But there is one important class of exceptions that need to be considered: cases where environmental influences heavily influence the phenotype.

4.14.2 The role of environmental influences

In this chapter I have nowhere considered any influence of the environment on the phenotype (let alone on the genotype), even if the environment does influence certain phenotypic traits and may even have an impact on the genotype. (In this respect, the above account is not better than that of Kitcher (1989), Darden (1991), or Balzer and Lorenzano (2000).) In this section I will proceed as follows. Firstly, I will show that classical geneticists did take the possibility of environmental influences into account. Then I will show that environmental influences can easily be incorporated in my framework. Finally I will show that this need not cast serious doubts upon the results of this chapter (although *prima facie* it does).

Morgan refers *passim* to the role of environmental influences. Let me first cite some passages regarding the influences on the phenotype. For example, he cites an experiment on *Drosophila* that disconfirms the presence and absence hypothesis that had been advanced to explain the difference between dominant and recessive characters (see footnote 47). In the experiment a temperature effect was recorded.

> A mutant race of Drosophila is called vestigial [...] because only vestiges of the wings are present, *but if the larvae are reared at a temperature of about* 31° C. *the rudiments are quite long* and in extreme cases may be almost as long as the wings of the wild type. If the gene for producing long wings is absent, how can a high temperature bring it back again? (Morgan, 1928, 75, my emphasis)

Later in his monograph, Morgan cites the findings of Hurst on polyploid roses.

[55] An example of two factors that jointly produce one character is given in section 7.2.2.

> Hurst, who has studied species of Rosa, both wild and cultivated, thinks that the wild diploid species consist of five primary groups [...]. Many combinations of these five fundamental types are recognizable. [...]
>
> Hurst states that each member of the five primary series has at least 50 diagnostic characters. These can be recognized in combinations in the hybrids. *The environmental conditions may alternately favor the expression of one or the other set of characters.* (Morgan, 1928, 163–165, my emphasis)

Regarding the causes responsible for sex reversal, Morgan writes:

> In recent years there has been some discussion concerning the reversal of sex, which means, by implication, that a male, already determined as such, can become changed into a female, and vice versa. It has even been suggested that, if this can be shown to occur, the genetic interpretation of sex is discredited or even overthrown. *It is scarcely necessary to point out that there is nothing in the theory of sex as determined by sex-chromosomes or genes contradictory to the idea that other influences may so affect the development of the individual* as to change or even reverse the balance normally determined by the genes. *To fail to appreciate this is to fail entirely in grasping the ideas that underlie the theory of the gene; for this theory postulates no more than that in a given environment such and such effects are expected as a result of the genes present.* (Morgan, 1928, 261)

Let me now turn to the case of environmental influences changing an organism's genotype.

> Stockard carried out a prolonged series of experiments on the effects of alcohol on guinea pigs. The guinea pigs were treated by placing them in closed tanks over strong alcohol. They breathed the air saturated with alcohol, and after a few hours became completely stupefied. The treatment was carried over a long time. Some of the guinea pigs were bred while undergoing treatment, others only at the end of the treatment. The results were essentially the same. Many young were aborted or absorbed, others were born dead, others showed abnormalities, especially in the nervous system and eyes [...]. Only those that themselves showed no defects could be bred. From these, abnormal young continued to appear along with other individuals normal in appearance. In

> later generations abnormals continued to appear, but only from certain individuals. (Morgan, 1928, 307–308)
>
> Stockard [...] interprets his result to mean that *an injury of some sort to the germ-cells has been produced by the alcohol – an injury to some part of the machinery that is involved in heredity.* (Morgan, 1928, 308, my emphasis)

Stockard's findings were not unique.

> More recently Little and Bagg have carried out a series of experiments on the effects of radium on pregnant mice and rats. When the treatment is properly administered, the young mice in utero may develop abnormally. [...] Some of these embryos die before parturition, and are absorbed, others are aborted. Still others are born alive and some of these survive and may procreate. The offspring often show serious defects in the brain or the appendages. [...] Bagg has bred some of these mice and finds that they produce many abnormal offspring that show defects similar, in a general way, to those induced directly in the original embryos. (Morgan, 1928, 308–309)

The role of environmental influences can be incorporated in my framework in the following way. In figure 4.4 the joint graph of all credal nets discussed so far is embedded in a larger graph that also contains variables denoting environmental influences. Each group (the two parental groups and the filial group) is given an environmental variable E_i. To make the graph not too complex, I have only added causal relations from the environment to the phenotype, and between the environments themselves. It is obvious that the environment of, say, the paternal plants may be causally related to the filial environment, and that the paternal environment and the maternal environment may be influenced by a common cause (that is why I have introduced the variable E_0.

An important problem arises immediately here. If the graph in figure 4.1 can be embedded in the larger graph 4.4, it appears that the former did not satisfy the causal Markov condition. For example, in figure 4.1 the variables PT_1 and PT_2 do not causally influence each other, nor do they have a common cause. Hence by the causal Markov condition (definition 3.2), PT_1 should be independent from PT_2 conditional on GT_1. However, with respect to figure 4.4 the causal Markov condition does not imply this independence relation. There, PT_1 and PT_2 share a common cause, viz. E_0.[56]

[56]In other words, the graph in figure 4.1 is not causally sufficient (cf. chapter 8).

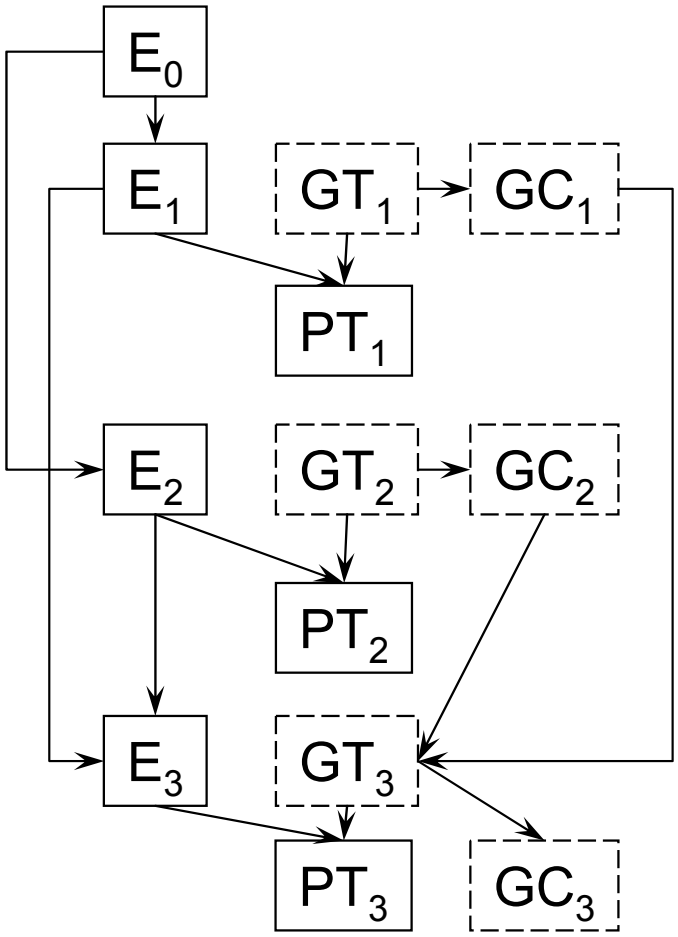

Figure 4.4: *The causal structure of classical genetics, including environmental influences on the phenotype variables*

Shouldn't we conclude from this that the above sections were in vain? No. All genetic crosses cited above were performed in 'standard' (or controlled) environments. Part of the experimental methodology was focussed on keeping the environmental influences as stable as possible. In the language of Woodward (2003b) this can be expressed as follows: by means of interventions, the environmental variables E_1, E_2 and E_3 were set at particular values $E_1 = e_s, E_2 = e_s$ and $E_3 = e_s$ corresponding to these 'standard' environments. This procedure had both a big advantage and a big disadvantage.

Here is the advantage. By implicitly conditioning on the 'standard' environments $E_1 = e_s, E_2 = e_s$ and $E_3 = e_s$, the variables PT_1 and PT_2 are rendered independent. The set $\{E_1, E_2, E_3\}$ d-separates PT_1 and PT_2.[57] Moreover, by conditioning on this set it is assured that any dependencies between PT_1 and PT_3, and between PT_2 and PT_3 are due to the causal relations in figure 4.1. The legitimacy of the theory of classical genetics is assured in the 'standard' or controlled environment.

However, there is one important drawback. The 'standard' environment is artificial and knowledge amassed in artificial contexts does not necessarily extend to other contexts. Unless scientific knowledge can be extended to other contexts, it will be of little practical value.

> In general, what is required for usable knowledge is some claim that one can detach from the particulars of a given observational or experimental situation and export to other contexts [...]. (Mitchell, 2000, 249)

> In order to apply less than ideally universal laws, one must carry the evidence from the discovery and confirmation contexts along to the new situations. As the conditions become less stable, more information is required for application. (Mitchell, 2000, 257)

The theory of classical genetics certainly was not useless and a large part of the enthusiasm for this discipline in the beginning of the twentieth century derived from its prospects regarding applications in e.g. horticulture, agriculture and stock breeding. But such applications were not straightforward.

[57] For the concept of d-separation, see definition 8.5.

4.15 Causal structure and pragmatic laws in classical genetics

In this chapter, I have analysed the causal structure of classical genetics by means of causal nets. It emerged that many aspects of the "Theory of the Gene" fit this framework. Firstly, Mendel's crosses on tall and dwarf pea plants and the principles used to account for the resulting phenotypic distributions and their relation to genotypes and gametic make-ups can be described by the members of one credal net, $\mathfrak{B}^1$. Similar monohybrid crosses with complete dominance can be described by the members of other credal nets, e.g. of $\mathfrak{B}^2$ in the case of eye colour in humans. These credal nets are distribution-identical (and hence also value-isomorphic) to $\mathfrak{B}^1$. I used the generic credal net $\mathfrak{B}^\alpha$ to abstractly represent these credal nets. In the language of Kitcher-Darden, Mendel's crosses on pea plants and the principles that were invoked to describe them formed an exemplar $\mathfrak{B}^1$. From this exemplar a general argument pattern can be obtained, which is represented semantically here by $\mathfrak{B}^\alpha$.

Other crosses comprised anomalies to $\mathfrak{B}^\alpha$. For example, monohybrid crosses with incomplete dominance comprised a small model anomaly which was accounted for by means of different principles. Semantically, cases of monohybrid crosses with incomplete dominance can be accounted for by means of the members of credal nets that are distribution-identical to the generic credal net $\mathfrak{B}^\beta$. In the language of the structuralists, $\mathfrak{B}^\alpha$ and $\mathfrak{B}^\beta$ are two different theory-elements, each of which have a limited domain of intended application and each of which is faced with exceptions or anomalies that are accounted for by the other.

Although $\mathfrak{B}^\alpha$ and $\mathfrak{B}^\beta$ are strongly different (they are not even value-isomorphic), they are nevertheless interestingly similar: they are isomorphic and share the same causal structure. This causal structure is also present in the other theory-elements of classical genetics: $\mathfrak{B}^\gamma, \mathfrak{B}^\delta, \ldots$. All these generic credal nets (and the 'non-generic' credal nets corresponding to them) belong to the same causal scheme! This shared causal structure somehow unifies the domain of intended application of genetics as a whole.

All these concepts not only serve to explicate the causal structure of classical genetics. They also allow us to tackle the concept of pragmatic laws. The sets of causal models of classical genetics were defined by means of conditional probability tables (together with the graph in figure 4.1). All causal models of classical genetics should satisfy several such tables. (Which tables have to be satisfied depends on the theory-element or the domain of intended application in question.) These conditional probability tables

are lists of probabilistic generalizations and they represent the principles of classical genetics.[58] The principles of classical genetics do not satisfy the criteria listed in section 1.1. They are not necessary, as they are the contingent outcome of the evolutionary history. Also, they are not universal or exceptionless. For example, many pairs of traits in many organisms do not relate to each other as dominant and recessive. Likewise, there are cases in which Mendel's first law fails.

All this shows that the principles of classical genetics are not strict laws. By consequence, this makes them excellent candidates for the label 'pragmatic law' and in this chapter I did not eschew terming them so. But are they indeed pragmatic laws? If so, they should be useful for prediction, explanation and/or manipulation. In the following chapters I will show that indeed they can play these roles in scientific practice.

[58]Since they are lists of generalizations, these conditional probability tables are to be considered (part of) the *theory* of classical genetics. The causal nets satisfying them are the *models*, in the sense of logical models, of this theory. To some extent, these models in turn represent things 'out there' and thus can be considered representational models too. They mediate between the theory and 'reality'.

Chapter 5

Explanation in Genetics: a Many-Headed Monster

In chapter 4 I analysed the causal structure of classical genetics and I termed the principles of classical genetics 'pragmatic laws'. If they really are 'pragmatic' laws, they should be useful for prediction, explanation, and/or manipulation. But this has not been shown adequately yet. Admittedly, in the previous chapter I have deliberately used 'description' and 'explanation' on a par. And I have repeatedly called the principles of classical genetics 'explanatory principles'. But this was little more than mere stipulation. I will now provide an account of scientific explanation as it figured in classical genetics. In section 5.1 I will first do some preliminary work and figure out what were the explananda in classical genetics. In section 5.2 I will present an account of explanation of phenotypic distributions, both semantically and syntactically. In section 5.3 I will show that this account is epistemic (or derivational) and that it incorporates mechanistic, causal-probabilistic and interventionist aspects of explanation. In section 5.4 I will show that it provides ontologically grounded derivational unification as a bonus. Finally, in section 5.5 I will discuss the status of other types of explananda in classical genetics.

I will not enlarge in too much detail on the different models of explanation that are, or were, on the philosophical market. I will touch upon the D-N model and the D-S model (Hempel, 1965), the D-N-P model (Railton, 1978, 1981), the mechanistic account (e.g. Machamer et al., 2000; Glennan, 2002) and the unificationist account (Kitcher, 1989) only insofar as is necessary. I will leave the I-S model (Hempel, 1965) and the S-R model (Salmon, 1984, 1989) untouched. I will also sidestep the simulacrum account (Cartwright, 1983, 143–162), although by section 3.1 many aspects of it have implicitly been assimilated. Nor will I systematically discuss the many counterexamples

that have been raised against these models. So I will mostly disregard the flagpoles, the ink-stains, the contraceptive pills and the barometers.

5.1 The explananda in classical genetics

In classical genetics, two general kinds of explananda were addressed. On the one hand, geneticists aimed at explaining phenotypic distributions. On the other hand, they used classical genetics to explain singular events. The first kind of explananda outweighed the second kind and will receive most attention here.

How were phenotypic distributions explained? Recall the phenotypic distribution obtained in cross 2 of section 4.5: self-fertilization of the hybrid F_1-generation (i.e. tall pea plants that originated from a cross between short pea plants and true-breeding tall pea plants) resulted in offspring (F_2) 75% of which was tall, the other 25% being short. At the end of section 4.5, I stated that this distribution can be explained by $\mathcal{B}^3 = \langle G^1, P^3 \rangle$, where $P^3 \in \mathbb{P}^1$ is such that $P^3(GT_1 = ts) = P^3(GT_2 = ts) = 1.00$, since then it follows that $P^3(PT_3 = tall) = 0.75$ and $P^3(PT_3 = short) = 0.25$. I also added that this corresponds to the data.

Two caveats are in order here. The first caveat concerns the fact that, whereas I characterize explanation semantically here, I will also discuss it on the syntactical level (see section 5.2).

The second caveat is more pressing. The probabilities in question are the product of idealization, and the data rarely fitted these ratio's exactly. In Mendel's experiment, of the 1064 F_2-plants, 787 were tall and 277 were short (Mendel, 1865, 13). The corresponding relative frequencies are 73,97% and 26,03%. In Mendel's paper (but also in the writings of many of the later Mendelians) the step from relative frequencies to theoretical probabilities was very intuitive. He merely aggregated and rounded off the results of his seven experiments on *Pisum* (involving seed shape, albumen colour, ..., and stem length respectively).

> When the results of all experiments are summarized, the average ratio between the number of forms with the dominating trait and those with the recessive one is 2,98:1, *or 3:1.* (Mendel, 1865, 13, my emphasis).

I will not pursue the matter of statistical inference in detail here. Nor will I tackle the question why Mendel's data fitted his theoretical explanation too good (see Weldon (1902) and Fisher (1936); see also Wright (1966) and Meijer

(1983, section 3) for possible explanations).[1] I only want to draw attention to the following fact: in the context of phenotypic distributions, my account is primarily aimed at the explanation of idealized statistical generalizations, not of frequency distributions in limited samples.

In the vocabulary of Bogen and Woodward (1988), this amounts to the following: the theory of classical genetics can be used to explain the *phenomenon* of the 3:1 ratio in cross 2, but it cannot be used to explain directly the *data*, or the reported relative frequencies $\frac{787}{1064}$ and $\frac{277}{1064}$, which play the role of evidence for the existence of this phenomenon. Bogen and Woodward do not say that data cannot be explained. Only, data are idiosyncratic to particular experimental contexts and are produced (better: co-determined) by the experimental design, the measurement or detection techniques, the data gathering procedures, etc. It follows that explanations of data do not exhibit detailed patterns of dependency (of the explanandum on the explanans) and that they don't provide unification or systematization. Bogen and Woodward deem both requirements quintessential for explanation.

This distinction between data and phenomena, or between frequency distributions in limited samples and statistical generalizations, is necessary given the following remarks by Wesley Salmon. These remarks are aimed at Hempel's D-S model of explanation (cf. infra), but they apply equally well to the account to be outlined below.

[1]Meijer's explanation is interesting in its own right, since it casts light on the state of statistical inference in Mendel's days. According to Meijer (1983, 128), Mendel's mathematical training was rooted mainly in the works of Doppler, Baumgartner and von Ettingshausen. Concerning statistical inference and the reliability of data, Baumgartner and Ettinghausen wrote the following in their *Die Naturlehre nach ihrem gegenwärtigen Zustand mit Rücksicht auf mathematische Begründung* (I quote Meijer's translation):

> Even the shrewdest observer with the best instruments never gets results which are totally devoid of error, and in order to come as close as possible to the truth, the only thing one can do is to repeat the operation as many times as necessary [Meijer adds: sic!] *and to select from all results those which have the smallest error.* (Meijer, 1983, 128, original emphasis)

Contemporary statistical theory requires that the number of measurements (or the size of the sample) is determined beforehand. Also, outliers (i.e. extreme observations) should only be discarded if good reasons to do so are present:

> A safe rule frequently suggested is to discard an outlier only if there is direct evidence that it represents an error in recording, a miscalculation, a malfunctioning of equipment, or a similar type of circumstance. (Kutner et al., 2005, 108)

It is evident from this quote that deviation from the model to be tested does not provide a good reason.

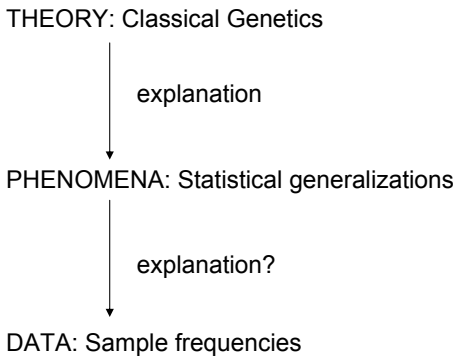

Figure 5.1: *Theory, phenomena, data, and explanation*

> [...] it would be possible to construct a D-S explanation of the fact that, in general, 10 milligrams of radon almost always contain about 2.5 milligrams of radon 7.64 days later. However, it is impossible to deduce from the information given anything about the radon content of this particular sample (beyond, perhaps, the conclusion that the value must be somewhere between 0 and 10 milligrams). (Salmon, 1998, 150–151)

> Since the D-S species of D-N explanations must appeal to statistical laws, and since such laws typically apply to potentially infinite classes, we can say, in principle, that there can be no D-S explanations of frequency distributions in finite samples. (Salmon, 1998, 153)

As I will show in section 5.3, my account to a large extent resembles the D-S model. Hence, to escape Salmon's criticism, I need at least to leave open the possibility of explanation of data.[2]

5.2 Explanation of phenotypic distributions

Let us now turn to the explanation of phenotypic distributions (in the sense of phenomena or statistical generalizations). How is the 3:1 ratio in cross 2,

[2]At present, I have no fully fledged account of explanation of data. It would appear to me that much hinges on the choice between classical and Bayesian statistics. I am also convinced that in the explanation of a set of data $\{a_1, \ldots, a_n\}$ more is involved than a series of D-N-P explanations, one for each a_i (see section 5.5 for the D-N-P model of explanation).

i.e. the pair of generalizations $P(PT_3 = tall \mid PT_1 = tall, PT_2 = tall) = 0.75$ and $P(PT_3 = short \mid PT_1 = tall, PT_2 = tall) = 0.25$, explained by the theory of classical genetics? I will first approach the problem syntactically. Then I will show that this fits the semantic characterization of chapter 4.

Syntactically, one introduces assumptions (hypotheses) regarding the genotype of the parental plants (the F_1 hybrids) and one tries to find a theory-element of classical genetics that allows to derive the observed phenotypic ratio. Which theory-element is appropriate cannot be determined a priori (cf. some alleles show complete dominance, others show incomplete dominance, some genes have more than two alleles, some characters depend on multiple factors, etc.) and the choice of theory-element is closely linked to the genotypic assumptions that are introduced (or: the introduction of the genotypic assumptions is closely linked to the chosen theory-element).

For the explanation of the 3:1 ratio in cross 2, Morgan assumes that both the paternal and the maternal phenotype are caused by a pair of alleles, say a tall-producing allele and a short-producing allele, and that the former is dominant to the latter. This step is *abductive*, but not arbitrary.[3] It is partly based on background knowledge concerning the ancestry of both the paternal and the maternal plant(s) – together with the explanatory principles of $\mathfrak{B}^1$ (cf. genotypes and phenotypes are $\mathbf{T}$-theoretical (definition 4.1)). This background knowledge derives from cross 1 of section 4.5, where it was assumed that the parental organisms had similar genes. At the same time, Morgan assumes this cross belongs to the domain of the theory-element of reciprocal monohybrid crosses with complete dominance. From the above genotypic assumptions, together with the principle of complete dominance (table 1), Mendel's first law (table 2), the principle of combination (table 3) and the assumption of random fertilization, and together with the appropriate mathematical machinery, the 3:1 ratio can be derived. If necessary, the reliability of the genotypic assumptions can be checked by performing test crosses (see e.g. cross 3 of section 4.5), where it is assumed that in these crosses the same explanatory principles are valid.[4]

[3]It will be seen in chapter 7 that this abductive character initially put explanation in genetics in bad light.

[4]In Neapolitan (2004, 216–230), abductive inference in Bayesian networks is treated probabilistically. Let $\mathcal{B} = \langle\langle V, E\rangle, P\rangle$ be a Bayesian network, and let $M \subseteq V, D \subseteq V$, and $M \cap D \neq \emptyset$. M is called the manifestation set, D is called the explanation set. Let $m \in [M]$ and $d \in [D]$ be sets or configurations of values of the variables in M and D, respectively. Then a set of values of the variables in D that maximizes $P(d \mid m)$ is called a most probable explanation (MPE) for m (relative to $\mathcal{B}$). The process of determining such a set is called abductive inference.

To some extent, this fits the reasoning involved in the explanation of the 3:1 ratio in cross 2. The manifestations are that both the paternal plant and the maternal F_1-plant are

In chapter 4 I handled explanation semantically instead of syntactically. For example, I stated that the phenomenon of the 3:1 ratio in cross 2 could be explained by providing a model, $\mathcal{B}^3 = \langle G^1, P^3 \rangle$, that makes this phenomenon true. I will first show that there is a good reason to look at explanation from a semantic point of view. Then I will argue that this dual approach does not result in two different kinds of explanation.

Focussing on the semantic level helps to get a better grip on the causal interpretation of symmetric laws. For example, the relations between the volume, the temperature, and the pressure of some particular sample of gas are described by a symmetric equation, $pV = nRT$ (the volume V, the pressure p and the temperature T may vary; n, the amount of gas (in mol) and R, the gas constant are fixed for the sample). But the symmetry breaks down in any particular instantiation. In one case, a change in temperature may cause a change in volume and/or pressure. In another case, the change in temperature may itself be caused by changes in the volume and/or the pressure. Hence, while the linguistic (mathematical) law is symmetric, each model of the law incorporates asymmetric (i.e. fully directed) graphs (cf. the relation between the language and the semantics of **ALIC** in chapter 8).

All this need not imply, however, that I favour two different kinds of explanation. For any theory-element x, the explanatory principles of x are sound with respect to the models of x. This follows from the characterization of the models.[5] Hence, if some statistical generalization G follows from a set of hypotheses and the explanatory principles of x, then in any model of x in which the hypotheses are true, G will also be true.

5.3 Features of the above account

What features does such an explanation have? The explanandum consists of two related statistical generalizations: $P(PT_3 = tall \mid PT_1 = tall, PT_2 = tall) = 0.75$ and $P(PT_3 = short \mid PT_1 = tall, PT_2 = tall) = 0.25$. The explanans implicitly incorporates statistical generalizations too. These are constitutive of the credal set $\mathbb{P}^1$. And the relation between the explanans

tall, *and* that they each result from a cross of short plants with true-breeding tall plants. Given the theoretic principles incorporated in $\mathfrak{B}^1$, the probability that both the paternal and the maternal F_1-plant are heterozygous is approximately unity. This can be further corroborated via test crosses. It should be noted, however, that abductive reasoning in classical genetics is far more complex than Neapolitan's framework suggests. Neapolitan presupposes that the Bayesian net $\mathcal{B}$ is *given*. In classical genetics, a large part of the abductive inference consists in *finding* a suitable $\mathcal{B}$.

[5] I apologize for the fact that I will not endeavour to prove this claim. So it should rather be considered a conjecture.

and the explanandum is deductive. In these respects, the above explanation fits the D-S model.

The D-S model of explanation is a special kind of deductive-nomological explanation. In deductive-nomological explanation (D-N explanation), the *explanandum* is either a sentence describing a particular fact, or a general law (a universal conditional in first order logic, i.e. a strict law). The *explanans* must consist of (true!) sentences describing particular facts and at least one general law and it must have empirical content (i.e., it must be capable, at least in principle, of test by experiment or observation). Finally, the explanandum must be a logical consequence of the explanans. (Hempel, 1965, 247–249, 336)

In deductive-statistical explanation (D-S explanation), the explanandum is a statement in the form of a statistical law.[6] The explanans is a set of premises that contains indispensably at least one law or theoretical principle of statistical form. And the deduction is effected by means of the mathematical theory of statistical probability. (Hempel, 1965, 380–381)

In my account, the relation between the explanans and the explanandum is deductive, as in the D-S model, but it involves more than the mere theory of statistical probability. It involves the constraints in tables 1, 2 and 3 that are constitutive of $\mathbb{P}^1$, and that represent the explanatory principles listed in section 4.5. *But these are inextricably joined to the causal structure embodied in* G^1. Each $P \in \mathbb{P}^1$ must satisfy the causal Markov condition with respect to G^1.

So the concept of explanation that implicitly figured in chapter 4 can be treated as D-S-*cum*-causation, where the causal component is ingrained in the account. This is highly important. *It is the parental genotypes, not the parental phenotypes, that explain the distribution of the filial phenotypes.* (Better: it is the descriptions of the former that explain the description of the latter.)

Of course, many accounts of explanation have incorporated a causal component and, very often, different accounts of explanation were coupled with different accounts of causation. My account relies both on the notion of 'invariance' (or cP-laws) and on the concept of 'complex-system mechanism' (cs-mechanism) we met in chapter 2. As such, it can be considered a causal-mechanistic variant of D-S explanation.

It is easily seen that the generalizations represented by the conditional probability tables 1, 2 and 3 (relating genotypes, phenotypes and gametic

[6] In Hempel (1965, 376–379), a statement has the form of a statistical law if it is formulated in terms of statistical probabilities. It may be a law of basic statistical form, such as '$P(G \mid F) = r$' (where G and F denote statements, not random variables), or it may be more complex.

make-up), which together constitute $\mathbb{P}^1$ and thus partly determine $\mathfrak{B}^1$, are (or are to be deemed) invariant. Recall that for a generalization to be invariant, it must at least be stable under interventions, although it may also and typically will be stable under other sorts of changes too (see section 1.3).

Let us have a look at the relation between GT and PT. Suppose that some pea plant has the genotype $GT = ts$ and the phenotype $PT = tall$. Then if by some ideal intervention its genotype were changed to $GT = ss$, its phenotype would change to $PT = short$ – or so we would expect from the theory of classical genetics. Hence, the relation between GT_1 and PT_1 described by conditional probability table 1 is invariant ... according to classical genetics.

The invariant relations satisfied by the members of $\mathfrak{B}^1$ together describe a complex-system mechanism. I will elaborate this claim in a minute. First, however, I need to clear up two conceptual ambiguities. The first conceptual ambiguity is found in the works of Woodward. On the one hand, Woodward defines mechanisms (better: representations of mechanisms) as follows (against the background of a structural equation interpretation of directed acyclic graphs):

> It is natural to suppose that if a system of equations correctly and fully represents the causal structure of some system, then those equations should be modular. One way of motivating this claim appeals to the idea that *each equation in the system should represent the operation of a distinct causal mechanism.* (Correlatively, *each complete set of arrows directed into each variable in a directed graph should also correspond to a distinct mechanism.*) (Woodward, 2003b, 48, my emphasis)

Under this view, a system of equations is modular if each equation is independently changeable. In section 4.5, the conditional probability tables 1, 2 and 3 play the role of the equations in question. According to the above characterization, $\mathfrak{B}^1$ contains or represents seven mechanisms (let me call them *local mechanisms*), one for each endogenous variable: $GT_1 \rightarrow PT_1, \ldots, GT_1 \rightarrow GC_1, \ldots,$ and $GC_1 \rightarrow GT_3 \leftarrow GC_2$. On the other hand, Woodward characterizes representations of mechanisms along the lines of the complex-systems account:

> **(MECH)** a necessary condition for a representation to be an acceptable model of a mechanism is that the representation (i) describe an organized or structured set of parts or components, where (ii) the behavior of each component is described by a generalization that is invariant under interventions, and where (iii)

> the generalizations governing each component are also independently changeable, and where (iv) the representation allows us to see how, in virtue of (i), (ii) and (iii), the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves. (Woodward, 2002b, S375)

Let me call mechanisms covered by representations satisfying **(MECH)**, *global mechanisms*. Condition (iii) clearly indicates what is the relation between global and local mechanisms: the former are composed of (or contain) the latter. According to **(MECH)**, $\mathfrak{B}^1$ contains or represents a single global mechanism.

The second conceptual ambiguity concerns the status of mechanisms in classical genetics. On the one hand, the ensemble of genotypes, phenotypes and gametic make-ups, governed by the principles cited in section 4.5, constitutes a global mechanism (a *cs*-mechanism) whose behaviour (the phenotypic distributions) is explained by the behaviours of the components. On the other hand, Morgan et al. (1915) and Morgan (1919, 1928) also treat cytological entities and activities (the chromosomes, meiotic division, physical crossing-over, etc.) as the 'mechanism' underlying Mendel's first law, Mendel's second law, and the principles of linkage and crossing-over. I would not label these cytological entities and activities a complex-system mechanism underlying these laws and principles. Rather, the cytological mechanism provides the physical realization of the genetic mechanism – on the same ontological level.

Given these conceptual clarifications, my claim that the invariant generalizations in question together describe a complex-system mechanism comes down to the following: the genotypes, the phenotypes and the gametic make-ups together constitute a global mechanism. In accordance with the complex-systems literature on mechanistic explanation, the description (the representational model) of the organized behaviours of these components explains the resulting phenotypic distributions.

5.4 Ontologically grounded unification as a bonus

Up till now, I have shown that the above account amounts to a causal-mechanistic kind of D-S explanation. I stick to the notion of D-S explanation since deduction and probabilistic inference played an important role in classical genetics – at least in the case of explanation of phenotypic distributions. In this section, I will show how explanation in genetics also provides

ontologically grounded unification.

But let me first indicate that ontological unification is not only a by-product of explanation in classical genetics. The assumption of the unity of the world was ingrained in Mendel's world view. In the concluding remarks of his *Versuche*, Mendel asks himself to what extent his laws for *Pisum* are generalizable, both to other *Pisum* plants and to other genera. His answer is subtle. On the one hand, inductive generalization should be based on additional research. On the other hand, he believes that to some extent the world is unified:

> Whether variable hybrids of other plant species show complete agreement in behavior [with *Pisum*] also remains to be decided experimentally; one might assume, however, that no basic difference could exist in important matters since *unity* in the plan of development of organic life is beyond doubt.[7] (Mendel, 1865, 43, original emphasis)

Recall that I represented the theory of classical genetics by means of interrelated credal nets $\mathfrak{B}^1, \mathfrak{B}^2, \ldots$, each of which corresponds to an exemplar in *The Theory of the Gene*. Recall also that, according to Darden (1991, 18–19), these exemplars may serve in the construction of Kitcher-style general argument patterns (abstract explanatory patterns or schemas) and that these patterns abstractly characterize mechanisms. In her view, these mechanisms, when they are operating, produce observable data-points as output and fitting an observation into a pattern is a way of explaining it. The set of exemplary patterns in *The Theory of the Gene* then constitutes the explanatory repertoire of classical (or Mendelian) genetics. (As we have seen in section 5.3, the credal nets $\mathfrak{B}^1, \mathfrak{B}^2, \ldots$ indeed represent mechanisms. However, as we have seen in section 5.1, it is advisable to distinguish between data and phenomena. So, we may say, these mechanisms are idealized mechanisms that produce phenomena which are indirectly uncovered via the data.) Finally, recall also that Kitcher directly applies his account of explanatory unification to the theory of classical genetics. He distinguishes between four

[7] The German text is as follows:

> Ob die veränderlichen Hybriden anderer Pflanzenarten ein ganz übereinstimmendes Verhalten beobachten, muss gleichfalls erst durch Versuche entschieden werden; indessen dürfte man vermuthen, dass in wichtigen Punkten eine principielle Verschiedenheit nicht vorkommen könne, da die *Einheit* im Entwicklungsplane des organischen Lebens ausser Frage steht. (Mendel, 1933, 42–43, original emphasis)

patterns: `Mendel`, `Refined Mendel`, `Morgan`, and `Watson-Crick`.[8] For example, the pattern `Mendel` consists of the following items:

> (1) There are two alleles A, a. A is dominant, a recessive.
>
> (2) AA (and Aa) individuals have trait P, aa individuals have trait P'.
>
> (3) The genotypes of the individuals in the pedigree are as follows: i_1 is G_1, i_2 is G_2, ..., i_N is G_N {(3) is accompanied by a demonstration that (2) and (3) are consistent with the phenotypic ascriptions in the pedigree.}
>
> (4) For any individual x and any alleles yz if x has yz then the probability that x will transmit y to any one of its offspring is 0.5.
>
> (5) The expected distribution of progeny genotypes in a cross between i_j and i_k is D; the expected distribution of progeny genotypes in a cross ...{ continued for all pairs for which crosses occur }.
>
> (6) The expected distribution of progeny phenotypes in a cross between i_j and i_k is E; the expected distribution of progeny genotypes in a cross ...{ continued for all pairs in which crosses occur }.
>
> *Filling instructions.* A, a are to be replaced with names of alleles, P, P' are to be replaced with names of phenotypic traits, $i_1, i_2, \ldots, i_N$ are to be replaced with names of individuals in the pedigree, $G_1, G_2, \ldots, G_N$ are to be replaced with names of allelic combinations (e.g. AA, Aa, or aa), D is replaced with an explicit characterization of a function that assigns relative frequencies to genotypes (allelic combinations), and E is to be replaced with an explicit characterization of a function that assigns relative frequencies to phenotypes.
>
> *Classification.* (1), (2), and (3) are premises: the demonstration appended to (3) proceeds by showing that, for each individual i in the pedigree, the phenotype assigned to i by the conjunction of (2) and (3) is that assigned in the pedigree; (4) is a premise; (5) is obtained from (3) and (4) using the principles of probability; (6) is derived from (5) and (2). (Kitcher, 1989, 439)

In Kitcher's theory, unification consists in showing that many different phenomena can be accommodated by (derived from) a small number of such patterns. But, as Mäki (2001) forcefully argues, this account is problematic. I will first present Mäki's claims. Then I will show to what extent I can side with his arguments and to what extent his position needs to be revised. Finally, I will show how explanation in genetics gives rise to ontologically

[8] In the rest of this section, I will only discuss Kitcher's work. Darden's emphasis on diagrammatic representations will not solve the problem put here, viz. that Kitcher's account is not (or not necessarily) ontologically grounded. Diagrammatic representations may or may not be ontologically grounded. But this is a contingent matter of fact. Hence, the arguments in this section equally apply to Darden.

grounded unification, how this unification comes in degrees and how it is nicely captured by the concepts introduced in chapter 3 (isomorphism, value-isomorphism, distribution-identity and causal schemes).

Mäki distinguishes between two kinds of unification, viz. derivational unification and ontological unification (see also Marchionni (2005)).[9] When this distinction is applied to Kitcher's theory, it emerges that the latter advocates "unification as a derivational accomplishment without ontological groundings."[10] (Mäki, 2001, 497) According to Kitcher, explanation is not a matter of describing causal relations in the world. By contrast, causal relations are a function of explanatory relations.

> What is distinctive about the unification view is that it proposes to ground causal claims in claims about explanatory dependency rather than vice versa. (Kitcher, 1989, 436)

> [...] there is no sense to the notion of causal relevance independent of that of explanatory relevance and ... there is no sense in the notion of explanatory relevance except that of figuring in the systematization of belief in the limit of scientific inquiry, as guided by the search for unification. (Kitcher, 1989, 499)

Mäki is dissatisfied with these suggestions. In his view, derivational explanatory unification is one option, but not the only one. As an alternative, he proposes *ontological unification* (see also Mäki, 1990):

> In contrast to derivational unification, ontological unification is based on the referential and representational capabilities of theories, while derivational unification is based on their inferential capabilities [...]. Ontological unification is a matter of redescribing apparently independent and diverse phenomena as manifestations (outcomes, phases, forms, aspects) of one and the same small number of entities, powers, and processes. Those phenomena are thereby revealed to be only apparently independent; as a matter of actual fact, they are dependent on the same underlying structure of entities, forces, and processes [...]. The notion of ontological unification, unlike that of mere derivational unification, is presumed to include some deeper idea of why exactly unification would be a virtue worth pursuing. (Mäki, 2001, 498)

[9] Mäki primarily focusses on explanatory unification in economics, but his arguments may be generalized to explanatory unification in other special sciences as well.

[10] Strictly speaking, Mäki aims this quote against Robert J. Aumann, a game theorist who advocates an instrumentalist interpretation of scientific theories. But, Mäki (2001, 497) writes, "In this respect [Aumann's view] resembles Kitcher's "Kantian" account."

> Unification, according to this picture, is not just a matter of derivational success but rather a matter of *successfully representing how things are related in the causal order of things in the world*. (Mäki, 2001, 500, my emphasis)

Where derivational unification mostly serves as a formal constraint on theories, ontological unification is a contingent matter of fact.

> If there is unity among a set of phenomena, it is a matter of their sharing the same ontic foundations (causes, origins, constituents). Unity among phenomena is a matter of what they are and how they come about, and it is a matter of discovery rather than [Kantian] imposition to establish this. (Mäki, 2001, 498)

Does all this imply that derivational and ontological unification are incompatible, one may ask? Mäki claims they are not. In this respect I would like to take his side. But, as I will argue, his answer is only partly satisfying.

According to Mäki, derivational and ontological unification are compatible:

> Note that ontological unification and *mere* derivational unification are supposed to be [...] contrasted – but putting the suggestion in this way leaves room for the possibility that derivational and ontological unification coincide, or perhaps that derivational unification has partial ontological grounds. (Mäki, 2001, 499, my emphasis)

But whether this is the case (in some particular context), or the extent to which this is the case (in that context), is contingent:

> My hunch is that this is a contingent issue; there is no necessity for the two kinds of unification to be related in one particular way or another. (Mäki, 1990, 499)

Mäki's hunch is plausible, provided derivation is tied to standard logical inference. In this respect I would like to take his side. Kitcher's argument patterns, e.g. `Mendel`, are phrased in natural language. But their expressive power does not exceed that of non-modal second order logic joined with probability theory.[11] Hence, whether Kitcher's argument patterns coincide

[11] Alleles and genotypes may be considered properties (of organisms) instead of objects. Then, in a sense, "There are two alleles A, a. A is dominant, a is recessive.", which is premise (1) of `Mendel`, is a second order sentence. I write 'in a sense', since the quantifier indirectly refers to the filling instructions (A and a are dummy letters, not common second order variables).

with "the causal order of things in the world" (where this causal order is not defined as a function of the explanatory relations), is indeed contingent. In the formula $(\forall x)(Ax \supset Bx)$, the predicate A may denote a property that is causally relevant to the property denoted by B, but this certainly needn't be the case.

If, by contrast, derivation is tied to causal reasoning (in the sense of inference with respect to causal nets), the tie between derivational unification and ontological unification is much stronger. In this sense, Mäki's answer is but partly satisfying. It should be stressed that there are good reasons to draw in causal reasoning. Unification, it is often said, is closely linked to explanation. But it is not clear how *mere* ontological unification, in the sense of *mere* reference and representation, would contribute to this. Representation is description, but description is not the same as explanation (otherwise, I needn't write this chapter). If derivation is tied to causal reasoning, the distinction between derivational and ontological unification largely dissolves.

The relations between the causal nets describing genetic crosses show that the various phenotypic distributions ('those phenomena') are 'only apparently independent' (cf. the quote on page 122). They are dependent on the same underlying structure of entities and processes. In other words, they are dependent on the same complex-system mechanism. The exemplars $\mathfrak{B}^1, \mathfrak{B}^2, \ldots$ each are value-isomorphic and distribution-identical to several other credal nets. By abstracting from their details, generic credal nets $\mathfrak{B}^\alpha, \mathfrak{B}^\beta, \ldots$ can be obtained. Generic credal nets can be regarded the causal (and semantic) analogues of Kitcher's argument patterns. (I will not give an elaborate counterpart for Kitcher's filling instructions here.)

In this framework, unification comes in degrees. The stronger the relations (isomorphism, value-isomorphism, distribution-identity) between credal nets, the stronger the corresponding phenomena are unified. But, and this is very important, all credal nets are isomorphic. All credal nets fit the same causal scheme:

$$\text{for all credal nets } \mathfrak{B} \text{ and } \mathfrak{B}' \text{ of classical genetics: } \mathfrak{C}(\mathfrak{B}) = \mathfrak{C}(\mathfrak{B}').$$

It is this causal scheme that represents the causal structure of classical genetics and that unifies all phenomena in its domain of application.

5.5 Other types of explananda

Before I conclude this chapter, I will briefly examine other types of explananda and see whether they can be explained by classical genetics (and

if so, how). Up to now, I have treated the explanation of statistical generalizations, e.g. why 75% of the offspring of hybrid pea plants in cross 2 is tall and why the rest (25%) is short. In section 5.1, I mentioned another type of explananda: singular events (better: descriptions of singular events). For example, one might want to know why Mary has blue eyes. How can we explain this fact?

A caveat is in order here. In line with the previous sections, I will treat explanation as an epistemic, derivational activity (which is not to say that (all) explanations are purely arguments – cf. infra). In my view, explanation is strongly tied to causation (at least in many contexts). But from this premise one should not conclude that we need an ontic account of explanation. So I disagree with Salmon's plea for such an ontic account (cf. Salmon, 1984, 17–18, 121–124).[12] Regarding the explanation of general regularities, in casu the explanation of those covered by Kepler's laws from those covered by Newton's, he writes:

> While it is true that Kepler's laws can be deduced from Newton's laws, it is also correct to say that the physical regularity exhibited by these specially situated bodies [i.e. relatively light bodies moving in an orbit around another, much more massive body] is part of the general regularity exhibited by all bodies of any sort that possess gravitational mass. The more restricted regularity (which is not a statement, but a physical fact) is part of the more general pattern (which is also a physical fact). Indeed whenever the physical part-whole relation obtains, that is also *a physical fact that is entirely independent of the behavior of language-users and of the epistemic states of scientists.* Thus there is no obstacle to our acceptance of the thesis that scientific explanation involves subsumption under laws and our rejection of the view that explanations are arguments. *It is the* physical *subsumptive relations, not the inferential relations of deductive or inductive logic, which is exhibited by our beautiful scientific paradigms.* The supposition that relations of subsumption *must* be interpreted in terms of logical argument forms is, I believe, one of the most unfortunate errors in modern philosophy of science. (Salmon, 1984, 92, emphases are changed)

According to the ontic conception, "[to] give scientific explanations is to show how events ... fit into the causal structure [or the causal nexus] of the world." (Salmon, 1984, 19) That is fine to me, but I fail to see how this

[12] Salmon's ontic account of explanation is still popular, see e.g. Craver, 2007.

can be done without making use of representational tools. I fear that by excising the mediating role of languages and conceptual schemes, of models, and of logics or inferential frameworks from our account of explanation we would deny ourselves the very possibility of being on our guard against their possibly biasing effects.

In spite of all this, I will sometimes elliptically use ontological language, but only in the interest of readability. For example, I may say that an explanandum-event is explained by its causes instead of saying that the explanandum-statement is explained by deriving it from explanans-statements (premises describing the causes).

So how can we explain why Mary has blue eyes ($PT = blue$)? The most straightforward explanation is in terms of her genotype. Mary's eye colour is caused (and hence explained) by her genotype (according to $\mathfrak{B}^2$ she must be homozygous for the blue producing allele: $GT = blbl$). It is not crucial that according to $\mathfrak{B}^2$ the relation between the genotype and the phenotype is deterministic. In section 4.3 I briefly discussed the concept of 'penetrance', which is defined as the percentage of individuals that show at least some degree of expression of a (mutant) genotype. Although penetrance is not discussed in Morgan (1928), I would not keep it out of the domain of classical genetics. In the glossary of Klug et al. (2006, A-11), penetrance is defined in more detail as follows:

> **(penetrance)** The frequency, expressed as a percentage, with which individuals of a given genotype [say, $GT = y$] manifest at least some degree of a specific mutant phenotype [say, $PT = x$] associated with a trait.

This percentage needn't always equal 100. And it seems to me that it is entirely acceptable to explain why, for example, some particular fruit fly manifests some degree of mutant phenotype x by showing (or hypothesizing) that it has genotype y, by stating that $f\%$ of the flies with genotype y manifest at least some degree of x (where $f < 100$) and by stating as a parenthetic addendum that, *by chance*, this fly manifests at least some degree of x. It is only required that the relation $GT \rightarrow PT$ is causal. But this is ingrained in the theory of classical genetics.

In the last paragraph I implicitly referred to Railton's D-N-P model (Railton, 1978, 1981). In the rest of this section I will do the following. First I will briefly present this model and show where I deviate from it. Then, on the basis of one of these deviations, I will discuss the limits of explanation in classical genetics.

Railton's D-N-P model treats singular probabilistic explanation in a deductive-nomological way (hence: deductive-nomological-probabilistic explanation). Schematically, we may D-N-P explain why e did become G at t_0 (or

why it did *not* become G at t_0) in the following way (Railton, 1978, section 4):

(1) $\forall t \forall x[F_{x,t} \rightarrow Prob(G)_{x,t} = p]$
"At any time, anything that is F has probability p to be G."
(2) A derivation of (1) from our theoretical account of the mechanism(s) at work.
(3) F_{e,t_0}
"e is F at time t_0"
(4) $Prob(G)_{e,t_0} = p$
"e has probability p to be G at time t_0"
(5) $(G_{e,t_0}/{\sim}G_{e,t_0})$
"(e did/did not become G at t_0)"

According to Railton, premise (1) need be a (true) law, but he leaves open the question what statements count as natural laws. He also explicitly leaves open the interpretation of '$\rightarrow$' in (1). My bet is that he would not have excluded the causal interpretation figuring in my account. (He deems causal laws useful, but not indispensable for explanation (Railton, 1978, 207).) The relation between the premises (1)-(3) on the one hand, and the conclusion (4) on the other hand, is deductive-nomological. Hence, the explanation of why e has probability p to be G at t_0 is an argument. (5) is not a conclusion, but a parenthetical addendum. It does not follow from (1)-(4). Therefore I stated in the beginning of this section that, although I deem explanation an epistemic, derivational activity, I do not wish to say that (all) explanations are purely arguments.

There are three ways in which my explanation of the fruit fly's manifesting some degree of x deviates from the D-N-P model. The first and the second I merely mention in the interest of comprehensiveness. The third way is especially interesting with respect to the limits of explanation in classical genetics. Firstly, Railton (1978, 213) demands that (1) is a genuinely indeterministic law. There may be no 'hidden variables' characterizing unknown initial conditions that suffice to account for G deterministically. I cannot guarantee that in all cases of incomplete penetrance such hidden variables are ruled out. Later, Railton weakened his position (Railton, 1981, 249–255). Even if (1) is not genuinely indeterministic, an 'argument' of the above form may be explanatory (in the sense of conveying explanatory information), even if it does not fit the D-N-P model. Secondly, my explanation regarding the fruit fly's manifesting some degree of x deviates from the D-N-P model in that the latter seems to be wedded to a propensionist interpretation of probability (Railton, 1978, 222), but the definition of penetrance is phrased in terms of frequencies. Thirdly, and most interestingly, Railton heavily stresses

the role of (2), the theoretical derivation of (1). For Railton, the existence of general, causal laws that cover the explanandum is not always sufficient: "the search for explanation has also taken the form of a search for mechanisms that underly these laws." (Railton, 1978, 207)

This clearly shows the explanatory limits of classical genetics. Classical genetics may explain why our fruit fly manifests some degree of x by reference to the genotype y. *But it cannot explain why* it is the case that y *causes* x. *Nor can it explain why* homozygosity for blue-producing alleles (either in general for all human beings, or in the case of Mary in particular) *causes* blue eyes. And it cannot be explained *why* blue-producing alleles are *recessive* with respect to their brown-producing counterparts. These are tasks for molecular genetics.[13] In molecular genetics, the mechanisms connecting the genotype, the phenotype *and* the environment (including transcription and translation of the genetic information, regulation of gene expression, etc.) are uncovered.

5.6 The use of pragmatic laws in classical genetics

The results of this chapter show that the label 'explanatory principle', which I have used in chapter 4, does not apply vainly to the principles of classical genetics. This is very important, since explanation is one of the central role of laws in Mitchell's framework. The principles of classical genetics were used to explain phenotypic distributions (in the sense of phenomena or statistical generalizations, not in the sense of frequency distributions in limited samples). This kind of explanation is akin to D-S explanation. Statistical generalizations are derived by means of other statistical generalizations (the explanatory principles of classical genetics). It is important to note, however, that these principles describe causal regularities that together make up a complex-system mechanism. They are causal pragmatic laws. Finally, we saw that explanation in classical genetics also provides ontologically grounded unification.

Phenotypic distributions do not exhaust the scope of possible explananda in classical genetics. Singular events, such as Mary's having blue eyes, can also be explained. Explanation of singular events in classical genetics is akin to Railton's D-N-P model. But there are also interesting explananda that cannot be accounted for by classical genetics. To give one example, the

[13]Note that Bateson's attempts to explain dominance and recessiveness in terms of the presence and absence of alleles were not successful (see footnote 47 of chapter 4).

principle of complete dominance can be used, together with a description of her genotype, to explain why Mary has blue eyes. But it cannot be explained why brown eyes are dominant to blue eyes.

Now that we have seen that non-strict laws may be used for explanation, let us see whether they can also be used for policy. This I will do in the next chapter, where I will consider pragmatic laws more generally (i.e. I will not restrict to the case of classical genetics).

Chapter 6

P-laws, cP-laws and Policy

Up till now we have seen that the concept of '(causal) pragmatic law' is a valuable alternative to the decried concept of 'strict law', and that it is compatible with (but not superseded by) the concept of '*cs*-mechanism'. We have also seen how $(c)P$-laws and the theories in which they figure can be represented by (interrelated sets of) causal nets. Finally, it emerged that causal P-laws may be used for explanation even if they are not strict, as was revealed by the case of explanation in genetics.

This last point was very important, since their use in explanation is one of the main conditions in Mitchell's framework to call non-universal and contingent generalizations (such as the principles of classical genetics) 'pragmatic laws'. But explanation is not the only role of laws in her framework. She also focuses on prediction and manipulation.

I will not discuss prediction explicitly here. At several points in chapter 4 the topic of prediction was dealt with implicitly. I do not intend to say that generating reliable predictions from non-strict laws is straightforward. If some generalization has limited stability and strength it often takes much effort to assess its use for prediction in some intended area of application. But in my view, all these difficulties are also present in the context of policy and manipulation. Therefore I will concentrate thereon.

The concept of manipulation might raise a problem for the present framework, however. In section 6.1 I will present the standard view on the practical value of causal knowledge, viz. that manipulation of causes is a good way to bring about a desired change in the effect (whereas, for example, manipulation of the effect is no good way to bring about a desired change in the causes).[1] If this is true, it can be asked whether non-causal P-laws are of any

[1]This chapter is based on Leuridan, Weber and Van Dyck, "The Practical Value of Spurious Correlations: Selective versus Manipulative Policy", to appear in *Analysis* (Leuridan et al., POL). To avoid change of style, I will continue writing in the first person singular,

use in policy at all. If not, it seems they have little value apart from their applicability for prediction, since they hardly played a role in explanation (except as explanantia). But then the lack of any clear distinction between causal pragmatic laws and non-causal ones should be counted heavily against Mitchell's framework and it would become hard to see what this framework adds to that of Woodward. Fortunately, the situation is not that bad. In section 6.2, I will argue that although the standard view is intuitively very plausible, there are common cases in which some desired goal is obtained on the basis of spurious correlations or non-causal P-laws. (From this one should not conclude, however, that causal knowledge is superfluous in these cases.) I will not concentrate solely on the principles of classical genetics, although these will be addressed too. In section 6.3 I will distinguish between manipulative policy and selective policy to account for the use of non-causal P-laws in policy.

(I will not give an overview of the technical literature on the prediction of policy-outcomes. I refer the reader to the *do*-calculus of Pearl, 2000, chapters 3, 4, and 7 and the manipulation theorem of Spirtes et al., 2000, 47–53.)

6.1 The standard view on the practical value of causal knowledge

In the past 25 years, many philosophers have endorsed the view that the practical value of causal knowledge lies in the fact that manipulation of causes is a good way to bring about a desired change in the effect. This view is intuitively very plausible. For instance, we can predict a storm on the basis of a barometer reading, but we cannot avoid the storm by manipulating the state of the barometer (barometer status and storm are effects of a common cause, viz. atmospheric conditions). In this section, I will present textual evidence which shows that this view is very popular.

In "Causal Laws and Effective Strategies", the first essay of *How the Laws of Physics Lie* (1983), Nancy Cartwright argues that the notion of "effective strategy" is intimately tied to causation. Intuitively, an effective strategy should be conceived of as an adequate way to achieve some desired goal. She distinguishes between *laws of association* and *causal laws*. The former just tell how often two qualities co-occur, "but they provide no account of what makes things happen" (Cartwright, 1983, 21). This is done by causal laws, which relate some cause C to some effect E – e.g. "Smoking causes lung cancer". Cartwright stresses "[...] that causal laws cannot be done

but of course the arguments to be presented are the outcome of this joint work.

away with, for they are needed to ground the distinction between effective strategies and ineffective ones" (Cartwright, 1983, 22). She also claims that "[...] spurious correlations are no grounds for action" (Cartwright, 1983, 34) and that "[...] if one wants to obtain a goal, it is a good (in the pre-utility sense of good) strategy to introduce a cause of that goal" (Cartwright, 1983, 36).

Daniel Hausman describes the practical value of causal knowledge as follows:

> Causes can be used to manipulate their effects, but effects cannot be used to manipulate their causes, and effects of a common cause cannot be used to manipulate one another. (Hausman, 1998, 1)

> Knowing the causes of an event helps one to bring it about. Knowing its effects does not. Knowing what things are related to an event as effects of a common cause does not help one to bring it about either. Causes are *means* and *tools*. People can use them to bring about their effects. (Hausman, 1998, 86, original emphasis)

The following quote of Jim Woodward reveals that he also ties the practical value of causation to manipulation of causes to bring about a desired effect:

> In particular, it is our interest in manipulation that explains why we have (or provides the underlying motivation for our having) a notion of causality that is distinct from the notion of correlation. If one asks why we single out those relationships between X and Y that persist under interventions on X [...], the answer is to be found in our practical interest in changing or controlling nature. Human beings often are (and are justified in believing that they are) in situations in which they can perform actions affecting some variable X [...] meeting the conditions [...] on interventions set out above, and in which their interest is in knowing what will happen to some other variable Y under such an intervention. (Woodward, 2003b, 150)

According to Jon Williamson, one of the reasons why it is convenient to represent the world in a causal way, is stated in the following principle, which he calls Strategy:

> Normally, instigating causes is a good way to achieve their effects. On the other hand instigating effects is not normally a good way to bring about their causes. (Williamson, 2005, 137)

In the following quote, the practical value of causal knowledge is described in the same way:

> [...] if A causes B then intervening to change the value of A can change the value of B but intervening to change the value of B cannot change the value of A. An *intervention* (sometimes called a *divine intervention*) on A is a change in the value of A that is brought about without changing the values of any of A's direct causes in V. Thus an intervention changes A via a causal pathway that is not captured by the modelling context V. (Williamson, 2005, 139, original emphasis)

Judea Pearl also endorses this view:

> What *difference* does it make if I told you that a certain connection is or is not causal? Continuing our example, what difference does it make if I told you that the rooster does cause the sun to rise? This may sound trivial. The obvious answer is that knowing "what causes what" makes a big difference in how we act. If the rooster's crow causes the sun to rise, we could make the night shorter by waking up our rooster earlier and making him crow – say, by telling him the latest rooster joke. (Pearl, 2000, 337, original emphasis)

In the following quote Pearl also refers to bringing about effects by manipulating causes:

> Causal effects permit us to predict how systems would respond to hypothetical interventions – for example, policy decisions or actions performed in everyday activity. As we have seen [...], such predictions are the hallmark of causal modeling, since they are not discernible from probabilistic information alone; they rest on – and, in fact, define – causal relationships. (Pearl, 2000, 65)

6.2 Why the standard view is too restrictive: three examples

I have used the label *standard view* for the thesis that the practical value of causal knowledge lies in the fact that manipulation of causes is a good way to bring about a desired change in the effect. I will now argue, by means of three examples, that this view is too restrictive.

table 11

	PT_i	
GT_i	*black*	*white*
BB	1.00	0.00
Bb	1.00	0.00
bb	0.00	1.00

table 12

	GC_i	
GT_i	B	b
BB	1.00	0.00
Bb	0.50	0.50
bb	0.00	1.00

Conditional probability tables for crosses on coat colour in guinea pigs (together with some analogue of table 3). The resulting credal set is distribution-identical to $\mathbb{P}^{\alpha}$.

For ages, agriculturists have sought to obtain ever better (in the sense of more productive, more beautiful, more easy to raise, more resistant to diseases, ...) livestock and crops. They sought to obtain some desired phenotype. It should be obvious from chapter 4 how this can be obtained. (I will illustrate it by means of the example used in Leuridan et al. (POL), which was taken from Stansfield, 1983, 24–25.)

Guinea pigs are black or white. The colour is determined by one pair of alleles (B and b), where B is dominant to b and codes for black coat colour. Hence, crossings on guinea pigs regarding coat colour can be described by means of the members of a credal net that is distribution-identical to $\mathfrak{B}^{\alpha}$.

With respect to the phenotypic expression of these genes, we have three causal P-laws. All BB animals are black. All Bb animals are black. All bb animals are white. With respect to the results of crossings, we also have several causal P-laws, e.g.: if x has BB, then if it is crossed with an arbitrary y, all offspring is black. In line with chapter 4, this is summarized in tables 11 and 12.

We also have several non-causal P-laws that can be explained by means of these causal P-laws. If x is white, then it has bb. If x is black, then it has BB or Bb. If x is black and one of its parents white, then it has Bb ... (In Leuridan et al. (POL) these non-causal P-laws are called 'symptom laws', i.e. laws that claim that some phenomenon is a symptom of the presence of a cause.) One important symptom law is the following: If x is black and belongs to a population that was produced by procedure P, x is almost certainly BB. The procedure P refers to a series of artificial selections: start with a population of black animals and let them mate; remove all white animals from the filial population, and let the filial population mate; repeat; repeat; ... A population that is produced this way is called "true-breeding" (cf. section 4.5).

Suppose that we want to produce a black guinea pig. The following strategy would help (though it is not completely reliable, as I will show):

Take a black female (x), and a black male (y). Let this male and

female mate.

The expected result depends on the genetic make-up of x (which by the symptom laws is either BB or Bb) and of y (which also is either BB or Bb). Suppose, on the one hand, that at least one of them has BB. Then all offspring will be black. Suppose, on the other hand, that both have Bb. Then on average 75% of the offspring will be black, 25% white. So it is to be expected that between 75% and 100% of the offspring of x and y will be black. (In about 99,6% of the litters with 4 offspring, which is the average number of young per litter, there will be at least one black offspring in the case of $Bb \times Bb$.)

Suppose we want to produce a white guinea pig. A possible strategy is the following:

> Take a white female and a white male. Let this male and female mate.

This strategy is completely reliable (barring cases of e.g. mutation). By the causal and non-causal P-laws specified above, all resulting offspring will be white.

Though the strategies are useful to obtain a desired goal, they do not rely on a directed causal relation. Instead they rely on a common cause structure, on a non-causal P-law that describes correlated effects of a common cause. The phenotypes of the parents and of the children are non-simultaneous effects of the same cause, viz. the genotype of the parents, as is evident from figure 4.1.

Let us turn now to a second example. Insurance companies ask you a lot of questions before they decide whether or not to give you a life insurance. They do this with the explicit goal to minimize future disbursements. For instance, they usually ask whether and how much you smoke, whether you drink alcohol (and how much) and whether you are pregnant (if you are female). And they ask about a whole list of diseases you might have or have had: tuberculosis, peptic ulcers, heart attack, multiple sclerosis, leukemia, diabetes, malaria, and many others. All these questions relate to possible (remote or more proximate) causes of death. However, the following questions were also found in a list of a Belgian insurance company:

> Were you refused by another company for a life insurance or a health insurance?
>
> Did you ever do an AIDS test?

With respect to the second question, it is important to note that the company already knows whether you have aids or not (this is one of the diseases they ask you about). So they use the fact that you *thought* you might have AIDS (and therefore did a test) as a criterion. But the AIDS test will not kill anyone. If there is a correlation between AIDS tests and death, it is spurious. The policy of the company is based on this spurious correlation. The first question also shows that the company uses spurious correlations as a basis for policy: being refused does not make one less healthy.

Let us now consider the third and last example. A company's productivity in large part depends on the quality and allocation of its personnel. The goal of staff business is therefore to set up an appropriate work force. In many sectors, especially in academia, diplomas and degrees still play an important role in the recruitment of new employees. Her having a diploma or a high degree does not, however, make an applicant fit. If there is a correlation between having a diploma or high degrees and future value for the employer, it certainly is spurious. Rather, it is considered a sign of intelligence, diligence, organizational skills, etc., which are in turn considered as causally relevant for the applicant's expected value for her employer.

6.3 Manipulative policy versus selective policy

As Daniel Hausman has pointed out (cf. the quote in section 6.1) manipulating an effect of a common cause is not a good strategy to bring about a change in another effect of the same cause. Whitening a guinea pig will have no effect on the phenotype of its descendants. Forcing subjects (chosen randomly) to do an AIDS test will not make them die sooner (nor will deprivation from such a test raise their life expectancy). And arbitrarily distributing diplomas or degrees will not improve the productivity of the fortunate receivers. However, the examples in section 6.2 reveal that there is a different way in which causal knowledge can be useful. We should distinguish between two kinds of policy. On the one hand, we have *manipulative policy*. To obtain a goal, one may actively manipulate certain cause variables to obtain a desired value of the goal variable. On the other hand, we have *selective policy*. To obtain a goal we select (from a given population) individuals with certain characteristics that are spuriously related to the goal state.

Manipulative policy is based on causal P-laws. Selective policy is based on non-causal P-laws. But it also involves complex causal knowledge. In the examples above, the policy was based on knowledge about common cause

structures. However, this is only the tip of the iceberg. Selective policy may be based on knowledge of causal structures that are very diverse in character. Consider the following scenario: a student's having a high degree may not only be caused by her being intelligent, diligent, etc., but also by her having enjoyed a lot of tutoring. A company that wishes to recruit a new employee may thus benefit from choosing (selecting) someone with a high degree *amongst those that did not enjoy much tutoring*. Most certainly, exploring the space of possibilities will reveal many interesting ways in which a desired goal can be obtained on the basis of non-causal P-laws, provided the latter are used in an appropriate way.

To conclude, even if non-causal P-laws are unserviceable for manipulation, they are useful for other kinds of policy. Prediction, explanation and manipulative policy do not exhaust the range of roles of laws in scientific practice. Hence the lack of any clear distinction between causal pragmatic laws and non-causal ones should not be counted against Mitchell's pragmatic approach.

Chapter 7

Experiments in Classical Genetics

Up to now I have discussed several topics relating to laws of nature and causality in the special sciences. I showed that there are little or no strict laws in the special sciences and proposed to use the concepts of '(causal) pragmatic law' and '(causal) pragmatic regularity' instead (chapter 1). I also showed that these are compatible with, but not superseded by the concepts of 'mechanism' and 'mechanistic model' (chapter 2). Then I introduced the language of causal modelling and used it to analyse the causal structure of classical genetics (chapters 3 and 4). In chapters 5 and 6 I showed that causal pragmatic laws can be used for explanation and for manipulative policy, and I argued that non-causal pragmatic laws may usefully underly selective policy (a point which is mainly neglected in the philosophy of science).

Here, and in the next two chapters, I will tackle a new set of issues. All are related to the problem of causal discovery. In this chapter, I will zoom in on the case of classical genetics again and show that the language of causal modelling is also suitable for analysing the 'logic' of experimentation. It is generally assumed that experiments and experimental data are superior to merely observational data as regards causal discovery. I subscribe to this assumption, and want to examine to what extent this bears upon causal discovery in classical genetics. In chapter 2 I gave a very general characterization of experiments. In section 7.1 I will elaborate on this characterization and show in what sense it distinguishes experiments from non-experimental designs. This distinction is commonplace in the methodology of the special sciences. Section 7.2 will serve three goals. Firstly, I will show that it was customary in the history of classical genetics to call genetic crosses 'experiments'. Secondly, I will show that these crosses are not experiments in the sense of section 7.1. Thirdly, I will show that genuine experiments were not

completely wanting in the history of classical genetics. In section 7.3 I will discuss the relevance of these findings as regards causal inference in classical genetics. The non-experimental nature of genetic crosses hampered causal discovery, as was argued by the early T.H. Morgan!

7.1 Experiments and non-experimental designs

In section 2.6 I gave the following general characterization of experiments:

(EXP) In an experiment, an *object* is placed in some *controlled environment*. It is *manipulated*, often using some *apparatus*, such that it assumes some definite property $X = x$. Then, again using some apparatus, the outcome is *measured* in some (other) property Y. More specifically, it is verified whether there is some relation between $X = x$ and $Y = y$ (for some or all possible values x of X and y of Y), and if so, what is its strength and how it can be characterized.

I added that the term object should be interpreted as broadly as possible, that an environment is 'controlled' if the relation between X and Y is not influenced or disturbed by other factors and that apparatus are often indispensable in experimental designs for manipulation and/or measurement and/or control of disturbing influences. (Although control of the environment is vital to experimentation, it also hampers the projection of experimental knowledge to, or the application of such knowledge in, other contexts – cf. section 4.14.2.) The central point of this characterization, however, is the notion of *manipulation* since this provides one of the main distinctions made in statistics and the methodology of the special sciences between experimental and observational studies. This is evident from Kutner et al. (2005), an elaborate statistics handbook covering the use of regression analysis and analysis of variance in business administration, economics, the social and behavioral sciences, engineering, the health and biological sciences, etc. (Kutner et al., 2005, vi, 2).

Kutner et al. (2005) define experiments as follows:

> Designed experiments are conducted to demonstrate a cause-and-effect relation between one or more explanatory factors (or predictors) and a response variable. The demonstration of a cause-and-effect relationship is accomplished, in simple terms, by *altering* the levels of the explanatory factors (i.e., the X's) and observing

> the effect of the changes on the response variable Y. (Kutner et al., 2005, 643, my emphasis)

The levels of the explanatory factors are often called treatments. The objects or entities to which the treatments are applied are called experimental units. In a proper experimental design the assignment of treatments to experimental units is governed by randomization. (In section 2.6 I indicated that randomization is an effective and widely-used way to control for disturbing factors.)

> Thus a characteristic feature of an experimental study is that the investigator exercises *control* over the assignment of treatments to the experimental units through the process of randomization. If important differences in the responses result between the treatment groups, we can attribute them to the treatments. (Kutner et al., 2005, 644, my emphasis)

In section 2.6 we saw that randomization occurs twice in experimental designs. Firstly, a sample S is randomly selected from the target population P. Secondly, the experimental units in S are randomly divided into an experimental group S^X (or several such groups) and a control group S^K.

Kutner et al. (2005, 643) cite the following example. In 1976 an experiment was conducted on the effects of vitamin C on the prevention of colds. A sample of 868 children was studied. (Kutner et al. do not mention whether or not these were selected at random from the target population.) Half of these were randomly selected for the experimental group. Children in this group *received* a 1,000-mg tablet of vitamin C daily for the test period ($X = 1$). The remaining children (the control group) received a placebo (an identical tablet containing no vitamin C; $X = 0$) also on a daily basis. The variable X is treated as a qualitative variable with two levels or values here.[1] The average number of colds per child was .38 for the first group, and .37 for the second group. (The response variable thus is a ratio variable.) The difference was not statistically significant.

By contrast, the assignment of treatments or factor levels to experimental units is not subject to randomization in observational studies. In other words, no second-phase randomization occurs. (The first-phase randomization preferably does occur.) Therefore, they are less suitable for demonstrating causal relations.

[1] It could also be treated as a quantitative variable with levels ranging from 0-mg to, say, 3000-mg (i.e. as a ratio variable).

> A cause-and-effect relationship between the explanatory factors and the outcome or response variable is difficult to establish in an observational study. Usually, evidence external to the observational study would be required to rule out possible alternative explanations for cause and effect. (Kutner et al., 2005, 645)

> [...] designed observational studies do not directly demonstrate cause-and-effect relationships between the explanatory factors and the response. They can establish association between explanatory factors and a response, and provide the basis for further study of potential cause-and-effect relationships. To infer causality, potential confounding variables would need to be identified, and subgroup analysis performed to try to rule out possible alternative factors. (Kutner et al., 2005, 666)

Several kinds of observational designs can be distinguished. I will focus on prospective designs, since these will turn out to be relevant in the following sections.[2]

> In a *prospective observational study*, one or more groups are formed in a nonrandom manner according to the levels of a hypothesized causal factor, and then these groups are observed over time with respect to an outcome variable of interest. Prospective studies answer the question: "What is going to happen?" (Kutner et al., 2005, 667)

In a prospective design, the groups are self-selected. Kutner et al. (2005, 645) cite the following example. The administration of a college of business offered its faculty the opportunity to participate in a summer workshop on case teaching methods. Faculty were not required to attend the workshop, but were asked to sign up on a first-come, first-served basis. Of the 110 faculty in the business school, 63 faculty elected to attend the seminar. At the end of the following academic year, the administration compared the recent teaching performances of faculty who attended the seminar ($X = 1$) to those who did not attend ($X = 0$). (Again, we have a qualitative variable. Since subscription was on a first-come, first-served basis, it seems to me

[2] In a *retrospective observational study*, groups are defined on the basis of an observed outcome, and the differences among the groups at an earlier point in time are identified as potential causes. Retrospective designs answer the question: "What has happened?" (Kutner et al., 2005, 667) In a *cross-sectional study* measurements are taken at a single point in time. The levels of the potential cause factors and the response variable are determined simultaneously. (Kutner et al., 2005, 666)

that it could also have been treated as an ordinal variable, e.g. subscribers 1 to 10, versus subscribers 11 to 20, versus ...) Students evaluated faculty on a 7-point scale (1=poor performance, ..., 7=outstanding). Faculty who attended the seminar were generally rated more highly by students than faculty who did not attend. The difference between the average ratings (5.76 versus 5.26) was statistically significant, but this in itself does not show that the workshop enhanced the faculty's teaching performances. It may be the case, for example, that attendance and teaching performances have a common cause such as the faculty's motivation to work.

This distinction between experimental designs and observational studies has frequently been discussed in philosophy of science (see inter alia Woodward (2003a,b) and Giere (1997)).[3] The implications of this distinction as regards causal discovery have been extensively treated in the literature on causal modelling: see for example Pearl (2000), Spirtes et al. (2000), Neapolitan (2004), Williamson (2005), and Korb and Nyberg (2006). In chapter 8 I will briefly zoom in on the results of this literature and present an adaptive logic for causal discovery. First, however, I will discuss the status of 'experiments' in classical genetics.

7.2 The status of 'experiments' in genetics

This section will serve three goals. Firstly, in section 7.2.1 I will show that it was customary in the history of classical genetics to call genetic crosses 'experiments'. To this end I will use quotations from Mendel, de Vries, Correns and Bateson. In many cases, I will add the original German or French version. In section 7.2.2 I will show that these crosses are not experiments in the sense of section 7.1. Finally, in section 7.2.3 I will show that genuine experiments were not completely wanting in the history of classical genetics.

7.2.1 'Experiments' in the works of Mendel, de Vries, Correns and Bateson

In the history of classical genetics it was customary to call genetic crosses 'experiments'. This is evident from the works of Mendel, de Vries and Correns. It is also evident from the works of Bateson, who explicitly joins the difference

[3]It should be noted that intervention or manipulation in experiments needn't always be material. Mary S. Morgan (2003) discusses the case of model experiments, virtual experiments and in-between hybrid experiments in which (little or) no material intervention occurs.

between experimentation and non-experimental research to the reliability of the resulting knowledge.

The German title of Mendel's main paper is "Versuche über Pflanzen-Hybriden". Eva R. Sherwood translated it as "Experiments on Plant Hybrids". The paper begins as follows:

> Artificial fertilization undertaken on ornamental plants to obtain new color variants initiated the *experiments* to be discussed here. The striking regularity with which the same hybrid forms always reappeared whenever fertilization between like species took place suggested further *experiments* whose task it was to follow the development of hybrids in their progeny. (Mendel, 1865, 1, my emphasis)

The original German text is as follows:

> Künstliche Befruchtungen, welche an Zierpflanzen deshalb vorgenommen wurden, um neue Farbenvarianten zu erzielen, waren die Veranlassung zu den *Versuchen*, die hier besprochen werden sollen. Die auffallende Regelmässigkeit, mit welcher dieselben Hybridformen immer wiederkehrten, so oft die Befruchtung zwischen gleichen Arten geschah, gab die Anregung zu weiteren *Experimenten*, deren Aufgabe es war, die Entwicklung der Hybriden in ihren Nachkommen zu verfolgen. (Mendel, 1933, 1, my emphasis)

Mendel discusses several such experiments:

> This paper discusses the attempt at such a detailed *experiment*. It was expedient to limit the *experiment* to a fairly small group of plants, and after a period of eight years it is now essentially concluded. Whether the plan by which the individual *experiments* were set up and carried out was adequate to the assigned task should be decided by a benevolent judgement. (Mendel, 1865, 2–3, my emphasis)

> Die vorliegende Abhandlung bespricht die Probe eines solchen *Detailversuches*. Derselbe wurde sachgemäss auf eine kleine Pflanzengruppe beschränkt und ist nun nach Verlauf von acht Jahren im Wesentlichen abgeschlossen. Ob der Plan, nach welchem die einzelnen *Experimente* geordnet und durchgeführt wurden, der gestellten Aufgabe entspricht, darüber möge eine wohlwollende Beurtheilung entscheiden. (Mendel, 1933, 4, my emphasis)

In 1900, Correns, de Vries and Tschermak all referred to Mendel's work or that of their own as 'experiments'. In section 4.9.1 I mentioned the following quote by de Vries:[4]

> My *experiments* have led me to the two following statements: [...]. [...] As my *experiments* show, they possess generalized validity for true hybrids. (de Vries, 1900a, 110)

In the same year 1900, de Vries wrote a French paper, "Sur la Loi de Disjonction des Hybrides", which begins as follows:

> According to the principles which I have expressed elsewhere (*Intracelluläre Pangenesis*, 1889), the specific characters of organisms are composed of separate units. One is able to study, *experimentally*, these units either by the phenomena of variability and mutability, or by the production of hybrids. (de Vries, 1950, 35, my emphasis)
>
> D'après les principes que j'ai énoncés ailleurs (*Intracellulare Pangenesis* [sic], 1889), les charactères spécifiques des organismes sont composés d'unités bien distinctes. On peut étudier *expérimentalement* ces unités soit dans des phénomènes de variabilité et de mutabilité, soit par la production des hybrides. (de Vries, 1900b, 845, my emphasis)

After discussing several monohybrid crosses, he concluded as follows:

> The totality of these *experiments* establishes the law of segregation of hybrids and confirms the principles that I have expressed concerning the specific characters considered as being distinct units. (de Vries, 1950, 38, my emphasis)
>
> L'ensemble de ces *expériences* met donc en évidence la loi de disjonction des hybrides et vient confirmer les principes que j'ai énoncés sur les charactères spécifiques considérés comme des unités distinctes. (de Vries, 1900b, 847, my emphasis)

When Correns read this French paper, he immediately submitted a reaction:[5]

> The latest publication of Hugo de Vries: "Sur la Loi de Disjonction des Hybrides", which through the courtesy of the author

[4] I did not consult the original German text.
[5] I did not consult the original German text.

> reached me yesterday, prompts me to make the following statement: In my hybridization *experiments* with varieties of maize and peas, I have come to the same results as de Vries, who *experimented* with varieties of many different kinds of plants, [...]. (Correns, 1900, 119, my emphasis)

> [...] the Abbot Gregor Mendel in Brünn, had, during the sixties, not only obtained the same result through extensive *experiments* with peas, which lasted for many years, as did de Vries and I, but had also given exactly the same explanation, as far as that was possible in 1866. (Correns, 1900, 120, emphases adjusted)

Bateson pursued the same course and added some zealous flavour. But what is most interesting, he joined the difference between experimentation and non-experimental research to the topic of the reliability of scientific knowledge.

> I venture to express the conviction, that if the facts now before us are carefully studied, it will become evident that the *experimental* study of heredity, pursued on the lines Mendel has made possible, is second to no branch of science in the certainty and magnitude of the results it offers. (Bateson, 1902, ix, my emphasis)

> The study of variation and heredity, *in our ignorance of the causation of those phenomena*, *must* be built of statistical data, as Mendel knew long ago; but, as he also perceived, the ground must be prepared by specific *experiment*. The phenomena of heredity and variation are specific, and give loose and deceptive answers to any but specific questions. That is where our *exact* science will begin. Otherwise we may one day see those huge foundations of "biometry" in ruins. (Bateson, 1902, xi, emphases adjusted)

7.2.2 Genetic crosses are no genuine experiments

How should we evaluate Bateson's judgement? When we put the results of chapter 4 and of section 7.1 together, it immediately emerges that genetic crosses (whether they are monohybrid or multihybrid crosses) are not genuine experiments.

The data that were obtained in such crosses concern both the parental and the filial phenotypes. But the former do not cause the latter. Rather, they each are effects of a common cause: the parental genotypes. (The paternal and the filial phenotype are effects of the paternal genotype; the

maternal and the filial phenotype are effects of the maternal genotype.) But this common cause was not observable. It is **CG**-theoretical (where **CG** stands for classical genetics) so it could not be determined without relying on the principles of classical genetics. Worse still, this common cause was not *manipulated* or intervened on. Before the advent of genetic engineering there was no way to purposively alter an organism's genotype.[6]

It can of course be objected that the parental genotypes were indirectly fixed by selecting 'appropriate' parental plants (e.g. short pea plants to make sure that $GT = ss$, or true-breeding tall plants to make sure that $GT = tt$ – cf. section 4.5). This is true, but it will not save the experimental status of genetic crosses. This procedure necessarily relies on the principles of genetics. It is of course possible to select short pea plants without relying on the principles or classical genetics (cf. *phenotype* is **CG**-non-theoretical). It is even possible to select true-breeding tall plants (defined phenotypically). Otherwise, Mendel could not have bought true-breeding seeds from seed dealers (see footnote 10). But unless you rely on the principles of genetics, you cannot in this way select organisms of some desired genotype. The structuralist notion of **T**-theoreticity thus is not only relevant with respect to observation, but also regarding selection and/or manipulation.

The **CG**-theoreticity of *genotype* has a serious drawback. The relation between genotypes and phenotypes is to some extent invariant. At least, that was part of classical genetics' claims. It is ideally possible to perform an intervention (i.e. an ideal intervention) on GT_i with respect to PT_i. At the time, however, it was not possible to perform a real intervention on GT_i with respect to PT_i. The selection of plants on the basis of their phenotype violates the definition of "intervention variable" **(IV)**, and hence of "intervention" **(IN)**. Whereas ideal interventions are all that is needed for the definition of causal relations, real interventions are indispensable for genuine experimentation. The absence of manipulation or intervention shows that genetic crosses are no genuine experiments. They are rather akin to prospective designs.

I write 'akin to', since genetic crosses do not have all the characteristics of

[6]This claim should be nuanced in two ways. Firstly, by saying that there was no way to *purposively* alter an organism's genotype, I explicitly leave room for non-purposive alterations. In fact, in section 4.14.2 we already encountered such non-purposive manipulations. By changing the environmental conditions (e.g. by administering alcohol or radium) changes in the genotype can be effected in mice. By calling these manipulations non-purposive, I mean that they could not be used to obtain some desired phenotype. Secondly, by saying that before the advent of genetic engineering there was no way to purposively alter an organism's genotype, I do not say that such manipulations now can be carried out *easily*.

prospective designs. Whereas in prospective designs there is an experimental group (or several such groups) and a control group, no such control group exists in genetic crosses. The plants in genetic crosses are compared with a theoretic expectation, not with unmanipulated counterparts. The degree of fit between the observational results and the theoretic expectation may then be measured by performing a χ^2 test.

Theoretic expectations of course do not come out of thin air. They are abductively posited or derived. In section 7.3 we will see this had a serious drawback.

7.2.3 Genuine experiments in classical genetics

From the fact that genetic crosses are no genuine experiments it does of course not follow that experiments were totally inexistent in classical genetics. In section 4.14.2 I already cited some examples of genuine experiments in classical genetics that are cited by Morgan (1928, chapter XVIII). In that chapter, Morgan tackles the question whether genes are stable.[7]

> Mendel's theory of heredity postulates that the gene is stable. It assumes that the gene that each parent contributes to the hybrid remains intact in its new environment in the hybrid. (Morgan, 1928, 292)

The experiments by Stockard and by Little and Bagg, which I briefly discussed in section 4.14.2 are such that the environment is manipulated, either by administering alcohol or by applying radium. Then the effects on the genotype are studied to see whether genes are stable under such influences. (Morgan, 1928, 307–310)

Of course, the **CG**-theoretical nature of the genotype still plays a role in these experiments.

> Since the gene cannot be studied directly by physical or chemical methods, our conclusions concerning its stability must rest on deductions [better: abductions] from its effects. (Morgan, 1928, 292)

7.3 Causal inference in classical genetics

What should we conclude from sections 7.2.1 and 7.2.2? Should we chide the early Mendelians for their poor methodology? That is certainly not my

[7] This question was very important given the then discussions about the inheritance of acquired characteristics.

intention. At the time, the theory of the design and analysis of experiments in the life sciences had not yet been developed. In 1925, Fisher published his *Statistical Methods for Research Workers*. It was a practical handbook, albeit somewhat cursory and elementary. It was based on the experience he acquired at Rothamsted Experimental Station. In this book the principles of randomization were first expounded. In 1935 he published *The Design of Experiments*, which was more elaborate.[8] (Yates and Mather, 1963)

Thus, at the time Mendel, Correns, de Vries and Bateson presented the results of genetic crosses, the theory of experimental design had not yet been developed. At Morgan's time, the theory was at its infancy. (But Fisher's *Statistical Methods for Research Workers* is not included in the bibliography of Morgan (1928).) If it is not my intention to chide the early Mendelians, what is? In the rest of this section I will elaborate the problem of causal discovery from non-experimental data with unobservable variables. It will be seen that this problem also engaged geneticists in the beginning of the twentieth century.

The non-experimental nature of genetic crosses hampers the use of such crosses for causal discovery. In order to infer the existence of causal relations between genotypes, gametic make-ups and phenotypes, a lot of background assumptions are needed. But given that the selection of parental genotypes depends on the theory of classical genetics itself, these background assumptions are not independently underpinned. Hence it is not surprising that Mendel and the later geneticists performed many test crosses to check them.

An objection seems to be in order here. Wasn't this problem quickly evanescent? After all, once the existence of pairs of 'genes' was well-confirmed and once their relation to the phenotype was well understood, the selection of genotypes on the basis of phenotypes was no longer problematic in practice.

To a certain extent, this objection is sound. The theory of classical genetics made it possible to purposively obtain offspring of some desired phenotype by crossing appropriate parental organisms – where the selection of the parental organisms is based on their phenotype (see chapter 6). The theory was thus invaluable for horticulture, agriculture and stock breeding. But we saw in chapter 4 that the history of classical genetics is replete with theory changes to account for recalcitrant anomalies. It often took a long time before anomalies were accounted for, and the proposed solutions frequently were not independently grounded.[9] Darden cites the following example:

[8]Unfortunately, I could not consult any of these two works.

[9]The chromosome theory and the resulting solution to the problem of linkage phenomena was an important exception to this finding. Independent grounding from cytology was quintessential for the development of the chromosome theory.

> East (1910)[10] while studying color inheritance of seeds in maize, found a ratio of 15 yellow to 1 white plant in the F_2 generation, with the yellow varying from dark yellow to light yellow. He explained this by postulating two pairs of factors that interacted in producing the yellow color. Thus the 9:3:3:1 ratio could be extended from cases of *two* independent pairs of factors producing *two* independent characters to cases in which *two* pairs of factors interacted in the production of *one* character in a "continuously" varying array in the F_2 generation. The 9:3:3 ratio collapsed to 15. (Darden, 1991, 66, original emphasis)

Similar ratios were explained in the same way, by invoking two, three or even more pairs factors (i.e. multiple factors). This posed methodological problems, and it kindled objections by opponents of Mendelian genetics. Originally, T.H. Morgan strongly opposed Mendelian genetics.[11] More specifically, he opposed the "modern factor-hypothesis", i.e. the assumption that phenotypic traits are caused by different material units (factors) which separate in the germ cells "after having lived together through countless generations of cells without having produced any influence on each other" (Morgan, 1909, 366) – see also Darden (1991, 67).[12]

[10]The paper referred to is East, E.M. (1910), "A Mendelian Interpretation of Variation That is Apparently Continuous," *American Naturalist*, 44: 65–82. I did not consult this paper myself.

[11]T.H. Morgan opposed Mendelism until about 1911.

[12]Morgan opposed to the preformation character of the factor-hypothesis or segregation-hypothesis.

> The factors have become entities that may be shuffled like cards in a pack, but cannot become mixed. The whole mechanism turns on the old preformation idea of the way the characters of the adult are contained in the egg. The success of the method as a ready means of explanation does not, in my opinion, justify the procedure [...]. (Morgan, 1909, 366)

Instead, Morgan favoured the epigenetic conception.

> I think that the condition of two alternative characters may equally well be imagined as the outcome of alternative states of stability (or of conditions) that stand for the characters that make up the individual. (Morgan, 1909, 366)

> The egg need not contain the *characters* of the adult, nor need the sperm. Each contains a particular material which in *the course of the development produces* in some unknown way the character of the adult. (Morgan, 1909, 367, original emphasis)

Morgan's partiality for the epigenetic conception was based on several grounds. Firstly, he was convinced that having more than one hypothesis or interpretation was favourable

> In the modern interpretation of Mendelism, facts are being transformed into factors at a rapid rate. If one factor will not explain the facts, then two are invoked; if two prove insufficient, three will sometimes work out. The superior jugglery sometimes necessary to account for the result, may blind us, if taken too naïvely, to the common-place that the results are often so excellently "explained" because the explanation was invented to explain them. (Morgan, 1909, 365)

Morgan continues by drawing attention to the abductive nature of explanation in genetics (cf. section 5.2) and by arguing in favour of an instrumental interpretation of the theory.

> We work backwards from the facts to the factors, and then, presto! explain the facts by the very factors that we invented to account for them. I am not unappreciative of the distinct advantages that this method has in handling the facts. I realize how valuable it has been to us to be able to marshal our results under a few simple assumptions [...] So long as we do not lose sight of the purely arbitrary and formal nature of our formulae, little harm will be done [...]. (Morgan, 1909, 365)

Not all Mendelians considered their formulae as purely arbitrary or formal. But where did these formulae come from? The theoretic expectations certainly did not come out of thin air. They were abductively posited or derived. But the scope of possible explanations was seemingly unlimited (see footnote 4 of chapter 5, where I contrasted the complexity of abduction in classical genetics with the simplicity of Neapolitan's framework for abductive reasoning in Bayesian networks). Hence Morgan's reticence was not unfounded. Only by performing many different crosses on many different kinds of organisms and on many different characters, by systematically performing test crosses to check the explanatory principles invoked, by repeatedly incorporating apparent anomalies, and by gradually incorporating the results of i.a. cytology, classical genetics eventually developed into a well-established theory. But nothing was straightforward.

---

for scientific progress: "[...] the more such [interpretations] we have, the less likely are we to become blind followers of one idea" (Morgan, 1909, 367). Secondly, genetic evidence indicated that Mendelian segregation is not universally or generally valid. Thirdly, he thought he had good cytological evidence for his epigenetic interpretation. Finally, experimental embryology provided indirect (better: analogical) evidence contra the preformation idea.

7.4 Causal discovery in the special sciences

We may conclude this chapter by stating that a large part of the empirical basis of classical genetics was non-experimental and that this hampered causal discovery. But classical genetics is not the only scientific discipline that is largely thrown back on non-experimental data. In epidemiology, in the social sciences, in medicine, ... experiments are sometimes feasible and permissible, but not always so. Fortunately, more and more algorithms for causal discovery from merely observational data are being developed. These algorithms are phrased within the framework of causal modelling. In the following chapter, I will discuss one such algorithm, show that it faces a problem that is important and widespread in the special sciences, and present a solution based on the adaptive logic framework.

Chapter 8

Causal Discovery and the Problem of Ignorance. An Adaptive Logic Approach.

In chapter 7 I presented the distinction between experimental designs and observational studies. Whereas the former involve manipulation, the latter do not. I also indicated that the former are more suitable for demonstrating causal relations (cP-regularities) than the latter. This does not mean, however, that non-experimental causal discovery is out of the question. Indeed, in the past decades, many interesting algorithms for causal discovery from (purely) observational data have been developed. But these algorithms still face several problems.

In this chapter,[1] I will zoom in on one such algorithm (Judea Pearl's **IC** algorithm), on one such problem it faces (a problem that I will call the 'problem of ignorance'), and on a possible solution (an adaptive logic for causal discovery). Besides, I will pass by the theory of classical genetics for a while, and seek alliance with occupational medicine, the social sciences, epidemiology, ...

8.1 Introduction: causal discovery and the problem of ignorance

Since the end of the 1980's, the interrelations between probability theory, graph theory and causal discovery have been studied by increasing numbers

[1] This chapter is based on my paper "Causal Discovery and the Problem of Ignorance. An Adaptive Logic Approach", to appear in the *Journal of Applied Logic* (Leuridan, CDPI).

of research groups. Different algorithms have been developed to infer causal relations from non-experimental statistical data.[2]

In this chapter, I will discuss Judea Pearl's **IC** algorithm (Pearl, 2000, 50–51). It is one of the best known algorithms for causal discovery, or '*i*nductive *c*ausation', and its merits can hardly be overrated. Nevertheless, it faces a hard and very important problem. In this section, I want to substantiate three related claims. Firstly, in scientific practice, the *problem of ignorance* is ubiquitous, persistent, and far-reaching. Intuitively, the problem of ignorance bears upon the following situation. A finite set of random variables V is studied but only *partly* tested for (conditional) independencies; i.e. for some variables A, B and for some sets of variables $\mathbf{Q}$ *it is not known whether* A *and* B *are independent (conditional on* $\mathbf{Q}$*).* So 'ignorance' should not be understood as 'probabilistic knowledge', or 'degree of belief < 1', but as the existence of 'undecided independencies'.[3] Secondly, the **IC** algorithm cannot be applied in cases of ignorance. It presupposes that a *full list* of (conditional) independence relations is on hand.[4] If it would be applied to *partial* lists, this would moreover lead to unsatisfactory results. Thirdly, the problem of ignorance can be solved without losing the strong points of the **IC** algorithm, viz. by means of an *adaptive logic for causal discovery.*[5]

Let me briefly dwell on the first claim. In scientific practice, for example in the social sciences or in epidemiology, the problem of ignorance is *ubiquitous.* This may be illustrated by an example. The influence of many different factors on cognitive skills and educational achievement are studied by many different research groups. Some research groups focus on cultural factors, some on sociological factors, still others on psychological ones. Other research groups focus on biological, chemical or other factors. In total, hundreds of variables are studied.[6] The combined research of all these groups gives rise

[2]See Spirtes et al. (2000, chapters 5–6) and Neapolitan (2004) for an overview of several different such algorithms.

[3]In 2006, Rolf Haenni and Stephan Hartmann devoted a special issue of *Minds and Machines* to the topic of Causality, Uncertainty and Ignorance. Unfortunately, none of the papers in question treated 'ignorance' in the sense just stated, viz. the presence of undecided independencies. (Haenni and Hartmann, 2006)

[4]In the rest of this chapter, I will use the following abbreviations. '*UIR*' will stand for 'unconditional independence relation'. Likewise, '*CIR*' will stand for 'conditional independence relation'. '*IR*' will be used as an umbrella term. Note that the relations in question are particular relations; they are relations between particular (sets of) variables.

[5]Since the adaptive logic to be presented is based on **IC**, it involves only finite sets of finite variables. Like **IC**, it also assumes causal sufficiency (i.e. if two variables under study share a common cause, this cause is observed too).

[6]The ISI Web of Knowledge cites hundreds of articles on "educational achievement", published between 2000 and 2007. Many of them report non-experimental data (e.g. the National Education Longitudinal Study). The scope of the factors studied is huge:

to the confirmation and disconfirmation of many *CIRs* and *UIRs*, as in each study several possible confounders are tested for.[7] However, not all possible (conditional) independencies are tested.

The problem of ignorance is moreover *persistent*. Even if many conditional independence relations can be ruled out *a priori*, on the basis of reliable background knowledge, micro-level knowledge, common sense arguments, ..., many others will still be undecided.

Finally, the problem of ignorance is *far-reaching*. If the causal interpretation of non-experimental data requires that all (conditional) independencies are decided, as in the **IC** algorithm, then observational science would lose a large part of its materiality. No causal knowledge could ever be obtained in the short run.[8]

In section 8.5, I will present **ALIC**, the adaptive logic for causal discovery which properly solves the problem of ignorance while doing justice to the merits of Pearl's **IC**. First, however, I will present the formal background to **IC** (section 8.2), and also the algorithm itself (section 8.3). Then I will present **LIC**, a non-adaptive logic for causal discovery (section 8.4). By itself, **LIC** has little to add to the **IC** algorithm, but its significance derives from the role it plays in the formulation of **ALIC**.

8.2 Formal background to causal discovery

In this section I will briefly present the formal background to **IC**. Part of this has already been dealt with in chapters 3 and 4. In the interest of readability, I will restate some of the concepts defined there. Section 8.2.1 deals with directed acyclic graphs and their relations with probability distributions. Section 8.2.2 is on faithful indistinguishability classes and patterns. Finally, section 8.2.3 treats of the graphoid axioms and their meta-theoretic properties (the relevance of which will prove in sections 8.4.4 and 8.5.5).

---

it ranges from television viewing, social capital and self-esteem over parasites, maternal smoking, birth order, ethnicity, ...

[7] I use 'confirmation' and 'disconfirmation' in a loose sense here.

[8] The problem of ignorance not only lurks in scientific practice, but also in everyday human reasoning. Humans often endorse or disaffirm *CIRs* and *UIRs* between many different variables, thereby leaving undecided a large number of *IRs*.

8.2.1 Directed acyclic graphs and probability distributions

As in the previous chapters, causal structures will be described by means of directed acyclic graphs, or *DAGs*. Let me briefly restate what *DAGs* are. A *graph* $G = \langle V, E \rangle$ consists of a finite set of *vertices* V and a finite set of *edges* E. In a *directed graph*, all edges are *directed* ($\rightarrow$). Two vertices A and B are *adjacent* ($A - B$) iff either $A \rightarrow B \in E$, or $B \rightarrow A \in E$ ('$-$' will also be called an undirected edge). There is a *path* between A and B iff there is a sequence of adjacent vertices, beginning with A and ending with B. A *directed path* ($A \Rightarrow B$) is a path that has no *colliders* ($X \rightarrow Y \leftarrow Z$) or *forks* ($X \leftarrow Y \rightarrow Z$). A path that contains no vertex more than once is *acyclic*. A *directed, acyclic graph* (*DAG*) is a directed graph that contains no directed, cyclic paths.

Directed acyclic graphs are closely connected to probability distributions. The vertices in V may denote random variables with a finite number of discrete values. Recall that random variables are represented by italicized capital letters: e.g. $A, B, C, \ldots$ or $X, Y, Z, \ldots$ Values of variables are represented by italicized small letters: e.g. $a, b, c, \ldots$ or $x, y, z, \ldots$ Sets of variables are denoted by bold capital letters: e.g. $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$ or $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots$ Configurations of values for all members of a set of variables are denoted by bold small letters: e.g. $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$ or $\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots$

Let $P(V)$ be a joint distribution over V. P may verify some independence relations (*IRs*).

Definition 8.1 ((Un)conditional Independence) *According to* P, $\mathbf{A}, \mathbf{B} \subseteq V$ *are independent conditional on* $\mathbf{Q} \subseteq V$, *in short* ($\mathbf{A} \amalg_P \mathbf{B} \mid \mathbf{Q}$), *iff* $P(\mathbf{a} \mid \mathbf{b}, \mathbf{q}) = P(\mathbf{a} \mid \mathbf{q})$ *for all* $\mathbf{a}, \mathbf{b}$ *and* $\mathbf{q}$ *(whenever* $P(\mathbf{b}, \mathbf{q}) > 0$*). Likewise,* $\mathbf{A}$ *and* $\mathbf{B}$ *are unconditionally independent, in short* ($\mathbf{A} \amalg_P \mathbf{B}$), *iff* $P(\mathbf{a} \mid \mathbf{b}) = P(\mathbf{a})$ *(whenever* $P(\mathbf{b}) > 0$*).*[9] *(Pearl, 2000, 11)*

A *DAG* $G = \langle V, E \rangle$ can be used to represent the *IRs* verified by $P(V)$. To that extent, P and G need to satisfy two conditions: the causal Markov condition and the Faithfulness condition. (The Markov condition we already encountered in chapter 3. In this chapter, I will often omit the label 'causal'.) These conditions are formulated in terms of kinship relations between variables. B is a *parent* of A iff $B \rightarrow A \in E$. $Parents(A)$ is the set of parents of A. Other kinship relations are defined likewise. By convention, A is its own child and descendant, even though $A \rightarrow A$ and $A \Rightarrow A$ are ruled out in *DAG*s.

[9] Conditional probability is defined as follows: $P(\mathbf{a} \mid \mathbf{b}) = P(\mathbf{a}, \mathbf{b})/P(\mathbf{b})$.

Convention 8.2 *$A \in Children(A)$, and hence also $A \in Descendants(A)$*

Definition 8.3 (Causal Markov Condition) *(Pearl, 2000, 30) $G = \langle V, E\rangle$ and $P(V)$ satisfy the causal Markov condition if and only if each variable is probabilistically independent (according to P) of its non-effects, conditional on its direct causes (where causes and effects are relative to G). In other words, for every $A \in V$,*

$$(A \amalg Nondescendants(A) \setminus Parents(A) \mid Parents(A))$$

Given any graph G, the Markov condition generates a set of *IRs*. But probability distributions that are Markov to G may verify extra *IRs* too. The faithfulness condition rules out such distributions.

Definition 8.4 (Faithfulness Condition) *Let $P(V)$ be a probability distribution generated by $G = \langle V, E\rangle$ according to the Markov condition. G and P satisfy the faithfulness condition iff every IR true in P is entailed by the Markov condition applied to G. (Spirtes et al., 2000, 13)*

As I stated earlier, a *DAG* $G = \langle V, E\rangle$ may be used to represent the *IRs* verified by some $P(V)$. If P is Markov and faithful to G, then all and only those *IRs* that are entailed by the Markov condition applied to G are true in P. It is difficult, however, to delineate the set of these relations. The graph-theoretical concept of d-separation provides an easy means to do this (Spirtes et al., 2000, 44).

Definition 8.5 (d-separation) *Let $G = \langle V, E\rangle$ be a DAG. If $\mathbf{Q} \subset V$ and $A, B \in V \setminus \mathbf{Q}$, then A and B are d-separated given $\mathbf{Q}$ in G, in short $(A \amalg_G B \mid \mathbf{Q})$ iff there is no path U between A and B, such that*

1. *for every collider $\ldots \rightarrow C \leftarrow \ldots$ on U, $Descendants(C) \cap \mathbf{Q} \neq \emptyset$,*[10]
2. *and no other vertex on U is in $\mathbf{Q}$.*

If $\mathbf{X} \neq \emptyset, \mathbf{Y} \neq \emptyset$ and $\mathbf{Z}$ are three disjoint sets, then $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{W}$ iff every member of $\mathbf{X}$ is d-separated from every member of $\mathbf{Y}$ given $\mathbf{Z}$.

[10]Note again that $C \in Descendants(C)$, by convention 8.2.

8.2.2 Faithful indistinguishability classes and patterns

The relation between conditional independence and d-separation is characterized by the following two theorems, the last of which lays at the basis of Judea Pearl's **IC** algorithm for causal discovery.

Theorem 8.6 *$P(V)$ is Markov and faithful to a DAG $G = \langle V, E\rangle$ iff for all disjoint sets $\mathbf{A}, \mathbf{B}$ and $\mathbf{Q}$, $(\mathbf{A} \amalg_P \mathbf{B} \mid \mathbf{Q})$ iff $(\mathbf{A} \amalg_G \mathbf{B} \mid \mathbf{Q})$*

Proof. See Spirtes et al., 2000, 385–393. ■

Theorem 8.7 *If $P(V)$ is faithful to* some *DAG, then it is faithful to the DAG $G = \langle V, E\rangle$ iff*

1. *for all $A, B \in V$, $A - B$ iff $\sim(A \amalg_P B \mid \mathbf{Q})$ for all $\emptyset \subseteq \mathbf{Q} \subseteq V \setminus \{A, B\}$;*
2. *and for all $A, B, C \in V$ such that $A - B - C$, but not $A - C$, $A \rightarrow B \leftarrow C$ is a subgraph of G iff $\sim(A \amalg_P C \mid \mathbf{Q} \cup \{B\})$ for all $\emptyset \subseteq \mathbf{Q} \subseteq V \setminus \{A, B, C\}$.*

Proof. See Spirtes et al., 2000, 393–394. ■

The **IC** algorithm takes as its input a probability distribution P that is Markov and faithful to some underlying *DAG* G_0.[11] The intended output is the *DAG* G_0. In most cases, however, several *DAGs* are statistically indistinguishable from G_0 – i.e. no non-experimental data can distinguish between G_0 and these other *DAGs*. *DAGs* that are statistically indistinguishable belong to the same *indistinguishability class.* (Spirtes et al., 2000, 59, 61) There are several different concepts of indistinguishability and corresponding indistinguishability classes. In the rest of this section, I will discuss *faithful indistinguishability.* Two graphs $G = \langle V, E\rangle$ and $G' = \langle V, E'\rangle$ are faithfully indistinguishable (f.i.) iff for every $P(V)$, $P(V)$ is Markov and faithful to G iff it is Markov and faithful to G'. Whether or not two *DAGs* are f.i. can be easily verified by the following graphic criterion:

Definition 8.8 (Faithful indistinguishability) *Two DAGs G and G' are faithfully indistinguishable iff (i) they have the same vertex set V, (ii) they have the same underlying undirected graph: $A - B$ in G iff $A - B$ in G' and (iii) they have the same unshielded colliders: if $A - B - C$ and not $A - C$ in G or in G', then $A \rightarrow B \leftarrow C$ in G iff $A \rightarrow B \leftarrow C$ in G'. (Spirtes et al., 2000, 61)*

[11] The requirement that P is Markov and faithful to some *DAG* G_0 will prove to be very important in section 8.3.2.

Faithful indistinguishability classes may be represented by a *pattern* $\Pi = \langle V, E \rangle$. A *pattern* is a partially directed graph: E may contain both directed $(\ldots \rightarrow \ldots)$ and undirected edges $(\ldots - \ldots)$ edges. Each pattern Π represents a set of graphs $Repr(\Pi)$. Whether $G = \langle V, E \rangle \in Repr(\Pi)$ may be determined by the following graphic criterion. $G = \langle V, E \rangle \in Repr(\Pi)$ iff

1. G and Π have the same adjacency relations;
2. if $A \rightarrow B$ in Π, then $A \rightarrow B$ in G;
3. if $A \rightarrow B \leftarrow C$ and not $A - C$ in G, then $A \rightarrow B \leftarrow C$ and not $A - C$ in Π.

8.2.3 The graphoid axioms, incompleteness and partial completeness

In this section, I will briefly dwell on some meta-theoretical results regarding the (semi-)graphoid axioms, since these are relevant for the rest of this chapter.[12]

A ternary relation $\amalg$ between *disjoint* subsets of V is a *semi-graphoid* over V iff it satisfies the following axiom schemata[13] (cf. Spohn, 1994, 176):

(G1) $(\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q}) \supset (\mathbf{B} \amalg \mathbf{A} \mid \mathbf{Q})$ (*Symmetry*)
(G2) $(\mathbf{A} \amalg \emptyset \mid \mathbf{Q})$ (*Trivial Independence*)
(G3) $(\mathbf{A} \amalg \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q}) \supset (\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q})$ (*Decomposition*)
(G4) $(\mathbf{A} \amalg \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q}) \supset (\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q} \cup \mathbf{C})$ (*Weak Union*)
(G5) $((\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q} \cup \mathbf{C}) \wedge (\mathbf{A} \amalg \mathbf{C} \mid \mathbf{Q})) \supset (\mathbf{A} \amalg \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q})$ (*Contraction*)

It is a *graphoid* over V iff it also satisfies the following extra schema:

(G6) $((\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q} \cup \mathbf{C}) \wedge (\mathbf{A} \amalg \mathbf{C} \mid \mathbf{Q} \cup \mathbf{B})) \supset (\mathbf{A} \amalg \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q})$ (*Intersection*)

The graphoid axioms are highly relevant for the following three reasons. Firstly, probabilistic conditional independence is a (semi-)graphoid (theorem 8.9). This means that the graphoid axioms may be used to derive *IRs* from other *IRs*. Secondly, however, probabilistic conditional independence is not in general completely axiomatizable (theorem 8.10). But this doesn't alter the fact that, thirdly, some interesting subclasses of (semi-)graphoids *are* completely axiomatizable (theorem 8.11).

[12] In fact, the (semi-)graphoid axioms are not axioms, but axiom schemata. But for reasons of readability, I will use 'axiom' and 'axiom schema' interchangeably.

[13] '$\supset$' denotes material implication; '$\wedge$' denotes classical conjunction.

Theorem 8.9 *For any probability measure P, $\amalg_P$ is a semi-graphoid. Moreover, if P is strictly positive (i.e., if $P(A) = 0$ only for $A = \emptyset$), then $\amalg_P$ is a graphoid.*[14] *(Spohn, 1994, 176)*

So the axioms **(G1)-(G5)** are *sound* for probabilistic conditional independence and **(G6)** is sound in case P is strictly positive. By contrast, **(G1)-(G6)** are *not in general complete* for probabilistic conditional independence. This follows from the following theorem of Milan Studený (1992):

Theorem 8.10 *There is no finite set of independent axioms which is complete for probabilistic conditional independence. More specifically, there is no finite set of rules of the form ($r \geq 0$):*

$$((\mathbf{A}_1 \amalg \mathbf{B}_1 \mid \mathbf{C}_1) \wedge \ldots \wedge (\mathbf{A}_r \amalg \mathbf{B}_r \mid \mathbf{C}_r)) \supset (\mathbf{A}_{r+1} \amalg \mathbf{B}_{r+1} \mid \mathbf{C}_{r+1})$$

such that for any set T of IRs on any set of variables V there is a probability measure $P(V)$ such that $\amalg_P = \amalg$, i.e. such that

$$(\mathbf{A} \amalg_P \mathbf{B} \mid \mathbf{Q}) \textit{ iff } (\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q}) \in Cl(T)$$

(where $Cl(T)$ is the closure of T under the given set of rules)

As I stated above, there are some interesting subclasses of (semi-)graphoids that may be completely characterized. For example:

Theorem 8.11 *For each semi-graphoid $\amalg$ generated by a list of total causes there is a probability measure P such that $\amalg_P = \amalg$. (Spohn, 1994, 180)*

A *list of total causes* is a set of *IRs* for which there is a linear ordering $X_1, X_2, \ldots$ of all variables in V such that the list contains for each X_k exactly one statement of the form $(X_k \amalg \{X_1, \ldots, X_{k-1}\} \setminus \mathbf{J} \mid \mathbf{J})$ for some $\mathbf{J} \subseteq \{X_1, \ldots, X_{k-1}\}$. A (semi)graphoid is *generated by* a list of total causes iff it is the closure of that list under the (semi)graphoid axioms. (Spohn, 1994, 179)

Let $P^*(V)$ be Markov to the *DAG* $G_0 = \langle V, E \rangle$, i.e. $P^*(V)$ is a distribution that verifies, for all $A \in V$, the *IR* $(A \amalg Nondescendants(A) \setminus Parents(A) \mid Parents(A))$ that results from applying the Markov condition to G_0. The set of all these *IRs* is a list of total causes. Hence, the (semi-)graphoid $\amalg$ generated by this list is completely axiomatized by the (semi-)graphoid axioms and for all P that are Markov and faithful to G_0: $\amalg_P = \amalg$.

[14] In this theorem, A denotes an event, not a variable.

8.3 The IC algorithm for causal discovery

In section 8.3.1, I will present the **IC** algorithm. Then I will show to what extent it is impotent regarding the problem of ignorance (section 8.3.2) and give a hint at the solution which I will present in the rest of this chapter (section 8.3.3).

8.3.1 The algorithm

The **IC** algorithm is based on the relations between *DAGs*, probability distributions and f.i. classes described in sections 8.2.1 and 8.2.2. It runs as follows (Pearl, 2000, 50–51):

Input A probability distribution $P^*(V)$ which is faithful to some underlying *DAG* D_0 (or the list of *IRs* that it verifies).

Output A pattern Π^* representing all *DAGs* that form a complete causal explanation for the *IRs* verified by $P^*(V)$; i.e. all *DAGs* G such that $(\mathbf{A} \amalg_{P^*} \mathbf{B} \mid \mathbf{Q})$ iff $(\mathbf{A} \amalg_G \mathbf{B} \mid \mathbf{Q})$. These *DAGs* form the f.i. class of D_0.

Algorithm The algorithm consists of three consecutive steps:

1. For all $A, B \in V$ search for a (possibly empty) set $\mathbf{Q} \subseteq V \setminus \{A, B\}$ such that $(A \amalg_{P^*} B \mid \mathbf{Q})$. Construct an undirected graph G^1 such that $A - B$ iff no such $\mathbf{Q}$ can be found.
2. For all $A, B, C \in V$ such that $A{-}C{-}B$ and not $A{-}B$ in G^1, check if $C \in \mathbf{Q}$. If it is, then continue. If it is not, then $A \rightarrow C \leftarrow B$. The resulting graph is G^2.
3. Starting from G^2, orient as many of the undirected edges as possible subject to two conditions: (i) the orientation should not create a new unshielded collider; and (ii) the orientation should not create a directed cycle. This is done by closing G^2 under the rules R3.1 - R3.4 depicted in figure 8.1.[15]

8.3.2 Taking stock of IC

The **IC** algorithm certainly is meritorious as it provides an interesting means to infer causal relations from non-experimental data. However, it is impotent regarding the problem of ignorance. Firstly, its possible inputs are restricted

[15]The rules R3.1 - R3.4 are necessary (Verma and Pearl, 1992) and sufficient (Meek, 1995a) for obtaining the pattern Π^* representing the intended equivalence class.

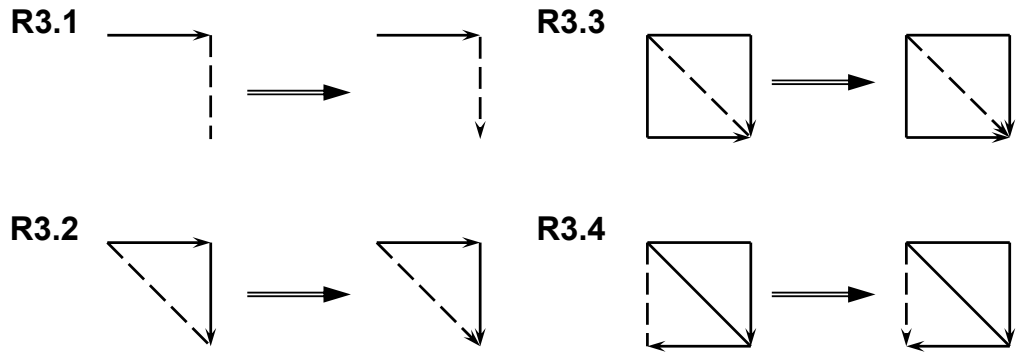

Figure 8.1: **Rules R3.1-R3.4** *According to each of the rules, if a graph contains the left-hand side as a subgraph, an arrow may be added to the dotted line so as to obtain the right-hand side.*

to *full lists* of *IRs*. Secondly, when applied to partial lists of *IRs*, **IC** would lead to unsatisfactory results.

The possible inputs of **IC** are restricted to *full lists* of *IRs*. This follows from the requirement that $P^*(V)$ is Markov and faithful to D_0. The graphoid $\amalg$ generated by applying the Markov condition to D_0 is completely axiomatized by **(G1)-(G6)**. Hence for all $A, B \in V$ and all $\mathbf{Q} \subseteq V \setminus \{A, B\}$ it is known whether or not $(A \amalg_{P^*} B \mid \mathbf{Q})$; all *IRs* are decided.

Now suppose that **IC** were applied to a *partial list* of *IRs*. A straightforward solution would be to take a *negation as failure* account: all *CIRs* $(A \amalg B \mid \mathbf{Q})$ and *IRs* $(A \amalg B)$ not occurring in the input (failure) should be taken to be false (negation). This approach, however, has a serious drawback: it treats conditional and unconditional sentences on a par. From a pragmatic point of view, I will argue, this is unsatisfactory. Whereas the negation as failure account is sensible with respect to the former, it certainly isn't with respect to the latter.

The negation as failure account is sensible regarding *conditional independence* (*CIRs*). Suppose that A and B are known to be unconditionally dependent, $\sim(A \amalg B)$, that all tested sentences $(A \amalg B \mid \mathbf{Q})$ were falsified, but that $(A \amalg B \mid \mathbf{Q}^*)$ is still undecided for some $\mathbf{Q}^*$. Even if correlation is no proof for direct causation, it is often regarded as a useful indicator – and so it should be, otherwise non-experimental causal research would be a rather idle enterprise (cf. the importance attached to epidemiological evidence by the IARC (2006)). This requirement may satisfactorily be met by considering all undecided $(A \amalg B \mid \mathbf{Q}^*)$ as false (negation as failure), *provided faulty applications of this heuristic can be detected and remedied quickly.*

By contrast, the negation as failure account is unsatisfactory for *unconditional independencies* (*UIRs*). A large part of scientific practice consists in finding models that are simple enough to be manageable and useful for prediction, explanation and/or intervention. Models are simulacra that share some, but not all the characteristics of the phenomena under study (Cartwright, 1983). They are the mediators between a theory and reality (Morgan and Morrison, 1999). As such, they do not perfectly mirror this world. Abstraction and idealization are non-negligible aspects of scientific modelling. Moreover, they are already implicitly present in Pearl's framework. The Markov and the Faithfulness condition together imply the Minimality condition (Spirtes et al., 2000, 31).[16] If $G = \langle V, E\rangle$ and $P(V)$ satisfy the Minimality condition, then each edge in G prevents some conditional independence that would otherwise obtain; apart from that, G does not contain any superfluous edges (Spirtes et al., 2000, 12). However, if all undecided $(A \amalg B)$-sentences are taken to be false (negation as failure), the resulting model would be gratuitously complex. In scientific practice, if $(A \amalg B)$ is undecided, then it is frequently or even mostly the case that for all $\mathbf{Q}^*$, $(A \amalg B \mid \mathbf{Q}^*)$ is undecided too. Hence, in view of the previous paragraph, negation as failure would allow to infer that $A - B$. Therefore negation as failure should not be applied to undecided *UIRs*. These should be considered as true, *provided faulty applications of this heuristic can be detected and remedied quickly.*

Before I give a short introduction to adaptive logics in section 8.3.3, I will first pursue some possible epistemological worries. My suggestions may come across as too rash to some readers. Why should we rely on default assumptions which will most probably be violated in many contexts? And what would be the consequences of any such violation? The following three remarks should help to remove such doubts. Firstly, in scientific practice defeasible assumptions are abundantly used for causal inference. In his short but highly influential paper, Sir Austin Bradford Hill (1965, 295) addresses the question how to pass from an observed association to a verdict of causation in occupational medicine, when no general body of medical knowledge provides a decisive answer. After listing nine viewpoints from which to investigate the association in question, he argues that no decisive criterion (or set of criteria) exists. Hence, the inference from association to causation is defeasible. Nevertheless, we are pragmatically forced to judge (whether the association is causal or not, whether the assumed cause is suitable for intervention or not), given that we have to take action:[17]

[16] $G = \langle V, E\rangle$ and $P(V)$ satisfy the Minimality condition if and only if for every $G' = \langle V, E'\rangle$ such that $E' \subset E$ (i.e. for every subgraph G' of G), $G' = \langle V, E'\rangle$ and $P(V)$ do not satisfy the Markov Condition. (Spirtes et al., 2000, 12)

[17] I do not claim that Hill argued in favour of the defaults incorporated in the adaptive

> All scientific work is incomplete – whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time. (Hill, 1965, 300)

Secondly, the use of default assumptions is not new in the literature on causal inference. For example, Williamson (2005, chapters 5,6) proposes to use the Maximum Entropy principle (*MaxEnt*) to handle the problem of ignorance. Suppose we are given a partial list of *IRs*. Then many different probability functions will satisfy this list. If the set of these functions is closed and convex, *MaxEnt* selects the single member which is maximally noncommital with regard to missing information. (The *MaxEnt* principle generalizes the Principle of Indifference; see also Paris and Vencovská, 1997, 2001 and Paris, 2005.) As the selected probability function will mostly be at odds with any new information, the suggested mechanism is defeasible and non-monotonic (cf. infra, section 8.5.6). Thirdly, the adaptive logic framework provides us with a dynamic proof theory which allows us (i) to trace the particular assumptions on which each inference is based, and (ii) to trace the consequences of the violation of each particular assumption (i.e. faulty applications of each heuristic can be detected and remedied quickly). As such, the dynamic proof theory allows us to cautiously apply these default assumptions. This will become clear in section 8.5.4, where I will present the proof theory of **ALIC** and discuss its epistemological implications.

8.3.3 Towards an adaptive logic solving the problem of ignorance

In the following sections, I will show how the problem of ignorance may be solved. I will develop an adaptive logic for causal discovery that properly gives shape to the findings of the previous section.

Using an adaptive logic (instead of, for example, some other default logic) has several advantages. Firstly, the standard format of adaptive logics provides a unified framework for handling various non-monotonic consequence relations (defaults, inconsistency-handling mechanisms, ...).[18] Many defeasible consequence relations have successfully been translated into the adap-

logic **ALIC**, only that he argued in favour of defeasible reasoning.

[18] Given this flexibility, it would be fairly easy to devise alternatives to **ALIC**, based on different rationales, to address the problem of ignorance.

tive logic framework (e.g. Meheus, 2003; Batens, 2000),[19] whereas no other approach is known to have such unifying power. This unified framework makes it possible to easily compare such consequence relations. Moreover, contrary to most other non-monotonic logics (cf. Makinson, 2005), adaptive logics provide a good proof theory that captures the dynamics involved in non-monotonic reasoning. Finally, as I will argue in section 8.5.4, **ALIC**'s dynamic proof theory gives rise to a pragmatic picture of causal inference in which proofs may act as a guide for both scientific research and policy.[20]

Adaptive logics are non-monotonic logics (Batens, 2001, 2004, 2007). In general, they are characterized by a triple $\mathbf{AL} = \langle \mathbf{LLL}, \Omega, \text{adaptive strategy} \rangle$. **LLL** is the *lower limit logic* of **AL**. It is the stable part of **AL**. Semantically, the adaptive models of Γ are a subset of its **LLL**-models (Γ being a premise set). Proof theoretically, all the axioms and inference rules of **LLL** may be applied unconditionally. Hence, if $\Gamma \vdash_{\mathbf{LLL}} \alpha$, then $\Gamma \vdash_{\mathbf{AL}} \alpha$. By contrast, the other axioms or inference rules of **AL** are conditional: they may be applied on the condition that certain other formulas (certain abnormalities) are not derivable. Ω is the set of *abnormalities*. These are formulas that are characterized by a (possibly restricted) logical form and that are presupposed to be false, unless and until proven otherwise. Intuitively speaking, an abnormality $\omega \in \Omega$ is presupposed to be false relative to Γ, unless and until it turns out that Γ forces you to give up this presupposition.[21]

Together, **LLL** and Ω define an upper limit logic **ULL**. Proof theoretically, the **ULL** is obtained by adding to **LLL** an axiom or inference rule that connects abnormality to triviality. Semantically, the **ULL**-models are obtained by selecting those **LLL**-models that verify no abnormality. The following theorem reveals a crucial relation between **LLL**, Ω and **ULL**. ($Dab(\Delta)$ denotes the disjunction of the members of Δ.)

Theorem 8.12 $\Gamma \vdash_{\mathbf{ULL}} \alpha$ *iff there is a finite* $\Delta \subseteq \Omega$ *such that* $\Gamma \vdash_{\mathbf{LLL}} \alpha \vee Dab(\Delta)$ *(Derivability Adjustment Theorem)*

Proof. See Batens (2007, 230–231) ■

[19] Other results can be found at `http://logica.UGent.be/centrum/writings/`.

[20] It has often been argued that adaptive logics fit actual human reasoning. More specifically, this part of the motivation underlying the adaptive logic for causal discovery presented in Van Dyck (2004). I will not endeavour to argue so. In non-scientific reasoning contexts, humans are faced with the problem of ignorance too (cf. footnote 8). But there is insufficient psychological evidence indicating that they solve it along the lines of **ALIC** (for the relation between Bayesian networks and human reasoning, see Gopnik et al., 2004, Beckers et al., 2006).

[21] As I will show in section 8.5.4, the proof theory of adaptive logics is such that it makes sense to write "unless *and until* it turns out that Γ forces you to give up this presupposition".

In general, the following proof theoretic relations hold between **AL**, **LLL**, and **ULL**: If Γ is normal (i.e. if no *Dab*-formulas are **LLL**-derivable from Γ), then

$$Cn_{\mathbf{LLL}}(\Gamma) \subset Cn_{\mathbf{AL}}(\Gamma) = Cn_{\mathbf{ULL}}(\Gamma)$$

If Γ is abnormal, then (except for limit cases[22])

$$Cn_{\mathbf{LLL}}(\Gamma) \subset Cn_{\mathbf{AL}}(\Gamma) \subset Cn_{\mathbf{ULL}}(\Gamma) = \text{ the set of all formulas}$$

One of the best known adaptive strategies is *reliability*. It determines how to treat minimal *Dab*-consequences. $Dab(\Delta)$ is a minimal *Dab*-consequence of Γ iff $\Gamma \vdash_{\mathbf{LLL}} Dab(\Delta)$ and there is no $\Delta^* \subset \Delta$ such that $\Gamma \vdash_{\mathbf{LLL}} Dab(\Delta^*)$. According to *reliability*, a formula is unreliable relative to Γ if it is a disjunct of a *minimal Dab-consequence* of Γ.[23]

8.4 LIC: the lower limit logic of ALIC

In this section I will present **LIC**, a non-adaptive logic for causal discovery. By itself, **LIC** does not add much to **IC**, but its significance derives from the role it plays in the formulation of **ALIC**. I will describe both its language (8.4.1), its semantics (8.4.2) and its proof theory (8.4.3). I will conclude this section by briefly discussing soundness and (in)completeness for **LIC** (8.4.4).

8.4.1 The language of LIC

Let V be a *finite set of finite random variables*. I will assume that all variables in V are different (different name = different variable) and that they are logically independent.[24] Let $\mathcal{W}^f$ be the set of *factual propositions*, i.e. the smallest set satisfying the following conditions:

(L1) If $X \in V$ and $x_i \in [X]$ (i.e. x_i is a value of X), then $X = x_i \in \mathcal{S}$
(L2) If $\alpha \in \mathcal{S}$, then $\alpha \in \mathcal{W}^f$

[22] In limit cases, such as when Γ is itself trivial,

$$Cn_{\mathbf{LLL}}(\Gamma) = Cn_{\mathbf{AL}}(\Gamma) = Cn_{\mathbf{ULL}}(\Gamma) = \text{ the set of all formulas}$$

[23] Another well-known strategy is *minimal abnormality*. The choice of strategy affects both the proof theory and the semantics of the adaptive logic. Here I will stick to reliability.

[24] So V cannot contain both G, 'gender', and K, 'being a king', since by definition, all kings are male.

(L3) If $\alpha, \beta \in \mathcal{W}^f$, then $\sim\alpha, (\alpha \wedge \beta), (\alpha \vee \beta) \in \mathcal{W}^f$

The elements of V and of $\mathcal{W}^f$ by themselves do not belong to the language of **LIC**, but they serve as the basis for the formulation of the latter. The atomic sentences of **LIC** are either *probabilistic* ($\mathcal{W}^p$) or *causal* ($\mathcal{W}^c$). Complex sentences ($\mathcal{W}$) are built from atomic ones by means of classical connectives. So $\mathcal{W}^p$, $\mathcal{W}^c$ and $\mathcal{W}$ are the smallest sets satisfying the following conditions:

(L4) If $\mathbf{A}, \mathbf{B}, \mathbf{Q} \subset V$ are disjoint sets of variables, then $(\mathbf{A} \amalg \mathbf{B})$, $(\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q}) \in \mathcal{W}^p$

Convention 8.13 *If $A, B \in V$ and $\mathbf{Q} \subseteq V \setminus \{A, B\}$, then I will write $(A \amalg B)$ and $(A \amalg B \mid \mathbf{Q})$ instead of $(\{A\} \amalg \{B\})$ and $(\{A\} \amalg \{B\} \mid \mathbf{Q})$. Likewise, if $\mathbf{Q}$ is a singleton $\{Q\}$, I will write Q instead of $\{Q\}$.*

(L5) If $A, B \in V$, then $A \to B, A - B$, and $A \Rightarrow B \in \mathcal{W}^c$

(L6) If $\alpha \in \mathcal{W}^p \cup \mathcal{W}^c$, then $\alpha \in \mathcal{W}$

(L7) If $\alpha \in \mathcal{W}$, then $\sim\alpha \in \mathcal{W}$

(L8) If $\alpha, \beta \in \mathcal{W}$, then $(\alpha \wedge \beta), (\alpha \vee \beta), (\alpha \supset \beta), (\alpha \equiv \beta), (\alpha \veebar \beta) \in \mathcal{W}$[25]

8.4.2 The semantics of LIC

The semantics of **LIC** should meet certain obvious requisites. For one thing, each model should assign appropriate truth values to the atomic probabilistic sentences, to the atomic causal sentences and to the classical complex sentences (see **(Sv1)**–**(Sv12)**). Moreover, no model should allow for cyclicity (see **(Sc.3)**). Finally, the probabilistic sentences verified by each model should be Markov and faithful to the causal sentences it verifies (see **(Sp9)**).

To make sure that the semantics hereunder is recursive, I need to introduce an ordering relation $\prec$ over V. This ordering may be done by lexicographic order, or by the Gödel numbers of the variables. Hence it is not the case that $A \prec B$ iff A is an ancestor of B in some DAG.

- There is no $A \in V$ such that $A \prec A$
- For all $A, B \in V$: either $A \prec B$, or $B \prec A$, but not both
- For all $A, B, C \in V$: $A \prec B$ and $B \prec C$ implies $A \prec C$

[25] The last logical connective, $\veebar$, is the exclusive disjunction, cf. **(Sv12)**. Although the exclusive disjunction can be easily omitted, I include it so as to present the **LIC**-axioms in an intuitive way.

An **LIC**-model is a triple $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$. $\mathbb{R}^+$ is the set of nonnegative real numbers (including 0). c is a function which determines the causal relations holding between members of V. p is a probability distribution over V.

c is a partial function that maps couples of variables $\langle A, B \rangle$ to the set $\{l, r, n\}$, subject to the following conditions:

(Sc1) $\mathsf{c} : V \times V \rightarrow \{l, r, n\}$.

(Sc2) $\mathsf{c}(\langle A, B \rangle)$ is only defined for $A \prec B$.

(Sc3) There is no $\{X_1, \ldots, X_n\} \subseteq V$ such that either $\mathsf{c}(\langle X_i, X_{i+1} \rangle) = l$ for all $1 \leq i \leq n-1$ and $\mathsf{c}(\langle X_1, X_n \rangle) = r$, or $\mathsf{c}(\langle X_i, X_{i+1} \rangle) = r$ for all $1 \leq i \leq n-1$ and $\mathsf{c}(\langle X_1, X_n \rangle) = l$.[26] (acyclicity)

In the interest of the readability of the metatheorems, however, I introduce the following convention:

Convention 8.14 *"$\mathsf{c}(\langle A, B \rangle) = l$" should be read as: "If $A \prec B$, then $\mathsf{c}(\langle A, B \rangle) = l$. If $B \prec A$, then $\mathsf{c}(\langle B, A \rangle) = r$." In the same manner, I will use "$\mathsf{c}(\langle A, B \rangle) \neq n$" as an abbreviation for "If $A \prec B$, then $\mathsf{c}(\langle A, B \rangle) \neq n$. If $B \prec A$, then $\mathsf{c}(\langle B, A \rangle) \neq n$." etc. etc.*

p is a function assigning a 'weight' to sentences in $\mathcal{W}^f$. It is defined as the composition of two other functions, m and o. m assigns a weight to the elements of a set $\mathfrak{S}$. o maps factual propositions to subsets of $\mathfrak{S}$.

(Sp1) Let $\mathfrak{S}$ be a set with at least the cardinality of the sample space defined by V.

(Sp2) $\mathsf{m} : \mathfrak{S} \rightarrow \mathbb{R}^+$

(Sp3) $\sum_{\alpha \in \mathfrak{S}} \mathsf{m}(\alpha) = 1$

(Sp4) For all $\beta \subseteq \mathfrak{S} : \mathsf{m}(\beta) = \sum_{\alpha \in \beta} \mathsf{m}(\alpha)$

(Sp5) For all $\beta \subseteq \mathfrak{S} : \mathsf{m}(\beta^c) = 1 - \mathsf{m}(\beta)$.

(Sp6) For all $\beta, \gamma \subseteq \mathfrak{S} : \mathsf{m}(\beta \cap \gamma) = \mathsf{m}(\beta) + \mathsf{m}(\gamma) - \mathsf{m}(\beta \cup \gamma)$.

(Sp7) $\mathsf{o} : \mathcal{S} \rightarrow \wp(\mathfrak{S})$, where $\mathcal{S}$ is the set of atomic factual propositions $X = x_i$.

o can be extended to a function mapping all factual propositions to subsets of $\mathfrak{S}$: $\mathsf{o} : \mathcal{W}^f \rightarrow \wp(\mathfrak{S})$, with

- $\mathsf{o}(\sim\alpha) = (\mathsf{o}(\alpha))^c$ for all $\alpha \in \mathcal{W}^f$,
- $\mathsf{o}(\alpha \wedge \beta) = \mathsf{o}(\alpha) \cap \mathsf{o}(\beta)$ for all $\alpha, \beta \in \mathcal{W}^f$, and

[26] Note that I use the shorthand notation for the values of c throughout this section (cf. convention 8.14).

- $\mathsf{o}(\alpha \vee \beta) = \mathsf{o}(\alpha) \cup \mathsf{o}(\beta)$, for all $\alpha, \beta \in \mathcal{W}^f$

p is composed of m and o. Moreover, by condition **(Sp9)**, p is Markov and faithful to c.

(Sp8) $\mathsf{p}(x) = \mathsf{m}(\mathsf{o}(x))$

(Sp9) If $A, B \notin \mathbf{Q}$, then $(A \amalg B \mid \mathbf{Q})$ iff every undirected path between A and B is d-separated by $\mathbf{Q}$. More precisely, $\frac{\mathsf{p}(A=a \wedge (B=b \wedge \mathbf{Q}=\mathbf{q}))}{\mathsf{p}(B=b \wedge \mathbf{Q}=\mathbf{q})} = \frac{\mathsf{p}(A=a \wedge \mathbf{Q}=\mathbf{q})}{\mathsf{p}(\mathbf{Q}=\mathbf{q})}$ for all values a of A, b of B and $\mathbf{q}$ of $\mathbf{Q}$ (whenever $\mathsf{p}(B = b \wedge \mathbf{Q} = \mathbf{q}) > 0$) iff

1. $\mathsf{c}(\langle A, B \rangle) = n$, and
2. for any n-tuple $\langle X_1, \ldots, X_n \rangle$ such that (i) $n \geq 3$, (ii) $X_1 = A$, (iii) $X_n = B$, and (iv) $\mathsf{c}(\langle X_i, X_{i+1} \rangle) \neq n$ for all $1 \leq i \leq n-1$, at least one of the following conditions is satisfied:
   (a) There is some $1 \leq i \leq n-2$ such that $\mathsf{c}(\langle X_i, X_{i+1} \rangle) = \mathsf{c}(\langle X_{i+1}, X_{i+2} \rangle)$ $(= r$ or $= l)$ and $X_{i+1} \in \mathbf{Q}$.
   (b) There is some $1 \leq i \leq n-2$ such that $\mathsf{c}(\langle X_i, X_{i+1} \rangle) = l$, and $\mathsf{c}(\langle X_{i+1}, X_{i+2} \rangle) = r$, and $X_{i+1} \in \mathbf{Q}$.
   (c) There is some $1 \leq i \leq n-2$ such that $\mathsf{c}(\langle X_i, X_{i+1} \rangle) = r$, and $\mathsf{c}(\langle X_{i+1}, X_{i+2} \rangle) = l$, and $X_{i+1} \notin \mathbf{Q}$ and for all X such that there is an m-tuple $\langle Y_1, \ldots, Y_m \rangle$ such that $Y_1 = X_{i+1}$, and $Y_m = X$, and for all $1 \leq i \leq m-1, \mathsf{c}(\langle Y_i, Y_{i+1} \rangle) = r$, $X \notin \mathbf{Q}$.

A valuation function v_M determined by a model $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$ is a function that satisfies the following conditions:

(Sv1) $\mathsf{v}_M : \mathcal{W} \to \{0, 1\}$

(Sv2) $\mathsf{v}_M((\mathbf{A} \amalg \mathbf{B})) = 1$ iff $\frac{\mathsf{p}(\mathbf{A}=\mathbf{a} \wedge \mathbf{B}=\mathbf{b})}{\mathsf{p}(\mathbf{B}=\mathbf{b})} = \mathsf{p}(\mathbf{A} = \mathbf{a})$ for all values $\mathbf{a}$ of $\mathbf{A}$ and $\mathbf{b}$ of $\mathbf{B}$ (whenever $\mathsf{p}(\mathbf{B} = \mathbf{b}) > 0$)

(Sv3) $\mathsf{v}_M((\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q})) = 1$ iff $\frac{\mathsf{p}(\mathbf{A}=\mathbf{a} \wedge (\mathbf{B}=\mathbf{b} \wedge \mathbf{Q}=\mathbf{q}))}{\mathsf{p}(\mathbf{B}=\mathbf{b} \wedge \mathbf{Q}=\mathbf{q})} = \frac{\mathsf{p}(\mathbf{A}=\mathbf{a} \wedge \mathbf{Q}=\mathbf{q})}{\mathsf{p}(\mathbf{Q}=\mathbf{q})}$ for all values $\mathbf{a}$ of $\mathbf{A}$, $\mathbf{b}$ of $\mathbf{B}$ and $\mathbf{q}$ of $\mathbf{Q}$ (whenever $\mathsf{p}(\mathbf{B} = \mathbf{b} \wedge \mathbf{Q} = \mathbf{q}) > 0$)

(Sv4) $\mathsf{v}_M(A \to B) = 1$ iff $\mathsf{c}(\langle A, B \rangle) = r$

(Sv5) $\mathsf{v}_M(A - B) = 1$ iff $\mathsf{c}(\langle A, B \rangle) \neq n$

(Sv6) $\mathsf{v}_M(A \Rightarrow B) = 1$ iff there is a series of variables $C_1, \ldots, C_m \in V$ $(m \geq 2)$ such that $A = C_1$, $B = C_m$, and for all $1 \leq i \leq m-1$: $\mathsf{c}(\langle C_i, C_{i+1} \rangle) = r$

(Sv7) $\mathsf{v}_M(\sim\alpha) = 1$ iff $\mathsf{v}_M(\alpha) = 0$

(Sv8) $\mathsf{v}_M(\alpha \wedge \beta) = 1$ iff $\mathsf{v}_M(\alpha) = \mathsf{v}_M(\beta) = 1$

(Sv9) $\mathsf{v}_M(\alpha \vee \beta) = 1$ iff $\mathsf{v}_M(\alpha) = 1$ or $\mathsf{v}_M(\beta) = 1$

(Sv10) $\mathsf{v}_M(\alpha \supset \beta) = 1$ iff $\mathsf{v}_M(\alpha) = 0$ or $\mathsf{v}_M(\beta) = 1$
(Sv11) $\mathsf{v}_M(\alpha \equiv \beta) = 1$ iff $\mathsf{v}_M(\alpha) = \mathsf{v}_M(\beta)$
(Sv12) $\mathsf{v}_M(\alpha \veebar \beta) = 1$ iff $\mathsf{v}_M(\alpha) \neq \mathsf{v}_M(\beta)$

Definition 8.15 (Truth in a model) *α is true in a model $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$ (abbreviated as $M \models \alpha$) $=_{df}$ $\mathsf{v}_M(\alpha) = 1$*

8.4.3 The proof theory of LIC

The proof theory of **LIC** should meet the following obvious requirements. Firstly, it should determine the behaviour of the classical connectives $\sim$, $\wedge$, ..., of the $\amalg$-relation and of the causal relations $\rightarrow$, $-$ and $\Rightarrow$. Secondly, it should mimic the **IC** algorithm. These requirements are met by the following axiom schemata and inference rules:

(R1) $\alpha, \alpha \supset \beta / \beta$
(A1) The axiom schemata of propositional classical logic
(A2) The (semi-)graphoid axiom schemata
(A3) $\sim(A \Rightarrow A)$
(A4) $A - B \equiv (A \rightarrow B \vee B \rightarrow A)$
(A5) $A \Rightarrow B \equiv (A \rightarrow B \vee \bigvee\{(A \Rightarrow C \wedge C \rightarrow B) \mid C \in V \setminus \{A, B\}\})$
(A6) $A - B \veebar \bigvee\{(A \amalg B \mid \mathbf{Q}) \mid \mathbf{Q} \subseteq V \setminus \{A, B\}\}$
(R2) $A - C, C - B, \sim(A - B) \;/\; (A \rightarrow C \wedge B \rightarrow C) \equiv \bigvee\{(A \amalg B \mid \mathbf{Q}) \wedge \sim(A \amalg B \mid \mathbf{Q} \cup \{C\}) \mid \mathbf{Q} \subset V \setminus \{A, B\}\}$

That the **IC** algorithm is adequately mimicked, can be easily ascertained. The steps in the algorithm are based on a few basic premises. Firstly, no directed cyclic paths are allowed. This is mimicked by **(A3)** and **(A5)**. Secondly, two variables are adjacent if *and only if* no disjoint set screens them off. This is mimicked by **(A6)**. Thirdly, three variables A, B, C form an unshielded collider (i.e. $A \rightarrow C \leftarrow B$ and not $A - B$) if *and only if* there is a $\mathbf{Q} \subset V \setminus \{A, B\}$ such that $(A \amalg B \mid \mathbf{Q})$ and not $(A \amalg B \mid \mathbf{Q} \cup \{C\})$.

8.4.4 Soundness, but no completeness for LIC

The inference rules and axiom schemata listed in section 8.4.3 are not complete with respect to the semantics of section 8.4.2. This follows from theorem 8.10, given that I will not restrict the possible sets of premises to those sets of *IRs* for which the graphoid axioms are complete. Consider a set T of *IRs* for which there is no P such that $(A \amalg_P B \mid \mathbf{Q})$ iff $(A \amalg B \mid \mathbf{Q}) \in Cl(T)$ (where

$Cl(T)$ is the closure of T under the graphoid axioms). From theorem 8.10 it follows that such a T exists. Hence each model $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$ of T verifies at least one $IR \notin Cl(T)$. So all models verify the disjunction of these additional *IRs*, but there is no guarantee that this disjunction is **LIC**-derivable from T.

Conversely, the rules and axioms in section 8.4.3 are sound with respect to the semantics in section 8.4.2.

Theorem 8.16 (Soundness for LIC) *If* $\Gamma \vdash \alpha$, *then* $\Gamma \models \alpha$

Proof. See Appendix 2 (pp. 182–184). ■

8.5 ALIC: the adaptive logic for causal discovery

Regarding the problem of ignorance, **LIC**, the non-adaptive logic presented in section 8.4, is almost on a par with the **IC** algorithm. In case the input or premise set Γ consists of a full list of *IRs*, neither **LIC** nor **IC** will run into difficulty and they will produce the same causal output or consequence set: $A - B \in Cn_{\mathbf{LIC}}(\Gamma)$ iff $A - B \in \Pi(\Gamma)$ or $A \rightarrow B \in \Pi(\Gamma)$ or $B \rightarrow A \in \Pi(\Gamma)$, and $A \rightarrow B \in Cn_{\mathbf{LIC}}(\Gamma)$ iff $A \rightarrow B \in \Pi(\Gamma)$. Likewise, neither **LIC** nor **IC** will lead to satisfactory results in case the input or premise set Γ consists of a partial list of *IRs*.

As I said, **LIC** is *almost* on a par with **IC**. Since any set of **LIC**-wffs (i.e. any $\Gamma \subset \mathcal{W}$) may serve as a premise set, combining observational knowledge with background knowledge, micro-level knowledge, common sense knowledge, etc. poses no technical problems. For example, if background knowledge shows that A causes B, or if common sense dictates that C cannot cause D, since it succeeds D in time, it is straightforward to include $A \rightarrow B$ or $\sim(C \rightarrow D)$ in the premise set.[27]

In this section, I will present **ALIC**, the adaptive logic for causal discovery which properly solves the problem of ignorance. In section 8.3.2, I

[27] In Meek (1995a) a framework is presented which also allows to combine the **IC**-algorithm with background knowledge. However, this framework cannot deal with the problem of ignorance. Moreover, it cannot deal with background knowledge consisting of complex formulas. In Meek (1995a), background knowledge consists of a pair $\mathcal{K} = \langle \mathbf{F}, \mathbf{R} \rangle$ in which $\mathbf{F}$ is the set of directed edges which are forbidden and $\mathbf{R}$ is the set of directed edges which are required. Hence it cannot incorporate background knowledge of the form $A \rightarrow B \vee C \rightarrow B$.

have outlined the basic heuristics governing **ALIC**. In the face of an undecided *CIR*, $(A \amalg B \mid \mathbf{Q})$, we should conceive of it as false, *provided faulty applications of this heuristic can be detected and remedied quickly.* If not, non-experimental causal research would be a rather idle enterprise. In other words, *CIRs* should be presupposed to be false, unless and until proven otherwise. This heuristic does not apply to *UIRs*. In the face of an undecided *UIR*, $(A \amalg B)$, we should conceive of it as true, *provided faulty applications of this heuristic can be detected and remedied quickly.* If not, our scientific models would be gratuitously complex. So *UIRs* should be presupposed to be true, unless and until proven otherwise. These heuristics will be formalized singly in sections 8.5.1 and 8.5.2, where I will present two auxiliary logics $\mathbf{ALIC^{RS}}$ and $\mathbf{ALIC^{At}}$. Then I will combine both auxiliary logics (and their respective heuristics) in one single logic: **ALIC**. First I will describe its proof theory (section 8.5.4; an example of of an **ALIC**-proof can be found in Appendix 1, p. 181). Then I will describe its semantics and some straightforward meta-theoretic results (section 8.5.5). Finally, I will briefly discuss the relation between **ALIC** and *MaxEnt*-based causal discovery (section 8.5.6).

8.5.1 $\mathbf{ALIC^{RS}}$: the Reckless Statistician's account of *CIRs*

The heuristic regarding conditional independence statements may metaphorically be called the heuristic of the Reckless Statistician. If all *CIRs* $(A \amalg B \mid \mathbf{Q})$ are undecided, and if this heuristic is added to **LIC**, then one may infer causation, $A - B$, from correlation, $\sim(A \amalg B)$, contrary to one of the best-known warnings in introductory statistics courses. Any statistician that would apply this heuristic blindly, would deservedly be called reckless. But as I will show in the following sections, the adaptive logical framework affords a way to apply it properly. The Reckless Statistician is tempered.[28]

The heuristic of the Reckless Statistician naturally leads to the following adaptive logic: $\mathbf{ALIC^{RS}} = \langle \mathbf{LIC}, \Omega^{\mathbf{RS}}, \mathit{reliability} \rangle$. The set of abnormalities $\Omega^{\mathbf{RS}}$ contains all *CIRs* regarding pairs of variables (not pairs of sets of variables).

$$\Omega^{\mathbf{RS}} = \{(A \amalg B \mid \mathbf{Q}) \mid A, B \in V, \mathbf{Q} \subseteq V \setminus \{A, B\}\}$$

By adding axiom schema **(A7)** to **LIC**, a suited upper limit logic $\mathbf{ULL^{RS}}$ is obtained.[29]

[28] The same intuition, viz. that $\sim(A \amalg B)$ implies $A - B$, unless and until some set $\mathbf{Q}$ is found such that $(A \amalg B \mid \mathbf{Q})$, lays at the basis of the logic presented in Van Dyck (2004).

[29] The reader can easily check that theorem 8.12 holds for **LIC**, $\Omega^{\mathbf{RS}}$ and $\mathbf{ULL^{RS}}$.

(A7) $\sim(A \amalg B \mid \mathbf{Q})$

The $\mathbf{ULL^{RS}}$-semantics consists of the **LIC**-models for which c is such that

$$\text{if } \mathsf{c}(\langle A, B\rangle) = \mathsf{c}(\langle B, C\rangle) = r, \text{ then } \mathsf{c}(\langle A, C\rangle) = r$$
$$\text{if } \mathsf{c}(\langle A, B\rangle) = \mathsf{c}(\langle B, C\rangle) = l, \text{ then } \mathsf{c}(\langle A, C\rangle) = l$$
$$\text{if } \mathsf{c}(\langle A, B\rangle) = l \text{ and } \mathsf{c}(\langle B, C\rangle) = r, \text{ then } \mathsf{c}(\langle A, C\rangle) \neq n$$

I will not discuss the proof theory, nor the semantics of $\mathbf{ALIC^{RS}}$ in detail as these are not significant for the present purposes and moreover are really straightforward (see Batens, 2007, sections 3 and 4).

8.5.2 $\mathbf{ALIC^{At}}$: the Atomist's account of $UIRs$

Where the first heuristic could be called the heuristic of the Reckless Statistician, the second one, regarding *UIRs*, may metaphorically be called the heuristic of the Atomist. If all *UIRs* $(A \amalg B)$ are undecided, and if this heuristic is added to **LIC**, then one may infer that no causal relations whatsoever exist; i.e., $\sim(A - B)$ for all $A, B \in V$. So the heuristic of the Atomist results in models that verify as little causal relations as possible.[30]

All this naturally leads to the following adaptive logic:

$$\mathbf{ALIC^{At}} = \langle \mathbf{LIC}, \Omega^{\mathbf{At}}, \mathit{reliability}\rangle$$

where $\Omega^{\mathbf{At}}$ contains all negations of *UIRs* regarding pairs of variables.

$$\Omega^{\mathbf{At}} = \{\sim(A \amalg B) \mid A, B \in V\}$$

By adding axiom schema **(A8)** to **LIC**, a suited upper limit logic $\mathbf{ULL^{At}}$ is obtained.[31] The $\mathbf{ULL^{At}}$-semantics consists of the **LIC**-models for which $\mathsf{c}(\langle A, B\rangle) = n$ for all $A, B \in V$.

(A8) $(A \amalg B)$

Again, I will not further discuss the proof theory or the semantics of $\mathbf{ALIC^{At}}$.

[30] An even better label would have been 'the Greedy Statistician', but 'the Atomist' was chosen to avoid entanglement with existing greedy search strategies (cf. Williamson, 2005, 38).

[31] The reader can again easily check that theorem 8.12 holds for **LIC**, $\Omega^{\mathbf{At}}$ and $\mathbf{ULL^{At}}$.

8.5.3 ALIC: outline

ALIC is the result of combining $\mathbf{ALIC}^{\mathbf{RS}}$ and $\mathbf{ALIC}^{\mathbf{At}}$. Its lower limit logic is **LIC** and its strategy is *reliability*. Its set of abnormalities is the union of $\Omega^{\mathbf{RS}}$ and $\Omega^{\mathbf{At}}$:

$$\mathbf{ALIC} = \langle \mathbf{LIC}, \Omega^{\mathbf{RS}} \cup \Omega^{\mathbf{At}}, reliability \rangle$$

As I will show in section 8.5.4, the proof theory of **ALIC** is dynamic. Lines of a proof may be marked (but also unmarked) as the proof continues. Formulas occurring on marked lines are not considered as derived. Notwithstanding this dynamics, the set of **ALIC**-consequences of some premise set Γ is fixed: $Cn_{\mathbf{ALIC}}(\Gamma) = Cn_{\mathbf{ALIC}^{\mathbf{RS}}}(Cn_{\mathbf{ALIC}^{\mathbf{At}}}(\Gamma))$.[32] This stability is reflected in the **ALIC**-semantics (see section 8.5.5).

8.5.4 The proof theory of ALIC

ALIC-proofs are dynamic. Lines in a dynamic proof consist of five elements: (i) a line number k, (ii) a formula α, (iii) the line numbers of the formulas from which α is derived, (iv) the rule by which α is derived, and (v) a *condition* Υ.[33] The condition is a (possibly empty) set of abnormalities. It determines whether α is derived or not. Intuitively, if all members of Υ may be considered as false, then α is derived on line k. Otherwise, line k is marked and α is no longer considered as derived.

The proof theory of **ALIC** consists of three generic deduction rules, and a marking definition. The deduction rules allow one to add a line to the proof. By adding a line, the proof is brought to a next *stage*. The marking definitions determine, at each stage s of the proof, which lines are marked and which are unmarked.

The generic deduction rules are as follows (Γ is a premise set):

PREM If $\alpha \in \Gamma$, one may add a line comprising of the following elements: (i) an appropriate line number, (ii) α, (iii) –, (iv) PREM, and (v) $\emptyset$.

RU If $\beta_1, \ldots, \beta_n \vdash_{\mathbf{LIC}} \alpha$ and each of the β_i occurs in the proof on lines $i_1, \ldots, i_n$ that have conditions $\Upsilon_1, \ldots, \Upsilon_n$ respectively, one may add a line comprising of the following elements: (i) an appropriate line number, (ii) α, (iii) $i_1, \ldots, i_n$, (iv) RU, and (v) $\Upsilon_1 \cup \ldots \cup \Upsilon_n$.

[32] Although the proof theories of $\mathbf{ALIC}^{\mathbf{RS}}$ and of $\mathbf{ALIC}^{\mathbf{At}}$ are also dynamic, $Cn_{\mathbf{ALIC}^{\mathbf{RS}}}(\Gamma)$ and $Cn_{\mathbf{ALIC}^{\mathbf{At}}}(\Gamma)$ are fixed as well.

[33] For a general characterization of the dynamic proof theory of adaptive logics, see Batens, 2007, 227–229. For the proof theory of combined adaptive logics, see Batens, 2001, 61–62.

RC If $\beta_1, \ldots, \beta_n \vdash_{\mathbf{LIC}} \alpha \vee Dab(\Theta)$ (for some $\Theta \subseteq \Omega^{\mathbf{RS}} \cup \Omega^{\mathbf{At}}$) and each of the β_i occurs in the proof on lines $i_1, \ldots, i_n$ that have conditions $\Upsilon_1, \ldots, \Upsilon_n$ respectively, one may add a line comprising of the following elements: (i) an appropriate line number, (ii) α, (iii) $i_1, \ldots, i_n$, (iv) RC, and (v) $\Upsilon_1 \cup \ldots \cup \Upsilon_n \cup \Theta$.

As I stated above, α is derived from Γ at stage s of a proof from Γ iff α occurs at some line k of the proof, and line k is not marked at stage s. Marking is governed by the marking definition, which is applied at each stage of the proof. Let $Dab^{\mathbf{At}}(\Delta)$ denote a Dab-formula with $\Delta \subseteq \Omega^{\mathbf{At}}$; and let $Dab^{\mathbf{RS}}(\Delta)$ denote a Dab-formula with $\Delta \subseteq \Omega^{\mathbf{RS}}$.

Definition 8.17 (minimal $Dab^{\mathbf{At}}$-formula) *$Dab(\Delta)$ is a minimal $Dab^{\mathbf{At}}$-formula at stage s of a proof iff*

(1) $\Delta \subseteq \Omega^{\mathbf{At}}$
(2) *At stage s, $Dab(\Delta)$ is the second element of a line i on the condition $\emptyset$*
(3) *There is no $\Delta' \subset \Delta$ such that $Dab(\Delta')$ satisfies condition (2)*

Definition 8.18 (minimal $Dab^{\mathbf{RS}}$-formula) *$Dab(\Delta)$ is a minimal $Dab^{\mathbf{RS}}$-formula at stage s of a proof iff*

(1') $\Delta \subseteq \Omega^{\mathbf{RS}}$
(2') *At stage s, $Dab(\Delta)$ is the second element of a line i' on the condition $\Theta \subseteq \Omega^{\mathbf{At}}$*
(3') *Line i' is unmarked at stage s*
(4') *There is no $\Delta' \subset \Delta$ such that $Dab(\Delta')$ satisfies conditions (2') and (3')*

Where $Dab(\Delta_{11}), \ldots, Dab(\Delta_{1n})$ are the minimal $Dab^{\mathbf{At}}$-formulas at stage s, $U_s^{\mathbf{At}}(\Gamma) = \Delta_{11} \cup \ldots \cup \Delta_{1n}$ is the set of unreliable **At**-formulas at stage s. Likewise, where $Dab(\Delta_{21}), \ldots, Dab(\Delta_{2m})$ are the minimal $Dab^{\mathbf{RS}}$-formulas at stage s, $U_s^{\mathbf{RS}}(\Gamma) = \Delta_{21} \cup \ldots \cup \Delta_{2m}$ is the set of unreliable **RS**-formulas at stage s.

Now everything is in place to present the marking definition, the application of which proceeds stepwise. At each stage s of the proof, lines are *first* marked/unmarked in view of $U_s^{\mathbf{At}}(\Gamma)$. *Then* lines are marked/unmarked in view of $U_s^{\mathbf{RS}}(\Gamma)$.

Definition 8.19 (Marking for ALIC) *Step 1: line i is marked at stage s iff, where Υ is its condition, $\Upsilon \cap U_s^{\mathbf{At}}(\Gamma) \neq \emptyset$. Step 2: after step 1, line i is marked at stage s iff, where Υ is its condition, $\Upsilon \cap U_s^{\mathbf{RS}}(\Gamma) \neq \emptyset$.*[34]

[34]The marking definition depends on the adaptive strategy, *in casu quo* reliability. Minimal abnormality would result in a different definition (see Batens, 2001, section 6).

Notwithstanding the dynamics of **ALIC**-proofs, the set of **ALIC**-consequences of some premise set Γ is fixed, well-defined and proof-independent (see definition 8.21).

Definition 8.20 *α is* finally derived *from Γ on line i of a proof at stage s iff (i) α is the second element of line i, (ii) line i is not marked at stage s, and (iii) any extension of the proof in which line i is marked may be further extended in such a way that line i is unmarked.*

Definition 8.21 $\Gamma \vdash_{\mathbf{ALIC}} \alpha$ *(α is* finally derivable *from Γ) iff α is finally derived on a line of a proof from Γ.*

Let us briefly return to the worries raised in section 8.3.2, viz. that the suggestions which eventually were incorporated in **ALIC** come across as too rash. I argued that default assumptions for causal inference are used both in scientific practice (cf. Hill, 1965) and in the literature on causal modelling (cf. Williamson, 2005). I also argued that the proof theory of adaptive logics allows us (i) to trace the particular assumptions on which each inference or (intermediate) conclusion is based (cf. the condition of each line in a proof), and (ii) to trace the consequences of the violation of each particular assumption (cf. the marking definition). This gives rise to the following pragmatic picture of causal inference: given a partial list of *IRs*, **ALIC** allows us to derive a set of consequences (giving rise to a *DAG* or a pattern representing our causal beliefs). Some formulas are derived on the empty condition and hence are indubitable relative to the premises. All other consequences may be accepted provisionally. Whether these may serve as a ground for action will depend upon circumstances. Some interventions may be based on relatively slight evidence, while others need fair or even very strong evidence (cf. Hill, 1965, 300).[35] Hence, if α is derived on line i on the non-empty condition Υ_i, we may either decide to take action on the basis of α, or we may find that more information is needed regarding the members of Υ_i. As such, **ALIC**'s dynamic proofs may act as a guide for both scientific research and policy.[36]

[35] I do not use 'intervention' in Woodward's sense here.

[36] This gives rise to some kind of 'reverse falsificationism'. As the acceptability of α depends on the falsehood of Υ_i's members, try to prove their truth (e.g. by gathering new data and performing new conditional independence tests). Accept α in case such proofs fail.

8.5.5 The semantics of ALIC

The semantics of **ALIC** bears out that final derivability is a stable notion.[37] For each premise set Γ, the set of its **ALIC**-models is a subset of its **LIC**-models: $\mathcal{M}^{\mathbf{ALIC}}(\Gamma) \subseteq \mathcal{M}^{\mathbf{LIC}}(\Gamma)$. This subset is obtained by a two-step selection. For any $M \in \mathcal{M}^{\mathbf{LIC}}$, two abnormal parts of M are defined as follows:

Definition 8.22 $Ab^{\mathbf{At}}(M) = \{\omega \in \Omega^{\mathbf{At}} \mid M \models \omega\}$ *and* $Ab^{\mathbf{RS}}(M) = \{\omega \in \Omega^{\mathbf{RS}} \mid M \models \omega\}$

Now the selection runs as follows:

$$\mathcal{M}^0(\Gamma) =_{df} \mathcal{M}^{\mathbf{LIC}}(\Gamma), \text{ where } \mathcal{M}^{\mathbf{LIC}}(\Gamma) = \{M \in \mathcal{M}^{\mathbf{LIC}} \mid M \models \Gamma\}$$

$Dab^{\mathbf{At}}(\Delta)$ is a $Dab^{\mathbf{At}}$-consequence of Γ iff $\Delta \subseteq \Omega^{\mathbf{At}}$ and $Dab^{\mathbf{At}}(\Delta)$ is verified by all $M \in \mathcal{M}^0(\Gamma)$. Where $Dab^{\mathbf{At}}(\Delta_{11}), Dab^{\mathbf{At}}(\Delta_{12}), \ldots$ are the minimal $Dab^{\mathbf{At}}$-consequences of Γ, $U^1(\Gamma) =_{df} \Delta_{11} \cup \Delta_{12} \cup \ldots$

$$\mathcal{M}^1(\Gamma) =_{df} \{M \in \mathcal{M}^0(\Gamma) \mid Ab^{\mathbf{At}}(M) \subseteq U^1(\Gamma)\}$$

$Dab^{\mathbf{RS}}(\Delta)$ is a $Dab^{\mathbf{RS}}$-consequence of Γ iff $\Delta \subseteq \Omega^{\mathbf{RS}}$ and $Dab^{\mathbf{RS}}(\Delta)$ is verified by all $M \in \mathcal{M}^1(\Gamma)$. Where $Dab^{\mathbf{RS}}(\Delta_{21}), Dab^{\mathbf{RS}}(\Delta_{22}), \ldots$ are the minimal $Dab^{\mathbf{RS}}$-consequences of Γ, $U^2(\Gamma) =_{df} \Delta_{21} \cup \Delta_{22} \cup \ldots$

$$\mathcal{M}^2(\Gamma) =_{df} \{M \in \mathcal{M}^1(\Gamma) \mid Ab^{\mathbf{RS}}(M) \subseteq U^2(\Gamma)\}$$

This concludes the selection of the **ALIC**-models of Γ:

$$\mathcal{M}^{\mathbf{ALIC}}(\Gamma) = \mathcal{M}^2(\Gamma)$$

The meta-theoretical properties of adaptive logics are straightforward and have been studied extensively (Batens, 2001, 2004, 2007). If the proof theory of the **LLL** is sound and complete with respect to its semantics, then so is the resulting adaptive logic's proof theory regarding to the adaptive semantics. However, the proof theory of **LIC** is sound, but not complete, with respect to the **LIC**-semantics. Hence the **ALIC**-proof theory is sound, but not complete, with regard to the **ALIC**-semantics.

[37] For a general characterization of the semantics of adaptive logics, see Batens (2007, 229–230). For the semantics of combined adaptive logics, see Batens (2001, 56–57).

8.5.6 ALIC and MaxEnt

Inspection of the **ALIC** semantics gives us a clear view on what it does at the level of probability functions. Recall that an **(A)LIC**-model is a triple $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$, where p is a probability distribution over V. Intuitively, the set of **ALIC**-models of Γ consists of those **LIC**-models of Γ that verify no more abnormalities ($\omega \in \Omega^{RS} \cup \Omega^{At}$) than required by Γ. Whether $M \models \omega$, wholly depends on p. So **ALIC** indirectly selects those probability functions p that satisfy the premises, but verify no more abnormalities than required.

The **ALIC** semantics and the *MaxEnt* principle discussed in section 8.3.2 are somehow similar in that they both provide a mechanism to select one or more members from a set of probability distributions (a credal set) satisfying Γ. However, if this credal set is closed and convex, *MaxEnt* selects one single probability function, whereas the set of **ALIC**-models of Γ will usually not be a singleton (and its members may also differ qua p). Moreover, there is no guarantee that for any $M \in \mathcal{M}^{\mathbf{ALIC}}(\Gamma)$, p has maximal entropy.[38]

Given this semantic difference, **ALIC** and *MaxEnt*-based causal discovery, such as the framework of Williamson (2005, §§5.6–5.7), will typically lead to different results if the premises consist of a partial list of IRs. In the following example it will be seen that **ALIC** tends to output unshielded colliders. Suppose that $V = \{A, B, C, D\}$ and that Γ is the following partial list of IRs:

$$\begin{aligned} \Gamma \;=\; & \{\sim(A \amalg B), \sim(A \amalg C), \sim(A \amalg C \mid D), \sim(A \amalg D), \\ & (A \amalg D \mid C), \sim(C \amalg D), \sim(C \amalg D \mid A)\} \end{aligned}$$

Which are the causal relations that are **ALIC**-derivable from Γ? The object level proof in Appendix 1 shows that $B \rightarrow A, C \rightarrow A, C - D, \sim(B - C), \sim(A - D)$ and $\sim(B - D)$ are derivable.[39] Of these consequences only '$\sim(A - D)$' is derived unconditionally, but all are finally derivable from Γ. The resulting graph is depicted in figure 8.2.

It is easily seen that the framework of Williamson (2005, §§5.6–5.7) results in a different output. One starts by constructing an undirected *constraint*

[38] Entropy is defined as follows:

$$H = -\sum_{v \in [V]} p(v) \log p(v)$$

(Williamson, 2005, 80)

[39] In the interest of readability, the fifth element of some of the lines in that proof will be abbreviated. For example, '$\Upsilon_6 \cup \Upsilon_8$' is shorthand for the union of the fifth element of line 6 and the fifth element of line 8.

graph $\mathcal{G}$ as follows (Williamson, 2005, 86). Take as vertices the set $V = \{A, B, C, D\}$ and include an edge between two variables if and only if they occur in the same premise (constraint). The underlying rationale is the following: if no premise gives information about the relation between two variables X_i and X_j, then we should assume them to be independent. The result is depicted in figure 8.2. A constraint graph $\mathcal{G}$ represents conditional independencies that a maximum entropy function satisfies.

Theorem 8.23 *For all* $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq V$*, if* $\mathbf{Z}$ *separates* $\mathbf{X}$ *from* $\mathbf{Y}$ *in the constraint graph (i.e. if every path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in the constraint graph goes through some vertex in* $\mathbf{Z}$*) then* $(\mathbf{X} \amalg_p \mathbf{Y} \mid \mathbf{Z})$ *for any* p *satisfying the premises which maximizes entropy.*

Proof. See Williamson (2005, 86–87) ■

The converse does not hold. The premises state that $(A \amalg D \mid C)$, but this does not correspond to any separation in our constraint graph $\mathcal{G}$. (This is why the edge between A and D is dotted in figure 8.2.)

The next steps in Williamson's framework are such that a directed acylic graph $\mathcal{H}$ is obtained by adding arrows to $\mathcal{G}$ (Williamson, 2005, 89–90). It is possible that more than one *DAG* $\mathcal{H}$ can be obtained in this way. The *partially directed constraint graph* depicted in figure 8.2 is the pattern representing all possible *DAGs* that can be obtained in our example. (The reader can easily check that this partially directed constraint graph is the result of first triangulating $\mathcal{G}$ and then applying 'Step 1' of Williamson (2005, 90).)

In this partially directed constraint graph there is an edge from A to C, $A \to C$, whereas $C \to A$ is **ALIC**-derivable from Γ. This clearly shows that **ALIC** may lead to different results than *MaxEnt*-based causal discovery. Which framework is most suitable is hard to determine *a priori*. (By '*a priori*' I mean 'apart from concrete application contexts'.) Note that Γ may be the result of observations from an underlying *DAG* in which $C \to A$ is the case, such as world 2 in figure 8.2. In that case **ALIC** produces the best output. But Γ may equally result from observations from world 1. In that case, *MaxEnt*'s output is the best. Which framework is most suitable is underdetermined by Γ.

8.6 Concluding remarks

In this chapter, I claimed that, in scientific practice, the problem of ignorance is ubiquitous, persistent and far-reaching. I also claimed that Pearl's **IC** algorithm cannot be applied in cases of ignorance. Finally, I put forward

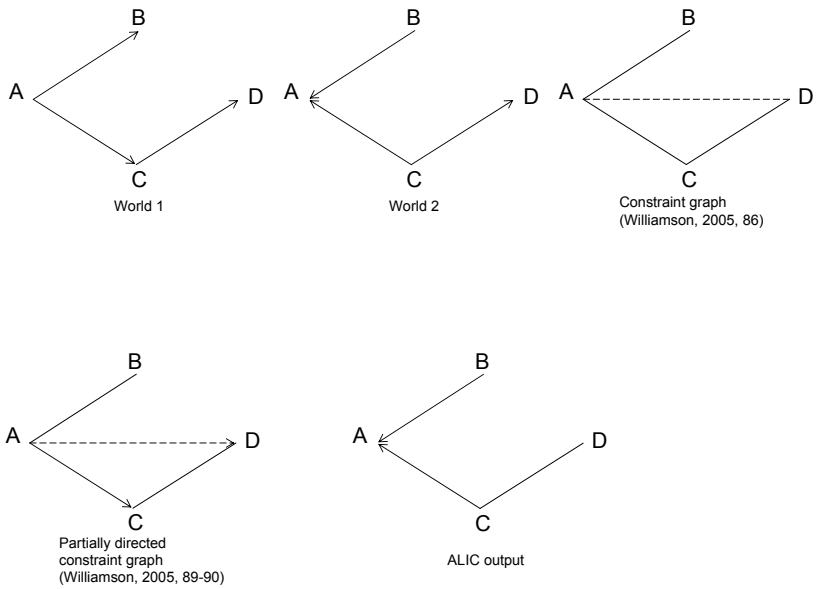

Figure 8.2: **ALIC** *and MaxEnt-based causal discovery.*

an adaptive logic, **ALIC**, which properly solves the problem of ignorance without thereby losing the strong points of **IC**.

ALIC allows one to derive both classical, probabilistic and causal conclusions from any set of probabilistic and/or causal premises. Hence it is greatly apt for combining observational knowledge with background knowledge, common sense knowledge, etc. (The use of reliable background knowledge should not be underrated, cf. the role of cytology in the development of classical genetics.) What is more important: **ALIC** assigns an adequate truth value to all undecided *UIRs* and *CIRs* (i.e. to all *IRs* that are undecided even in the light of the available background knowledge, common sense knowledge, etc). This assignment is based on two rationales: firstly, that scientific models should not be overly complex, and secondly, that correlation is a useful (but not infallible) indicator of causation. But what is most important, if the interpretation of an undecided *IR* turns out to be fallacious (e.g. in the light of new premises), **ALIC** adapts itself to the premises and faulty applications of the above rationales are remedied adequately.

Appendix 1: an ALIC-proof

Part I: to derive $A - C$ (conditionally), $A - B$ (conditionally) and $\sim(A - D)$ (unconditionally)

1.	$\sim(A \amalg C)$	–	PREM	$\{\}$
2.	$\sim(A \amalg C \mid D)$	–	PREM	$\{\}$
3.	$(A - C) \veebar ((A \amalg C) \vee (A \amalg C \mid B) \vee (A \amalg C \mid D) \vee (A \amalg C \mid B, D))$	–	RU	$\{\}$
4.	$(A - C) \veebar ((A \amalg C \mid B) \vee (A \amalg C \mid B, D))$	1, 2, 3	RU	$\{\}$
5.	$(A - C) \vee (A \amalg C \mid B) \vee (A \amalg C \mid B, D)$	4	RU	$\{\}$
6.	$A - C$	5	RC	$\{(A \amalg C \mid B), (A \amalg C \mid B, D)\}$
7.	$\sim(A \amalg B)$	–	PREM	$\{\}$
8.	$A - B$	7	RC	$\{(A \amalg B \mid C), (A \amalg B \mid D), (A \amalg B \mid C, D)\}$
9.	$(A \amalg D \mid C)$	–	PREM	$\{\}$
10.	$\sim(A - D)$	9	RU	$\{\}$

Part II: to derive $\sim(B - C)$ and $\sim(B - D)$ (conditionally)

11.	$(B \amalg C) \supset \sim(B - C)$	–	RU	$\{\}$
12.	$\sim(B \amalg C) \vee \sim(B - C)$	11	RU	$\{\}$
13.	$\sim(B - C)$	12	RC	$\{\sim(B \amalg C)\}$
14.	$(B \amalg D) \supset \sim(B - D)$	–	RU	$\{\}$
15.	$\sim(B \amalg D) \vee \sim(B - D)$	14	RU	$\{\}$
16.	$\sim(B - D)$	15	RC	$\{\sim(B \amalg D)\}$

Part III: to derive $C - D$ (conditionally)

17.	$\sim(C \amalg D)$	–	PREM	$\{\}$
18.	$\sim(C \amalg D \mid A)$	–	PREM	$\{\}$
19.	$C - D$	17, 18	RC	$\{(C \amalg D \mid B), (C \amalg D \mid A, B)\}$

Part IV: to derive $B \to A \wedge C \to A$ (conditionally)

20.	$A - B \wedge A - C$	6, 8	RU	$\Upsilon_6 \cup \Upsilon_8$
21.	$(B \to A \wedge C \to A) \equiv (((B \amalg C) \wedge \sim(B \amalg C \mid A)) \vee ((B \amalg C \mid D) \wedge \sim(B \amalg C \mid A, D)))$	20	RU	$\Upsilon_6 \cup \Upsilon_8$
22.	$\sim(((B \amalg C) \wedge \sim(B \amalg C \mid A)) \vee ((B \amalg C \mid D) \wedge \sim(B \amalg C \mid A, D))) \vee (B \to A \wedge C \to A)$	21	RU	$\Upsilon_6 \cup \Upsilon_8$
23.	$(B \to A \wedge C \to A) \vee \sim((B \amalg C) \wedge \sim(B \amalg C \mid A))$	22	RU	$\Upsilon_6 \cup \Upsilon_8$
24.	$(B \to A \wedge C \to A) \vee \sim(B \amalg C) \vee (B \amalg C \mid A)$	23	RU	$\Upsilon_6 \cup \Upsilon_8$
25.	$B \to A \wedge C \to A$	24	RC	$\Upsilon_6 \cup \Upsilon_8 \cup \{\sim(B \amalg C), (B \amalg C \mid A)\}$

Appendix 2: soundness for LIC

Theorem 8.16 (Soundness for LIC) *If* $\Gamma \vdash \alpha$, *then* $\Gamma \models \alpha$

Proof. Consider an **LIC**-proof of α from Γ. Each line of this proof either contains a premise, or an instance of **(A1)**-**(A6)**, or a formula which is derived from previous formulas by either **(R1)** or **(R2)**. I will show that all **LIC**-models of Γ verify α, so that $\Gamma \models \alpha$.

Consider an **LIC**-model $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$ such that v_M verifies all members of Γ. I will show that for each line i in the proof, if β is the formula derived on line i, then $v_M(\beta) = 1$. It follows that $v_M(\alpha) = 1$.

(PREM) If β is a premise, then $\beta \in \Gamma$ and, by hypothesis, $v_M(\beta) = 1$.

(R1) If $v_M(\beta') = v_M(\beta' \supset \beta) = 1$, then $v_M(\beta) = 1$ (by **(Sv10)**).

(A1) By **(Sv7)**-**(Sv12)**, $v_M(\beta) = 1$ if β is an instance of **(A1)** – i.e. an axiom of propositional classical logic.

(A2) Likewise, $v_M(\beta) = 1$ if β is an instance of **(A2)** – i.e. a (semi-)graphoid axiom. I will prove this for the case of **(G5)**. The proofs for the other (semi-)graphoid axioms are left to the reader.
Suppose that $v_M((\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{Q} \cup \mathbf{C})) = 1$ (*) and that $v_M((\mathbf{A} \perp\!\!\!\perp \mathbf{C} \mid \mathbf{Q})) = 1$ (**). From (*) it follows by **(Sv3)** that for all relevant values of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{Q}$,[40]

$$\frac{\mathsf{p}(\mathbf{A} = \mathbf{a} \wedge \mathbf{B} = \mathbf{b} \wedge \mathbf{Q} = \mathbf{q} \wedge \mathbf{C} = \mathbf{c})}{\mathsf{p}(\mathbf{B} = \mathbf{b} \wedge \mathbf{Q} = \mathbf{q} \wedge \mathbf{C} = \mathbf{c})} = \frac{\mathsf{p}(\mathbf{A} = \mathbf{a} \wedge \mathbf{Q} = \mathbf{q} \wedge \mathbf{C} = \mathbf{c})}{\mathsf{p}(\mathbf{Q} = \mathbf{q} \wedge \mathbf{C} = \mathbf{c})}$$

and from (**) it follows by **(Sv3)** that for all relevant values of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{Q}$,

$$\frac{\mathsf{p}(\mathbf{A} = \mathbf{a} \wedge \mathbf{Q} = \mathbf{q} \wedge \mathbf{C} = \mathbf{c})}{\mathsf{p}(\mathbf{Q} = \mathbf{q} \wedge \mathbf{C} = \mathbf{c})} = \frac{\mathsf{p}(\mathbf{A} = \mathbf{a} \wedge \mathbf{Q} = \mathbf{q})}{\mathsf{p}(\mathbf{Q} = \mathbf{q})}$$

From these equations it follows that

$$\frac{\mathsf{p}(\mathbf{A} = \mathbf{a} \wedge \mathbf{B} = \mathbf{b} \wedge \mathbf{Q} = \mathbf{q} \wedge \mathbf{C} = \mathbf{c})}{\mathsf{p}(\mathbf{B} = \mathbf{b} \wedge \mathbf{Q} = \mathbf{q} \wedge \mathbf{C} = \mathbf{c})} = \frac{\mathsf{p}(\mathbf{A} = \mathbf{a} \wedge \mathbf{Q} = \mathbf{q})}{\mathsf{p}(\mathbf{Q} = \mathbf{q})}$$

But then, by **(Sv3)**, $v_M((\mathbf{A} \perp\!\!\!\perp \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q})) = 1$.

(A3, A4, A5) If β is an instance of **(A3)**, **(A4)** or **(A5)**, then $v_M(\beta) = 1$. The proofs for these cases are straightforward and left to the reader.

[40] By 'relevant values' I mean those values for which the conditional probabilities in question are defined.

(A6) If β is an instance of **(A6)**, then $v_M(\beta) = 1$. For the first part, suppose that $v_M(A - B) = 1$. It has to be shown that for all $\mathbf{Q} \subseteq V \setminus \{A, B\}$, $v_M((A \amalg B \mid \mathbf{Q})) = 0$. So suppose that for some such $\mathbf{Q}^*$, $v_M((A \amalg B \mid \mathbf{Q}^*)) = 1$. By **(Sv3)** it follows that for all relevant values of A, B and $\mathbf{Q}^*$,

$$\frac{\mathsf{p}(A = a \wedge B = b \wedge \mathbf{Q}^* = \mathbf{q}^*)}{\mathsf{p}(B = b \wedge \mathbf{Q}^* = \mathbf{q}^*)} = \frac{\mathsf{p}(A = a \wedge \mathbf{Q}^* = \mathbf{q}^*)}{\mathsf{p}(\mathbf{Q}^* = \mathbf{q}^*)}$$

But then, by **(Sp9)**, $\mathsf{c}(\langle A, B\rangle) = n$. Hence, by **(Sv5)**, $v_M(A - B) = 0$, which contradicts our supposition.

For the second part, suppose that $v_M(A - B) = 0$. It has to be shown that for some $\mathbf{Q}^* \subseteq V \setminus \{A, B\}$, $v_M((A \amalg B \mid \mathbf{Q}^*)) = 1$ (i.e. that $\mathbf{Q}^*$ blocks all paths between A and B – cf. **(Sp9)** and **(Sv3)**). Define $\mathbf{Q}^*$ as follows (cf. Verma and Pearl, 1992, lemma 3.1):

$$\mathbf{Q}^* = \{X \mid A \neq X \neq B \text{ and either } X \Rightarrow A \text{ or } X \Rightarrow B\}$$

Suppose that some path $\mathcal{P} = \langle X_1, \ldots, X_n\rangle$ (with $n \geq 3$, $X_1 = A$, and $X_n = B$) is not blocked by $\mathbf{Q}^*$. This means that for this path, $\mathbf{Q}^*$ satisfies none of the conditions (a), (b) and (c) of **(Sp9)**. Hence, by the definition of $\mathbf{Q}^*$, for all X_i $(2 \leq i \leq n-1)$:[41]

($a^\sim$) if $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$ or $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$, then $\sim(X_i \Rightarrow A)$ and $\sim(X_i \Rightarrow B)$

($b^\sim$) if $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, then $\sim(X_i \Rightarrow A)$ and $\sim(X_i \Rightarrow B)$

($c^\sim$) if $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, then $(X_i \Rightarrow A)$ or $(X_i \Rightarrow B)$

Let us now consider all adjacency relations in $\mathcal{P}$, starting with X_1. If $X_1 \leftarrow X_2$, then $X_2 \in \mathbf{Q}^*$. But this contradicts either ($a^\sim$) (if $X_2 \leftarrow X_3$) or ($b^\sim$) (if $X_2 \rightarrow X_3$). So $X_1 \rightarrow X_2$ (§).

What about X_2 and X_3? If $X_2 \leftarrow X_3$, then $X_2 \Rightarrow B$ (by (§), ($c^\sim$) and acylicity). But then $X_3 \in \mathbf{Q}^*$, which contradicts either ($a^\sim$) (if $X_3 \leftarrow X_4$) or ($b^\sim$) (if $X_3 \rightarrow X_4$). It follows that $X_2 \rightarrow X_3$ (§§). If $n = 3$, $X_2 \in \mathbf{Q}^*$, which contradicts ($a^\sim$), so $n \geq 4$.

So what about X_3 and X_4? By the same reasoning, (§) and (§§) together imply that $X_3 \rightarrow X_4$ (§§§), and hence that $n \geq 5$. But this implies that $X_4 \rightarrow X_5$ (§§§§) and $n \geq 6$. etc. etc. So $\mathcal{P}$ consists of an infinite number of nodes, which is impossible.

To conclude, $\mathbf{Q}^*$ blocks all paths between A and B and $v_M((A \amalg B \mid \mathbf{Q}^*)) = 1$.

[41] Note that for all X_i in question, $A \neq X_i \neq B$.

(R2) If β is derived by means of **(R2)**, and if $v_M(\beta') = 1$ for all the β' used for this derivation, then $v_M(\beta) = 1$. Suppose that $v_M(A - C) = v_M(C - B) = v_M(\sim(A - B)) = 1$. It has to be shown that $v_M(A \rightarrow C \wedge B \rightarrow C) = 1$ implies that for some $\mathbf{Q}$, $v_M((A \amalg B \mid \mathbf{Q}) \wedge \sim(A \amalg B \mid \mathbf{Q} \cup \{C\})) = 1$, *and vice versa*.
For the first direction, suppose that $v_M(A \rightarrow C \wedge B \rightarrow C) = 1$. Given that $v_M(\sim(A-B)) = 1$, there is some $\mathbf{Q}^*$ such that $v_M((A \amalg B \mid \mathbf{Q}^*)) = 1$ (see the soundness proof for **(A6)**). Now suppose that $v_M((A \amalg B \mid \mathbf{Q}^* \cup \{C\})) = 1$ (†). Since $v_M(A \rightarrow C \wedge B \rightarrow C) = 1$, the triple $\langle A, C, B\rangle$ is a path between A and B. By (†), $\langle A, C, B\rangle$ must satisfy one of the conditions of **(Sp9)**. Trivially, it cannot satisfy (a) or (b). By condition (c), $C \notin \mathbf{Q}^* \cup \{C\}$, which is impossible. Hence, contra (†), $v_M((A \amalg B \mid \mathbf{Q}^* \cup \{C\})) = 0$ and, by **(Sv7)**, $v_M(\sim(A \amalg B \mid \mathbf{Q}^* \cup \{C\})) = 1$. So, by **(Sv8)**, $v_M((A \amalg B \mid \mathbf{Q}^*) \wedge \sim(A \amalg B \mid \mathbf{Q}^* \cup \{C\})) = 1$.
The proof for the reverse direction is left to the reader.

■

Chapter 9

Galton's Blinding Glasses. Modern Statistics Hiding Causal Structure in Early Theories of Inheritance

In the previous chapters I have dealt with the discovery of pragmatic regularities (more specifically with causal discovery) from two perspectives. Instead of trying to give an overview of the plethora of topics relating to causal discovery, I focussed on two themes. First I touched upon the superiority of experimental over non-experimental studies as regards causal discovery and showed that this was relevant in the history of classical genetics. Then I turned away from classical genetics to discuss algorithms for causal discovery from non-experimental data. Again, I did not give an elaborate overview of all such algorithms. I focussed on Pearl's **IC** algorithm, showed that it faced a serious problem (the problem of ignorance), and presented an adaptive logic, **ALIC**, that properly handles that problem.

The present chapter[1] is also dedicated to causal discovery. It serves two different goals. The major goal is to show that notwithstanding the fact that contemporary statistical techniques are highly valuable and widely used for causal discovery, we should be on our guard for their possibly biasing role. To that end I will exploit another case study from the history of science, viz. Sir Francis Galton's theory of Ancestral Inheritance. (We will see that the Mendelians did not have the monopoly of the concept of 'law' in the context

[1]This chapter is based on my paper "Galton's blinding glasses. Modern statistics hiding causal structure in early theories of inheritance", which appeared in F. Russo and J. Williamson (eds.), *Causality and Probability in the Sciences*, Texts in Philosophy series, pp. 242–262. College Publications, London. (Leuridan, 2007a).

of inheritance. Several different generalizations or equations were called laws by Galton and the later biometricians. Some of these were clearly causal P-laws (such as the law of ancestral heredity), others (such as the law of filial regression) were clearly non-causal.)

The theory of ancestral inheritance rivalled the theory of classical genetics for a long time. This will allow me to pursue a second aim. I will use the concepts of chapters 3 and 4 to state precisely how Galton's theory differed from that of classical genetics. Their respective models had a different causal structure (they were not isomorphic) and thus belonged to different causal schemes.

9.1 The problem: probability and statistics as a tool for discovering causal patterns

Mendel published the results of his genetic crosses on *Pisum* and their theoretic explanation in 1865/1866. His works again came to the fore from 1900 onwards. In the intervening years, the then influential scientist and statistician Francis Galton tried to analyse hereditary phenomena statistically. He discovered interesting phenomenological regularities and posited a theoretical or causal mechanism of hereditary transmission to explain them. But, as I will argue, his causal ideas were perniciously biased by the statistical techniques he used.

As we have seen in chapter 8, it is now commonplace to attribute to probability theory and statistical inference a central place in the philosophy of causality. The algorithms presented in Pearl (2000), Spirtes et al. (2000), Neapolitan (2004), and Williamson (2005) make it possible to discover causal relations on the basis of knowledge of (conditional) (in)dependence relations between variables, together with some graph theoretical theorems and a set of assumptions. Their theories are tightly linked with contemporary statistical techniques such as structural equation modelling (SEM).[2] To put it metaphorically: *probability and statistics are viewed as glasses through which we can see or detect causal relations.*

This raises a problem. If I succeed in showing that Galton's knowledge of statistics impeded him to successfully develop a biological theory of inheritance (and thus acted as *blinding glasses*), it could be called into question whether contemporary statistical techniques are neutral with respect to their

[2]'Structural equation modelling' refers to many related techniques. Other, more or less equivalent labels are: 'covariance structure analysis', 'covariance structure modelling', and 'analysis of covariance structures'. The term 'causal modelling' is considered somewhat dated. (Kline, 2005, 9)

domain of application (i.e. with respect to the theory being developed or tested).

So here's the plan. In sections 9.2 and 9.3, I will discuss an important methodological difference between Mendel and Galton (the role of statistics in their scientific research) and I will briefly recapitulate the main characteristics of Mendel's theory. Then, in section 9.4, I will show how probability and statistics generated two *explananda* which in their turn generated constraints for any theory of inheritance. In sections 9.4.1 and 9.5, I will present Galton's *explanans* (his theory of heredity) and discuss the differences with Mendel's view on the matter. After parrying two possible counter-arguments to my reasoning (9.6.1, 9.6.2), I will summarize what this case-study shows us regarding the neutrality or non-neutrality of statistics in the work of Galton. In section 9.7 I will redeem my promises. I will analyse the consequences for contemporary statistical techniques like SEM and show that these may also have a blinding influence. Finally, in section 9.8 I will show that Galton's theory of inheritance strongly differed from the classical framework. Their causal schemes were strongly different. In other words, Galton's causal or credal nets are not isomorphic (let alone value-isomorphic or distribution-identical) to any of the causal or credal nets of classical genetics.

9.2 Gregor Mendel and Francis Galton: two different scientists

In the second half of the 19th century Francis Galton studied the processes of heredity. He started his research by considering 'hereditary genius' (Galton, 1869), but soon he turned to more easily observable characteristics. In *Natural Inheritance* (Galton, 1889) he bundled the results of more than twenty years of research on this topic. One of the most interesting aspects of Galton's work was the fact that he used and developed several modern statistical techniques, some of which are still used today (e.g. linear regression, which was mathematically developed by Karl Pearson and others).

At the time Galton started to work on the problem of heredity, Mendel had just finished his series of crosses with pea plants (*P. sativum*). In 1866 he wrote his *Versuche über Pflanzenhybriden* (Mendel, 1933) in which he meticulously presented his theory of inheritance. As this paper met little or no response in the biological community at the time, his ideas remained silent until they were rediscovered by Carl Correns (1900) and Hugo de Vries (1900a).[3] In the meantime, Galton independently developed his biometric

[3]For a nice demystification of this mysterious rediscovery and of the preceding neglect,

theory of ancestral inheritance. After 1900, these theories would become vehement rivals and it would take several years or even decades before the dispute between the Mendelians (e.g. Bateson, 1902) and the biometricians (e.g. Karl Pearson, but also Weldon, 1902) was settled.[4]

Contrary to Galton, Mendel made little use of statistical theory. Of course, he was trained mathematically and statistically (cf. footnote 1 of chapter 5), and he recognized the need for empirical data. But he was not trained in the 'new' tradition of Quetelet, Galton, etc. He inferred in a rather intuitive way from particular observations to general regularities (see section 5.1). So, given the current dominant role of statistics in the special sciences, shouldn't we expect Galton to have found the most 'true' regularities? It seems not. While Mendel is still considered the founding father of genetics, Galton's name is now only associated with dubious disciplines such as phrenology and eugenics.

9.3 Mendel's theory of inheritance

All of the phenotypic characteristics that Mendel observed in *P. sativum* were qualitative and discrete. He studied seven pairs of contrasting traits, such as seed shape (round or wrinkled) and stem length (tall or dwarf). His *explanandum* consisted of some very straightforward empirical regularities (phenotypic distributions). The task was to explain why after crossing short plants with true-breeding tall plants, all the off-spring in the first filial generation was tall (see cross 1 of section 4.5); or why, after selfing this off-spring, 75% of the second filial generation was tall, while 25% was short (see cross 2 of section 4.5). These explananda are visualized in figure 9.1 (compare this to figure 9.2).

Mendel's *explanans* is based on a causal mechanism invoking material bearers of hereditary traits. In each pair of contrasting traits, one is dominant and the other is recessive (round is dominant to wrinkled, tall is dominant to dwarf). These traits are caused and carried over from parental plants to filial plants by *unit factors* ('*Factoren*', see Mendel, 1933, 24). A unit factor either codes for the dominant trait or for the recessive trait. Factors occur *pairwise* in the individual pea plants, but singly in the gametes.[5] Roughly half of the

see Meijer (1983, sections 1 and 2) and also Darden (1991, section 4.3.).

[4]Two main factors brought the dispute to an end. Firstly, the theory of classical genetics gained more and more independent evidence from e.g. cytology. Secondly, Sir R.A. Fisher, the well-known statistician, proved that Galton's explananda could be explained within the Mendelian framework (Fisher, 1918).

[5]At least, this is what classical Mendelism has taught us. However, there is every

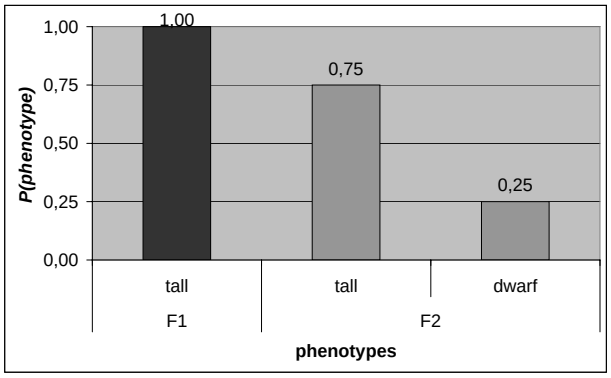

Figure 9.1: *Mendel's explananda illustrated. The F*1*-generation resulted from crossing dwarf plants with true-breeding tall plants. The F*2*-generation resulted from selfing F1.*

gametes carries (a copy of) the first factor, the other half carries (a copy of) the second factor (later this was called the '*principle of segregation*'). As emphasized by William Bateson in the beginning of the 20th century, no gamete carries an intermediate factor. This is dubbed '*the purity of the germ-cells*' in Bateson (1902, 108).[6]

A last point to be noted with respect to Mendel's theory is that the 'genotype' of a pea plant is *screened off* or *d-separated* (see figure 4.1) from the ancestral genotypes by the set of gametes that produced it. In figure 4.1,

$$(GT_3 \amalg \{GT_1, GT_2\} \mid \{GC_1, GC_2\}) \tag{9.1}$$

Once it is established what unit factors are carried by the germinal cell and the pollen cell, the origin of these gametes plays no further role. Remote ancestry has no influence, conditional on the gametes.

To summarize, these are the main features of Mendel's theory that we should bear in mind in the following sections and which we should contrast with Galton's. Traits are grouped in pairs, in which one is dominant, the other recessive. They are caused and transmitted by unit factors that occur pairwise in individual organisms. Each gamete, however, contains only one factor,[7] according to the principle of segregation. Moreover, the gametes

indication that according to Mendel factors occur pairwise in heterozygotes, but singly in homozygotes, since he considered them as discrete, uncountable fluids (Meijer, 1983).

[6]I will not discuss the 'principle of independent assortment', as Galton did not explicitly treat multi-hybrid crosses.

[7]Each gamete contains only one factor *of each pair of factors*. Recall, however, that I only discuss monohybrid crosses here.

are pure (they never carry intermediate factors), and they *screen off* the resulting individual from its ancestry.

9.4 Statistics generating two *explananda* for Galton's theory of heredity

9.4.1 Preview of Galton's theory of *Ancestral Inheritance*

Like Mendel, Galton invoked bearers of hereditary traits to set up a causal mechanism of heredity.[8] He called them 'elements' or 'particles'.[9] But contrary to Mendel's unit factors, these did not occur in pairs. In each individual organism, an *indefinite* or *incalculable number of elements* responsible for the same phenotypic trait is present. (If we want to assess the merits of Galton's work, we should bear in mind that cytological constraints were rather poor at Galton's time; see Darden, 1991, chapter 7.)

Mendel had a relatively clear view on the transmission of unit factors and on the possible genetic make-up of the gametes.[10] By contrast, the genetic make-up of gametes or individuals was never treated concretely in Galton's texts and the gametes played no inferential or predictive role in his theory.[11] Although he stated, in the beginning of his *Natural Inheritance*, that "there is no direct hereditary relation between the personal parents and the personal child" and that "the main line of hereditary connection unites the sets of elements out of which the personal parents had been evolved with

[8] I have found no explicit discussion of the concept of 'causation' either by Mendel or by Galton. Both their writings, however, reveal a 'mechanistic' view on causation.

Note that in this respect they strongly differed from e.g. Pearson who had a Machian view on causation in which, first, causation is defined as perfect correlation and in which, secondly, the existence of causal relations in the empirical world is excluded *a priori* (Pearson (1900, chapter IV) or (1911, chapter V)).

[9] Galton also used the word 'element' to refer to the phenotypic traits themselves, instead of the particles that caused them (see section 9.5.1).

[10] See 'The reproductive cells of hybrids' (in Stern and Sherwood, 1966, 23–32). See also the credal net $\mathfrak{B}^1$ and the generic credal net $\mathfrak{B}^\alpha$ in sections 4.5 and 4.7.

[11] In a later paper, Galton would refer to the germ-cells in his presentation of the law of ancestral heredity (for the definition of this law, see section 9.5.1):

> Now this law is strictly consonant with the observed binary subdivisions of the germ cells, and the concomitant extrusion and loss of one-half of the several contributions from each of the two parents to the germ-cell of the offspring. (Galton, 1897, 403)

the set out of which the personal child was evolved" (both are quoted from Galton, 1889, 19), he directly predicted the traits of issue from the traits of its ancestors. The ancestors not only included the parental generation, but also the grand-parents, great-grand-parents, ... Galton was convinced that all ancestry may in principle have an influence on the set of elements from which the organism is built, and thus on its phenotype (cf. infra, p. 197). The influence is neither screened off by the parents, nor by their gametes. Therefore, Galton's theory can be called the *theory of ancestral inheritance*.

This suffices to show that the causal mechanisms proposed by Mendel and by Galton were very different (see also section 9.8). In the following sections, I will show how this difference can (partly) be explained by laying bare the role played by Galton's statistical knowledge. (I write 'partly', since his views were also influenced by the theories of e.g. Weismann and Darwin.) His statistics generated two *explananda* which in their turn imposed influential constraints on any would-be *explanans*.[12]

9.4.2 The first *explanandum*: the normal distribution

Contrary to Mendel, Galton mostly studied continuous traits. In his *Natural Inheritance*, he paid a lot of attention to the schemes of distribution and the schemes of frequency of e.g. human strength, stature, span of arms, weight, breathing capacity, etc. (Galton, 1889, 35–50 and 200). All these characteristics, Galton noted, are normally distributed.

In the 19th century, the normal distribution played a very important role, not only in astronomy, but also in the biological and the social sciences. Adolphe Quetelet, the Belgian statistician and sociologist, discovered that measurements of e.g. human stature, birth ratios and crime rates were all normally distributed. The term 'Quetelismus' refers, then, to the exaggeration of the dominance of the normal distribution, i.e. to the view that "*all* naturally occurring distributions of properly collected and sorted data follow a normal curve" (Stigler, 1986, 201, my emphasis; see also 203–205). Galton explicitly acknowledged the influence of Quetelet and stated that the latter introduced the idea that the Law of Error (i.e. the normal distribution) might be applicable to human measures (e.g. Galton, 1877, 493 and 1889, 54–55).

Galton knew several ways to represent distributions of data. One way was to represent them graphically. Another way was to cite a series of eleven percentiles (the 5th, 10th, 20th, ..., 80th, 90th and 95th). But the most economical method, applicable in case the data were normally distributed (as seemed mostly the case), was to cite just two numbers: `M` and `Q`. `M` was

[12]In fact, Galton's statistics generated at least three explananda (see footnote 30).

the mean or median. Q he called the 'Prob. Deviation' and it was defined as one half of the interquartile range: $\mathtt{Q} = \frac{1}{2}(\mathtt{Q}_2 - \mathtt{Q}_1)$, where $\mathtt{Q}_2$ and $\mathtt{Q}_1$ are the third and the first quartile respectively. Q conveyed the dispersion of the distribution and thus played a role similar to the standard deviation σ.[13] Once M and Q were known, all percentiles could be calculated and the scheme of distribution could be drawn.

The normal distribution of Human Stature played a tremendously important role in Galton's biometrical work. He wrote:

> In particular, the agreement of the Curve of Stature with the Normal Curve is very fair, and forms a mainstay of my inquiry into the laws of Natural Inheritance. (Galton, 1889, 57)

Several sets of data, collected by Galton himself, indicated that the median male stature P was 68.25 inch (Galton used the symbol P to refer to the median in the context of Stature) and that $\mathtt{Q} = 1.7$ inch.

Obviously, the normal distribution generated a major constraint on Galton's theory of heredity. Whatever mechanism one was to propose to explain the processes of inheritance, it had to be able to explain why inherited traits are normally distributed. As Galton stated,

> The conclusion is of the greatest importance to our problem. It is, that the processes of heredity must work harmoniously with the law of deviation, and be themselves in some sense conformable to it. (Galton, 1877, 512)

What causes variables to be normally distributed? In 1810, Pierre Simon Laplace first introduced what was later to be called the *'Hypothesis of Elementary Errors'*: that the joint action of a multitude of independent 'errors' produces a normal distribution (Stigler, 1986, 201–202). Laplace's main topic of interest was the distribution of astronomical observations,[14] but his ideas forcefully influenced Quetelet and later also Galton. Galton wrote:

> The Law of Error finds a footing wherever the individual peculiarities are wholly due to the combined influence of a multitude of "accidents" [...]. (Galton, 1889, 55)

So the constraint imposed by the normal distribution was the following: Galton had to introduce what he called a '*variety of petty influences*' in his biological theory (Galton, 1889, 16–17).

[13] The standard deviation σ is by definition larger that Q, since $\mathtt{M} \pm \sigma$ includes about 68% of the observations, while $\mathtt{M} \pm \mathtt{Q}$ includes only 50% of them.

[14] If every measurement is the aggregate of many independent component measurements, each of them subject to small errors, the normal distribution of astronomical data is explained.

9.4.3 The second *explanandum*: regression towards the mean

Perhaps the most important one of Galton's contributions to modern statistical theory was his concept of '*regression towards the mean*'. It would later result in the theory of linear regression (elaborated by his protégé Karl Pearson). What exactly did Galton mean by this regression?

In the 1860's, he first tackled the topic of inheritance by studying human genius or talent and the way it was distributed within families. In *Hereditary Genius* (1869), Galton observed that, generally speaking, the relatives of gifted men (such as Johann Sebastian Bach or Jacob Bernouilli) are gifted too, but less so. And the more remote a relative is, the less he is talented (Stigler, 2002, 176–177). In other words, Galton found a 'regression towards mediocrity'.

Of course, genius or talent is not easily observable, let alone measurable, but other characteristics are.[15] In 1884, Galton gathered hundreds or even thousands of records about the human population, called the *R.F.F. Data* (*Record of Family Faculties*). They comprised information about the Stature, Eye-colour, Temper, Artistic Faculty, ... of whole families (spanning several generations). The major part of *Natural Inheritance* concerned Human Stature, and the main research question was whether it is inheritable or not, i.e. whether the offspring of tall people is – on average – tall, or not.[16] But what do we mean by 'tall'?

As I stated earlier, Galton knew that male stature was normally distributed with $\mathtt{P} = 68\frac{1}{4}$ inch and $\mathtt{Q} = 1.7$ inch (see page 191). Women are slightly smaller than men, but this difference disappears if their statures are multiplied or 'transmuted' by 1.08. These figures suggested a very straightforward criterion for 'tallness'. People are tall *iff* they have a stature larger than $\mathtt{P}$. Human stature could be written as the sum of two components:

$$\mathtt{Stature} = \mathtt{P} \pm \mathtt{D}, \tag{9.2}$$

in which $\mathtt{D}$ is the individual's deviation from the mean (Galton, 1889, 51–52 and Chapter VII). So the sign and size of $\mathtt{D}$ indicated whether and to what extent an individual was tall or small and Galton's research question could

[15] Galton took great pains, however, to classify men according to their talent, both on the basis of their 'reputation' and on the basis of their 'natural ability' (Galton, 1869, 37–38).

[16] The *R.F.F. Data* contained the statures of 205 pairs of parents and their 930 adult children. Other data sets also reported on Stature: the *Special Data* covered about 783 brothers from 295 families and the *Measures at the Anthropometric Laboratory* consisted of about 10,000 data (Galton, 1889, 71–82).

be rephrased as follows: does the issue of parents with a large D itself also have a large D?

The stature of the issue of unlike parents does not depend on the specific statures of the father and of the mother, as Galton's data revealed. It depends only on their *average* stature. Therefore, Galton introduced the concept of the 'Mid-Parent', which is defined as "an ideal person of composite sex, whose Stature is half way between the Stature of the father and the transmuted Stature of the mother." (Galton, 1889, 87) The Mid-Parental Statures are normally distributed with $\mathtt{P} = 68\frac{1}{4}$ inch and $\mathtt{Q} = 1.21$ inch.[17]

$$\mathtt{Stature\ Mid\text{-}Parent} = \frac{\mathtt{Stat.Father + Transm.Stat.Mother}}{2}, \tag{9.3}$$

(Maybe the reader cannot help sniggering here. We should recall, however, that these procedures now are well-established. Transformations of variables (such as the transmutation of the female statures) may be useful for a variety of reasons (Kutner et al., 2005, 127–137) and can be easily performed in e.g. SPSS.) The *R.F.F. Data* revealed that the off-spring of tall Mid-Parents is on average taller than P and that the relation between the Mid-Parental Stature and the *average* Stature of the Son is stable (where the Son refers to both the sons and the transmuted daughters). If the Mid-Parental deviation is D, then the filial deviation is on average $\frac{2}{3}$ D:

> I call this ratio of 2 to 3 the ratio of "Filial Regression." It is the proportion in which the Son is, on the average, less exceptional than his Mid-Parent. (Galton, 1889, 97)

Galton could now describe with numerical precision what he had discovered years before: the inheritance of characteristics is subject to 'regression towards the mean' (which should be labeled a 'phenomenon' in the language of chapter 5).

$$\mathtt{Stat.Son} = \mathtt{P} \pm \frac{2}{3}\ \mathtt{D},\ \text{where D is the Mid-Parent's deviation} \tag{9.4}$$

Take for example a Mid-Parent that is very tall, say 71.25 inch ($\mathtt{D} = 3$ inch). Equation (9.4) predicts and figure 9.2 illustrates that its issue will, on average, be 70.25 inch ($\mathtt{D} = 2$ inch $= \frac{2}{3} \times 3$ inch).[18]

[17] Note that 1.21 is a theoretically predicted number, as $\frac{1.7}{\sqrt{2}} = 1.21$. According to the *R.F.F. Data*, the Mid-Parental Q is 1.19. Galton considered the agreement between these numbers to be excellent (Galton, 1889, 92–94 and 208).

[18] Figures 9.1 and 9.2 nicely illustrate the difference qua explananda between Mendel's theory and Galton's.

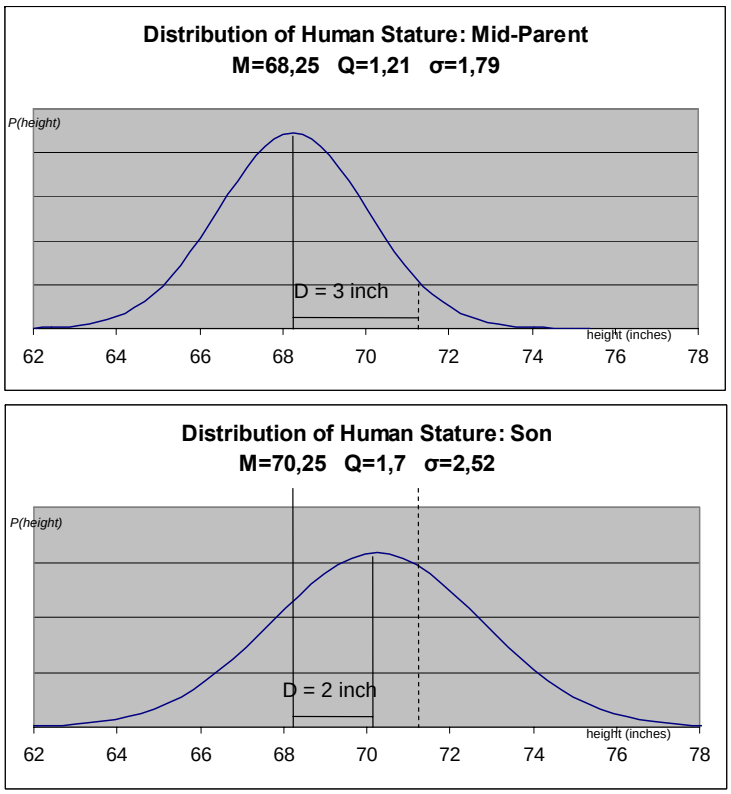

Figure 9.2: *Filial Regression illustrated*

Galton treated the equation describing filial regression not merely as a nice generalization. He considered it a law:

> By the use of this word [D or 'a deviate'] and that of 'mid-parentage,' we can define the *law* of regression very briefly. It is that the height-deviate of the offspring is, on the average, two-thirds of the height-deviate of its mid-parentage. (Galton, 1885, 508, my emphasis)

So here we are presented with a second set of constraints. Every theory of heredity should be able to explain the law of filial regression. In the following section, I will show how Galton's theory of Ancestral Inheritance did this.

'Filial Regression' was not the only kind of regression that Galton discovered (Galton, 1889, 99–110). Related notions which will prove to be relevant in the following sections are 'Mid-Parental Regression' (equation

9.5), 'Parental Regression' (equation 9.6) and 'Fraternal Regression' (equation 9.7):

$$\texttt{Stat.Mid-Parent} = \texttt{P} \pm \frac{1}{3}\,\texttt{D}, \text{ where } \texttt{D} \text{ is the deviation of the Son,} \tag{9.5}$$

$$\texttt{Stat.Son} = \texttt{P} \pm \frac{1}{3}\,\texttt{D}, \text{ where } \texttt{D} \text{ is } \ldots \text{ of one of his Parents,} \tag{9.6}$$

$$\texttt{Stat.Brother} = \texttt{P} \pm \frac{2}{3}\,\texttt{D}, \text{ where } \texttt{D} \text{ is } \ldots \text{ of a known man.} \tag{9.7}$$

As we will see, the equations (9.4), (9.5), (9.6) and (9.7) turn out to be spurious correlations from the point of view of Galton's theory of Ancestral Inheritance, just like the phenotypic distributions considered in chapters 4 and 5 turned out to be spurious from the point of view of classical genetics. Hence, if equation (9.4) is a law, it is not a causal law. In the language of chapter 1: it is a pragmatic law (a P-law), but not a causal pragmatic one (not a cP-law).

9.5 Galton's theory of ancestral inheritance as an *explanans*

In this section I will show how Galton's theory of inheritance satisfied the constraints generated by the normal distribution and regression towards the mean.

9.5.1 Particulate inheritance and the hypothesis of elementary errors

Mendel's causal mechanism responsible for the processes of heredity consisted of unit factors in pairs. In Galton's theory, the situation was less clear-cut. He called it the theory of '*particulate inheritance*' and used the words '*elements*' and '*particles*' several times. But sometimes these elements seemed to denote (elements of) phenotypic traits, other times they might have referred to carriers of hereditary traits. Nonetheless, as Galton was influenced by August Weismann's theory of the germ-plasm and Charles Darwin's concept of Pangenesis, a 'material' interpretation of the elements or particles is certainly justified (Galton, 1889, 7–9 and 192–193).

Galton's hereditary particles are transmitted from parents to offspring,[19] but in principle every ancestor may contribute to an individual's elements.

[19]Since phenotypic traits sometimes seem to skip a generation, he distinguished between *personal* elements (causing the traits they code for to be present), and *dormant* or *latent*

So the parents or their gametes do not screen off the offspring from the rest of its ancestry. The influence of remote ancestry is obviously smaller than the parental influence. Nevertheless, it exists and plays an important role in Galton's predictive inferences. The separate contribution of each ancestor follows a very simple rule, which would later be called the '*Law of Ancestral Heredity*' by Karl Pearson:[20]

> [...] the influence, pure and simple, of the Mid-Parent may be taken as $\frac{1}{2}$, and that of the Mid-Grand-Parent as $\frac{1}{4}$, and so on. Consequently the influence of the individual Parent would be $\frac{1}{4}$, and of the individual Grand-Parent $\frac{1}{16}$, and so on. (Galton, 1889, 136)

Taken together, the set of all ancestors fully determines the Son's set of elements, as

$$(2 \times \frac{1}{4}) + (4 \times \frac{1}{16}) + (8 \times \frac{1}{64}) + \ldots = 1.$$

At first glance, this picture seems paradoxical. Although all hereditary influence passes through the parents, there is still room for the influence of the grand-parents, great-grand-parents, etc. This semblance of paradox is dissolved if we distinguish between *personal allowance* and *ancestral allowance* (cf. section 9.6.2). Personal allowance is the allowance 'pure and simple' and it is governed by the Law of Ancestral Heredity (so that e.g. the

---

elements ('unused' elements, having no phenotypic influence). *Prima facie*, this strongly resembles the Mendelian distinction between dominant and recessive traits or factors. There is an important difference, however. The relation of dominance/recessiveness is fixed for each pair of contrasting traits (round seed shape is always dominant to wrinkled seed shape in *Pisum*). By contrast, Galtonian elements that are latent in one organism can be personal in another. Galton had no definite answer to the question what determined whether an element would be latent or personal. He thought there were three possible answers:

> [...] first, that in which each element selects its most suitable immediate neighbourhood, in accordance with the guiding idea in Darwin's theory of Pangenesis; secondly, that of more or less general co-ordination of the influences exerted on each element, not only by its immediate neighbours, but by many or most of the others as well; finally, that of accident or chance [...]. (Galton, 1889, 19)

[20]Galton explicitly discussed the validity of this 'Law of Ancestral Heredity' for the inheritance of personal elements. At the end of his *Natural Inheritance* he hypothesized that it would also apply to the latent elements (Galton, 1889, 187–191).

father's personal allowance is $\frac{1}{4}$). The ancestral allowance comprises all influence that is just passed through an ancestor. An individual's total allowance thus is the sum of its personal and its ancestral allowance.[21]

Now it is easy to explain why characteristics like human stature are normally distributed. In principle, an infinite number of ancestors influences the set of hereditary particles of a man. And his 'genotype' consists of an indefinite or incalculable number of elements. This makes sure that the inheritance of traits is determined by a '*variety of petty influences*' (Galton, 1889, 16–17).[22] If it is assumed that these are to some degree independent of one another (as Galton did), a physical basis is provided for the Hypothesis of Elementary Errors. So the theory of Ancestral Inheritance is capable of explaining the normal distribution of phenotypic traits (Galton, 1889, 84–85).

9.5.2 The Law of Ancestral Heredity, dilution and taxation

Now what is the cause of Filial Regression (equation 9.4)? Why is it that offspring tends to be more mediocre than its parents? Galton proposed an answer in his paper "Regression towards mediocrity in hereditary stature" (Galton, 1886), which he recapitulated in *Natural Inheritance* (Galton, 1889).

Suppose some Mid-Parent has Stature $\mathtt{P} \pm \mathtt{D}$ and call $\mathtt{D}$ her peculiarity. From equation (9.5), Galton stated, it follows that the peculiarity of the Mid-Grandparent is $\frac{1}{3}\mathtt{D}$, that of the Mid-Great-Grand-Parent $\frac{1}{9}\mathtt{D}$, etc. If each generation would contribute its whole peculiarity, we should expect the Son to inherit $\mathtt{D}(1 + \frac{1}{3} + \frac{1}{9} + \&c.) = \mathtt{D}\frac{3}{2}$. This contradicts the expected Filial Regression of $\frac{2}{3}$ (Galton, 1889, 134).

So Galton considered the possibility that the bequests of the successive generations are somehow taxed or diminished. His data did not allow to directly measure the size of this tax, but he had two limiting hypotheses. On the one hand, if the bequest of every generation is taxed just once, the tax rate has to be $\frac{4}{9}$ (since $\mathtt{D}\frac{2}{3} = \mathtt{D}\frac{3}{2} \times \frac{4}{9}$). On the other hand, if the tax is repeated at each successive transmission, the rate should be $\frac{6}{11}$.[23] Galton's data did

[21] Galton did not use the labels 'personal allowance' and 'ancestral allowance' in *Natural Inheritance*. They appear one time in his "The Average Contribution of each several Ancestor to the total Heritage of the Offspring" (Galton, 1897, 441).

[22] Galton incorporated two more sources of 'petty influences' in his theory. First, whether or not an element will be personal or dormant depends on very numerous influences (Galton, 1889, 22). Secondly, a trait such as Human Stature is not one element, but "a sum of the accumulated lengths or thicknesses of more than a hundred bodily parts" (Galton, 1889, 83–84), and each element or length of a body part is subject to errors or environmental effects.

[23] $\mathtt{D}\frac{2}{3} = 1\mathtt{D} \times \frac{6}{11} + \frac{1}{3}\mathtt{D} \times (\frac{6}{11})^2 + \frac{1}{9}\mathtt{D} \times (\frac{6}{11})^3 + etc.$

not allow to choose between these hypotheses. But as both values differed but slightly from $\frac{1}{2}$, he decided that this would be a very good approximation (Galton, 1889, 134–136).

Galton's reasoning lacked logical rigour and can be challenged from several sides.[24] Nevertheless, one should not consider it as totally *ad hoc*. Both in 1886 and in 1889, Galton concluded that, as the tax rate should be estimated to be $\frac{1}{2}$, the Mid-Parent contributes half of his peculiarity, the Mid-Grand-Parent one quarter, etc. That is, he combined it with the Law of Ancestral Heredity (see page 197).[25] In 1897, Galton published an extra argument for this law. If each ancestor *may* contribute to the heritage of the offspring ('as is shown by observation'), if remote ancestry contributes less than near ancestry ('as is well known'), if the contribution of the parents to the children is the same as that of the grand-parents to the parents, etc. ('as is reasonable to believe'), and if the total amount contributed equals 1 ('as is necessarily the case'), then only the series of $\frac{1}{2} + (\frac{1}{2})^2 + (\frac{1}{2})^3 + \textit{etc.}$ can describe the share of the Mid-Parent, the Mid-Grand-Parent, etc. (Galton, 1897, 403). Michael Bulmer deems it very plausible that Galton had this argument in mind in 1886 (Bulmer, 2003, 246).[26] If he is right, as I think he is, Galton's choice for the tax rate of $\frac{1}{2}$ was not totally *ad hoc*.[27]

So now we see how Galton's theory explains regression towards the mean. Since the peculiarity $\mathtt{D}$ of the Mid-Parent is mixed with the smaller peculiarities of more remote ancestry (viz. $\frac{1}{3}\mathtt{D}, \frac{1}{9}\mathtt{D}, \ldots$), the Son's deviation from $\mathtt{P}$ is smaller than $\mathtt{D}$. Galton used a very powerful metaphor to illustrate this:

> [The] effect resembles that of pouring a measure of water into a vessel of wine. The wine is diluted to a constant fraction of its alcoholic strength, whatever that strength may have been. (Galton, 1889, 105)

The exceptionality of the parents is diluted by the mediocrity of the rest of the ancestry (hence, I call this the *Dilution Theory*), so that their offspring is more mediocre too. But, as we have seen, the Dilution Theory needed to

[24] For a crushing discussion of Galton's derivation, see Bulmer (1998) and Bulmer (2003, 243).

[25] Note that the Law of Ancestral Heredity matches only the second of both limiting hypotheses, namely that the tax is repeated at each successive transmission.

[26] In the bibliography of Bulmer (2003), however, this article is wrongly dated to 1885.

[27] Bulmer's conviction is based on the following quote by Galton: "These and the foregoing considerations were referred to when saying that the law might be inferred with considerable assurance *à priori* [...]." (Galton, 1897, 403) Galton had indeed stated in 1886 that his law might have been deductively foreseen (Galton, 1886, 253).

be completed with the concept of *Taxation* to get the correct ratio of Filial Regression.

9.6 Two possible objections and a conclusion

Before I turn to the conclusion, I want to anticipate two possible objections to my thesis that statistics played a blinding role in the development of Galton's theory of heredity. I will show that neither the observational nature of the data on Human Stature, nor the continuous or blending nature of the observed characteristics can be cited as alternative explanantia for Galton's failure to arrive at the 'right' theory of heredity.

9.6.1 The nature of Galton's data

One of Mendel's major merits was that he paid a lot of attention to the design of his genetic crosses.[28] By carefully selecting a well-suited organism (pea plants) and manageable pairs of opposing characteristics, and by meticulously planning the right monohybrid as well as multihybrid crosses, he was able to confirm his causal theory of inheritance – a theory that still is considered the basis of modern genetics, although it has been subject to a vast amount of changes, specifications and additions.

It is certainly true that Galton's data were gathered in far less controlled circumstances. He preferred data about humans because he considered them more interesting or relevant. As a consequence his subjects were less easily controllable.

In the 1870's, however, after having published *Hereditary Genius* (1869), but before collecting the *R.F.F. Data*, Galton did study plants (*sweet peas*, not to be confused with *P. sativum*). He weighted thousands of seeds to determine their size and then selected several sets for planting. Each set consisted of seventy seeds, divided in seven packets of ten seeds of exactly the same weight $(K, L, \ldots, Q)$. The K-class contained very heavy seeds, L the next heaviest, and so on. He sent these sets to his friends throughout the United Kingdom and asked to plant them according to very minute instructions and to collect the produce of each class separately. Seven plantings

[28] It should be noted that this section is slightly different in my paper "Galton's Blinding Glasses" (Leuridan, 2007a). There it was called 'Observational versus experimental data'. It met the possible objection that Galton's data were observational, whereas Mendel's were allegedly experimental. The paper was written before chapter 7, and now the distinction between Galton's observational data and Mendel's experimental data is superseded. I was deluded by the statements of Mendel and others regarding the experimental nature of their designs. The results of chapter 7, however, by no means affect the present argument.

(with in total $7 \times 7 \times 10 = 490$ parental seeds) succeeded (Galton, 1877, 512–514 and 1889, 79–82, 225–226).

The data showed that large seeds beget large seeds. But, as was the case with hereditary genius, '*Reversion*' could be observed (the label 'regression towards the mean' was not yet introduced in 1877). In 1877, he gave no exact value for the regression coefficient, stating only that it is constant. In "Regression towards mediocrity in hereditary stature", which dates nine years later, this lacuna is removed:

> It will be seen that for each increase of one unit on the part of the parent seed, there is a mean increase of only one-third of a unit in the filial seed; and again that the mean filial seed resembles the parental when the latter is about 15.5 hundredths of an inch in diameter.[29] (Galton, 1886, 259)

This suffices to show that the difference between Mendel's and Galton's scientific practice should not be sought in the nature of their data (highly controlled in Mendel's case, poorly controlled in the case of Galton).

9.6.2 Alternative inheritance versus blended inheritance

Can't we explain the difference between Galton and Mendel by looking at the variables, i.e. the phenotypic traits, they studied? After all, Human Stature and the Size of sweet peas are continuous variables, while Mendel observed pairs of discrete, opposing characteristics.

It is certainly true that Galton paid a lot of attention to continuous traits and it is equally true that that was the best way to discover regression-phenomena.[30] In *Natural Inheritance*, however, he took great pains to ar-

[29] See also Galton (1889, 225).

[30] As I phrase it here, it looks as if Galton did not come upon the phenomenon of regression by merely analysing his data, but that he actively sought for it. In fact, this was indeed the case. Galton needed 'Regression towards the Mean' to explain another statistically inspired *explanandum*.

In 1877, he published "Typical Laws of Heredity" (Galton, 1877). One of the main explananda in that paper was the fact that the distribution of characteristics (specifically the Size of sweet peas) remained more or less constant in each successive generation. Using the terminology of 1889, it was to be explained why both its `M` and `Q` remained constant.

In the absence of regression, the dispersion `Q` would increase from generation to generation. The offspring of tall men would on average be as tall as its parents, some of it would even be taller; in the next generation, even taller issue would result, ... But this would contradict Galton's data about sweat peas, as well as the findings of Quetelet, Galton and others on human characteristics (Venn, 1889, 415). Regression was the perfect candidate to solve this problem, as it would act as a counterbalance to this dispersive tendency.

gue that the theory of Ancestral Inheritance could encompass the transmission of both *blended heritages* and *alternative heritages* (Galton, 1889, 12–14 and 138–153). To prove this point, Galton studied the transmission of Eye-Colour:

> If notwithstanding this two-fold difference between the qualities of Stature and Eye-colour, the shares of hereditary contribution from the various ancestors are alike in the two cases, as I shall show they are, we may with some confidence expect that the law by which those hereditary contributions are found to be governed, may be widely, and perhaps universally applicable. (Galton, 1889, 139)

How could the Law of Ancestral Heredity be used to predict the distribution of Eye-colour in issue, conditional on the Eye-colour of its parents, grand-parents, etc.? Galton distinguished between three types of Eye-Colour: light, hazel and dark, and then treated the problem as if it concerned Stature.

Suppose you want to predict the stature of some man, S, but that you only have information about one of his parents, F, having peculiarity $\mathtt{D}$. By equation (9.5), the parents of F have on average the peculiarity $\frac{1}{3}\mathtt{D}$, while his grand-parents (i.e. the great-grand-parents of S) have $\frac{1}{9}\mathtt{D}$, ... From the Law of Ancestral Heredity it follows that F transmits only $\frac{1}{4}$ of his peculiarity to S; his parents transmit $\frac{1}{16}\mathtt{D}$, etc. So the total calculable or predictable heritage that is transmitted through F is[31]

$$\mathtt{D}\{1 \times 1 \times \frac{1}{4} + 2 \times \frac{1}{3} \times \frac{1}{16} + 4 \times \frac{1}{9} \times \frac{1}{64} + \&c.\} \approx \mathtt{D} \times 0.30,$$

consisting of F's known personal allowance ($0.25\ \mathtt{D}$) and its predictable ancestral allowance ($(0.30 - 0.25)\mathtt{D} = 0.05\ \mathtt{D}$). By analogy, two parents have a known total allowance of 0.60, "leaving an indeterminate residue of 0.40 due to the influence on ancestry about whom nothing is either known or implied" (Galton, 1889, 149). The residue is a direct consequence of the fact that the ancestral influence is not screened off by the parents or the parental gametes in Galton's theory.

These results can be easily re-interpreted in the context of Eye-colour. We only need to interpret the personal allowance or ancestral allowance as fractions of the total number of children in a family that will inherit some specific trait (Galton, 1889, 149–150). If a parent has dark eyes, 30% of his children will have dark eyes. Of the children of two dark-eyed parents, 60%

[31] See Galton (1889, 148–149). Note that Galton's formula on p. 149 contains a printing error.

will be dark-eyed. Of the residue, 40% in this case, Galton assigns 28% to dark eyes and 12% to light eyes, proportionally to their overall ratio in the population. Galton's reasoning gives rise to the following table, from which it is easy to predict the distribution Eye-colour in issue, conditional on the Eye-colour of its parents and grand-parents.[32]

For example, from the premise that in a family there are two light-eyed parents, three light-eyed grand-parents and one hazel-eyed grandparent, you can calculate that on average 91% of the children will have light eyes. The rest, 9%, will be dark-eyed.[33]

Contribution to the heritage from each	Data limited to the eye-colour of the					
	2 parents		4 grand-parents		2 parents and 4 grand-parents	
	I.		II.		III.	
	Light	Dark	Light	Dark	Light	Dark
Light-eyed parent	.30				.25	
Hazel-eyed parent	.20	.10			.16	.09
Dark-eyed parent		.30				.25
Light-eyed grandparent			.16		.08	
Hazel-eyed grandparent			.10	.06	.05	.03
Dark-eyed grandparent				.16		.08
Residue, rateably assigned	.28	.12	.25	.11	.12	.06

Thus we can conclude that, even if Galton paid heavy attention to continuous variables, he also tried to incorporate the transmission of discrete characteristics.

9.6.3 The neutrality of Galton's statistics

I have shown that, in the second half of the 19th century, Galton's views on the mechanism of heredity were so much constrained by his statistical explananda, that he failed to discover the Mendelian scheme.[34] It follows that statistics *can* bias scientific research. Moreover, Galton's case can be supple-

[32]Note that only the distribution of dark eyes and of light eyes is calculated. This table is reproduced from Galton (1889, 213).

[33]$2 \times 0.25 + 3 \times 0.08 + 1 \times 0.05 + 0.12 = 0.91$,
$2 \times 0.00 + 3 \times 0.00 + 1 \times 0.03 + 0.06 = 0.09$ (see also Galton, 1889, 215, Table 19).

[34]By this I do of course not mean that Mendel's theory was either unconditionally true or that it is still used in its original form. I only mean that he laid the fruitful basis of modern genetics, even if the development of genetics involved a lot of changes to his basic tenets.

mented with examples from outside biology. In his utmost interesting book, *The Taming of Chance* (1990), Ian Hacking has argued that the practice of descriptive statistics by the bureaucracies of nation states affected the way people conceived of society, of other people and of themselves. So statistics has biased people's world views and it has (partly) constrained theory development in the social sciences too. So clearly, probability and statistics were *not* always *neutral* scientific tools.

9.7 Consequences for contemporary statistics

In the literature on causal discovery, probability and statistics play an important role, as several flourishing research programmes use conditional (in)dependence relations as indicators of the presence and absence of causal relations and maintain strong ties to statistical techniques such as SEM (Pearl, 2000; Spirtes et al., 2000). In the SEM-literature, however, one is frequently warned not to draw causal inferences from structural models too quickly. Causal inference is only justified on the basis of models that fit the data well. But good model fit is not enough. It may indicate that the model accurately reflects reality, but not necessarily so. It leaves open the possibility that the model is equivalent to one that corresponds to reality but itself is incorrect, or that it fits the data from a nonrepresentative sample but has poor fit in the population, or that it has so many parameters that it cannot have poor fit (Kline, 2005, 321). In addition to fitting the data well, the model should also correctly describe the causal relations between its variables.

A SEM-model has to be specified independently before it is confronted with the data. In the SEM-literature, heavy stress is laid on the need for reliable background knowledge or theory in this process.[35] But this presupposes that this background knowledge is not perniciously biased by statistics itself.

So what about the neutrality of contemporary statistics? I would like

[35]The following quote illustrates this and shows at the same time that, contrary to what Pearl has argued, the exclusion of causal interpretations from SEM cannot be completely put down to SEM-practitioners seeking respectability by keeping causal assumptions implicit, or to the unsuitability of algebraic language for making or expressing causal assumptions (Pearl, 2000, 137–138).

> It is only from a solid base of knowledge about theory and research that one can even begin to address [the] requirements for inferring causation from correlation. Although facility with the statistical details of SEM is essential, it is not a substitute for what could be called wisdom about one's research area. (Kline, 2005, 95)

to make two statements in this respect. First of all, we should not throw out the baby with the bathwater. One of statistics' fundamental strengths is that it has much thought for the presuppositions and internal limitations of each of the techniques developed (presuppositions with respect to the distribution of the data or the relation between variables, robustness against missing values and outliers, requisites concerning sample sizes, etc.). Add to this that tests have been developed for most of these presuppositions (tests which were not available to Francis Galton). Secondly, however, we should not be blindly optimistic. Galton's problem still is relevant today. Specifying a SEM-model involves that we prestructure our domain of interest. Examples of such prestructuring can be found at many places in Kline's introductory work (Kline, 2005). SEM-modelling involves marking out variables, specifying their possible values and developing and including methods for measuring them. It involves fixing covariance relations and perhaps imposing constraints on (covariances between) disturbance variables. It involves specifying the directionalities of presumed causal effects ...

Consequently, SEM (and like methods) should be used very carefully in practice. One should always bare in mind the possibly *blinding* influence of the techniques themselves.

9.8 Galton's causal scheme

Up till now, I have pursued the first goal of this chapter, viz. to show that statistical techniques (including contemporary ones), although valuable and widely used for causal discovery, may be perniciously biasing. In this final section I will briefly discuss the difference between Galton's theory and the theory of classical genetics. In figure 9.3 the causal structure of Galton's theory of inheritance is sketched by means of the inheritance of stature. The variable $elem_1$ denotes the set of elements of some human. I have not distinguished between her latent elements and her personal elements. Both are joined in one variable. The variable $elem_1$ causes the variable $stat_1$, which denotes her stature. The states of $elem_1$ are in principle determined by an infinite number of ancestral 'genotypes'. I have included six ancestors: two parents (numbers 2 and 3), and four grandparents (numbers 4 up to 7). This numbering is governed by the recursive rule Galton outlined in his "A Diagram of Heredity" (Galton, 1898).[36] The influence of the rest of

[36]Galton set forth his rule as follows:

> The Subject of the pedigree is numbered 1. Thenceforward whatever be the distinctive number of an ancestor, which we will call n, the number of its sire is $2n$, and that of its dam is $2n + 1$. All male numbers in the pedigree

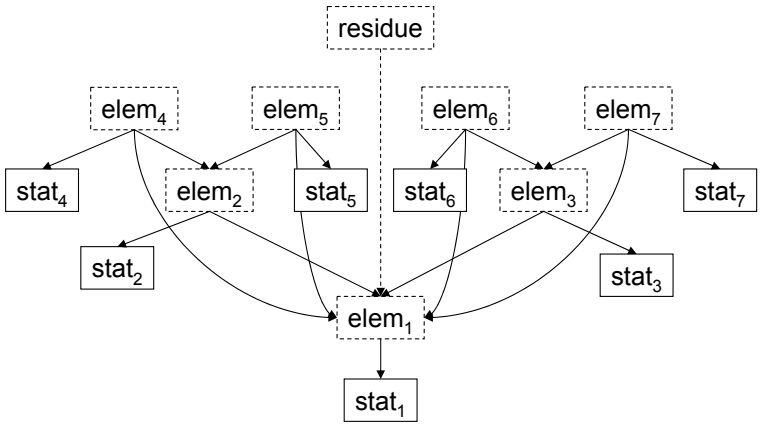

Figure 9.3: *The causal structure of Galton's theory of inheritance (sketch)*

her ancestry is represented by the variable *residue* (cf. the table in section 9.6.2). This is of course a stopgap (this is why I have dotted the edge between *residue* and $elem_1$.) The stature $stat_i$ of each ancestor is determined by the corresponding variable $elem_i$. All $elem_i$-variables are dotted because they are **T**-theoretical with respect to the theory of ancestral inheritance.

The graph in figure 9.3 is common to all pedigrees relating to the inheritance of stature in humans in which the statures of both parents and those of both grandparents are known (cf. column III. of the table in section 9.6.2). By specifying the values of *residue* and of all variables $elem_i$ ($2 \leq i \leq 7$), and by specifying the relations between these several variable (cf. taxation), the distribution over $elem_1$ and $stat_1$ can be derived.

Two remarks are in order here. Firstly, Galton never discussed the set of possible values (the space) of the variables $elem_i$ (let alone of *residue*) in detail. Whereas the Mendelians had (ever more) precise ideas about the possible values of the GT_i-variables, Galton only stated that the "there is no direct hereditary relation between the personal parents and the personal child" and that "the main line of hereditary connection unites the sets of elements out of which the personal parents had been evolved with the set out of which the personal child was evolved" (Galton, 1889, 19). Whereas abductive reasoning from phenotypes to genotypes was theoretically supported in classical genetics, it was absent in Galton's framework.

Secondly, from these quotes it is evident that Galton's phenotypic general-

are therefore even, and all female numbers are odd. To take an example – 2 is the sire of 1, and 3 is the dam of 1; 6 is the sire of 3, and 7 is the dam of 3. (Galton, 1898, 293).

izations, i.e. the equations (9.4), (9.5), (9.6), and (9.7), are spurious according to the theory of Ancestral Inheritance. Take the case of equation (9.6). The parental stature and the filial stature are effects of a set of common causes: the sets of elements of all ancestry. So where the law of ancestral heredity may be considered a causal P-law, the cited equations are non-causal P-laws.

Let us now compare figure 9.3 with figure 4.1. The graphs in these figures are not isomorphic. Hence the theory of Galton differs strongly from the theory of classical genetics. They have different causal schemes. It could of course be objected that in figure 4.1 I included but two generations (the parental and the filial generation), whereas in figure 9.3 I included three. This is true, but innocuous to my argument. Firstly, I included three generations to make clear the point that offspring is not d-separated from its remote ancestry by (the gametes of) its parents. Secondly, I could have restricted the graph in figure 9.3 to two generations. In that case, I would still have needed the variable *residue* (and its influence on $elem_1$ would then have been even stronger – cf. column I. in the above table). So anyhow, my reconstruction would have resulted in a graph that is *not* isomorphic to the graph in figure 4.1.

9.9 Concluding remarks

The present chapter served two goals. The first goal was to show that statistical techniques played a perniciously biassing role with respect to Galton's theory of ancestral inheritance and that these days the same might sometimes hold for e.g. structural equation modelling. The second goal was to show that the causal structure of Galton's theory strongly differed from that of classical genetics. Their respective causal models were not isomorphic and hence do no belong to the same causal scheme.

An important objection might be raised here. There could be a risk that the concepts which I used have equally played a biassing role. This would of course not follow directly from the results of the present chapter. After all, although I used concepts like 'causal net', 'credal net', and 'causal scheme', I did not use any statistical techniques. But the objection is legitimate and I will examine it in a minute.

Chapter 10

Concluding Remarks: the Biassing Role of Causal Models

In chapter 1 I presented the criteria for strict lawfulness that were proposed by i.a. Nagel, Hempel and Goodman. I also stated that these criteria were strongly shaped, or biassed, by the language and the inferential characteristics of first order classical logic – a logic the language of which moreover does not allow for a distinction between causal laws and non-causal ones. This bias was pernicious, as these criteria hardly allow for laws of nature in the special sciences (or even in physics!).

I proposed to use the concepts of 'P-law' and 'cP-law' as an alternative. These concepts are less strict and allow for laws of nature in the special sciences. They are formulated against the background of the causal modelling framework. This offered several advantages.

Firstly, the distinction between causal and non-causal laws could straightforwardly be expressed. Secondly, the causal structure of scientific theories, *in casu quo* classical genetics, could adequately be described. And it was seen that several related concepts from philosophy of science, such as anomalies, exemplars and theory-elements could fruitfully be adapted and incorporated. Thirdly, we saw that explanation on the basis of non-strict laws is nothing to be afraid of. Though explanation in genetics relies on causal pragmatic laws, it may account for different kinds of explananda (phenotypic distributions, singular events). It incorporates many different aspects commonly attributed to scientific explanation (derivational unification, ontological unification, mechanistic explanation, ...). Fourthly, non-strict laws are not only apt for explanation, but also for policy. It is well established in the literature on causal modelling that causal pragmatic laws can be used for manipulative policy. In addition, however, we saw that non-causal pragmatic laws are useful for policy (selective policy) too. Finally, the causal modelling frame-

work offered an interesting way to examine causal discovery. We saw that experimental data provide more solid grounds for discovering causal pragmatic laws than merely observational data and that classical genetics went to a large extent without experiments. Although experiments are useful for causal discovery, they are not indispensable. Many tools for observational causal discovery have been devised. My adaptive logic **ALIC** provides such a tool and it can handle an important and widespread problem: the problem of ignorance.

The framework of causal modelling also fits the methodology of many of the special sciences. It is closely related to statistical techniques used in the social sciences, in epidemiology, in biology, etc. It is even explicitly connected to structural equation modelling by Woodward, Pearl, etc. But as we saw in the last chapter, structural equation modelling should be applied carefully. Relying on the case of Galton's theory of inheritance I showed that statistical techniques (even contemporary ones) may bias our causal beliefs.

Do we have any guarantee, then, that the framework of causal modelling does not bias our causal beliefs as well? Or that it did not bias our view on causality and laws of nature in the special sciences? Quite the contrary! The concepts of causal modelling certainly are not innocent. In this final chapter, I will briefly discuss the biassing role of causal models with respect to both causal discovery and the concept of causality. I will not endeavour to glance through the plethora of objections that have been raised against causal modelling. I will concentrate on four central and closely related issues: the causal Markov condition, the faithfulness condition, invariance, and modularity. Whereas I discussed the first three concepts frequently, I did not touch upon modularity. But as it has been heavily debated in the past ten years, I will do it here.

10.1 The causal Markov condition

There are several key assumptions underlying both algorithms for causal discovery from non-experimental data and the causal modelling semantics. One central assumption is the *causal Markov condition*, viz. that each variable is probabilistically independent of its non-effects, conditional on its direct causes. This condition is not uncriticized. Jon Williamson (2005, 49–57) cites five sets of objections raised indirectly against the causal Markov condition.[1]

[1]In his view the objections affect the causal Markov condition if both causality and probability are interpreted physically (not as features of an agent's mental state). The discussion of these objections serves an overarching goal in Williamson (2005), viz. to argue for an objective Bayesian interpretation of probability and for an epistemic interpretation

The causal Markov condition is related to the *Principle of the Common Cause*. This principle is a development of John Stuart Mill's Fifth Canon of Inductive Reasoning and was formulated by Hans Reichenbach in close to its current form (Williamson, 2005, 51).

Definition 10.1 (Principle of the Common Cause) *If* $\sim(A \amalg B)$*, then* $A \Rightarrow B$ *or* $B \Rightarrow A$ *or there is a* $\mathbf{Q} \subseteq V$ *such that* $C \in \mathbf{Q}$ *implies* $C \Rightarrow A$ *and* $C \Rightarrow B$*, and* $(A \amalg B \mid \mathbf{Q})$.

The principle of the common cause is relevant for the present discussion, since it is implied by the causal Markov Condition. Hence any counterexample to the former calls into question the validity of the latter.

Theorem 10.2 *The causal Markov condition implies the Principle of the Common Cause.*

Proof. See Williamson (2005, 52) ■

According to the principle of the common cause, if two variables are probabilistically dependent, there must be a causal explanation for this. Either one causes the other, or they are effects of a common cause. Jon Williamson (2005, 52–57) cites five sets of objections to the principle of the common cause. Random variables may be probabilistically dependent without there being any causal explanation: (i) they may be so related by accident, (ii) or by having related meanings, (iii) they may be logically related, (iv) or mathematically, or (v) the dependence may result from non-causal physical laws or from local, non-causal constraints and initial conditions.

Firstly, probabilistic dependencies may arise by *accident*, as is shown by Elliott Sober's famous example about the sea level in Venice and the price of bread in Britain. Both have been on the rise in the past two centuries, and thus are strongly positively correlated. But none is a cause of the other. Nor are they effects of a common cause (unless one decides to treat time as a common cause). Probabilistic dependencies may also result from variables having related *meanings*. Analogously, they may result from variables being *logically related*. This is why I stipulated in chapter 8 that all variables in V should be different (different name = different variable) and that they should be logically independent. So V cannot contain both 'Gender' and 'Being a King' as a variable, but it can contain both 'Gender' and 'Being a royal head of state'. (By definition, kings are male.) Fourthly, probabilistic dependencies may result from *mathematical relations*. Williamson illustrates this by

of causality.

an example of a probabilistic expert system for colonoscopy. The aim was to produce a system for guiding an endoscope up a patient's colon, using a camera image to assist the guiding process. From the camera image, the colon centre (which shows up as a large dark region on the screen) had to be distinguished from pockets in the colon wall (which show up as small dark regions on the screen). In order to distinguish between both types of region, two variables were used: the mean light intensity M and the intensity variance V. (Williamson, 2005, 49) But the mean and the variance of a variable are related mathematically, not causally. Finally, probabilistic dependencies may arise from *non-causal physical laws* or from *local non-causal constraints* and *initial conditions*. The latter case is illustrated by an example from Nancy Cartwright.

> A typical case occurs when a cause operates subject to constraint, so that its operation to produce one effect is not independent of its operation to produce another. For example, an individual has \$10 to spend on groceries, to be divided between meat and vegetables. The amount that he spends on meat may be a purely probabilistic consequence of his state on entering the supermarket; so too may be the amount spent on vegetables. But the two effects are not produced independently. The cause operates to produce an expenditure of n dollars on meat if and only if it operates to produce an expenditure of $10 - n$ dollars on vegetables. Other constraints may impose different degrees of correlation. (Cartwright, 1994, 113–114)

10.2 The Faithfulness condition

The causal Markov condition is not the only assumption underlying algorithms for causal discovery that has been under attack. The *faithfulness condition*, which states that effects should be probabilistically dependent on their causes, has been criticized too.[2] (The faithfulness assumption is less central for the causal modelling semantics than is the causal Markov condition.) Nancy Cartwright (2001, 244–248) cites two problems for the faithfulness condition. The first problem is *Simpson's Paradox*: facts about probabilistic dependency (positive dependency, negative dependency, inde-

[2]For the exact interpretation of the faithfulness condition (which in fact is a binary relation between *DAG*s and probability distributions), see definition 8.4.

pendency) may be reversed in moving from populations to subpopulations.[3] Pearl illustrates this as follows:

> For example, we may find that students who smoke obtain higher grades than those who do not smoke but, adjusting for age, smokers obtain lower grades *in every age group* and, further adjusting for family income, smokers again obtain higher grades than nonsmokers *in every income-age group*, and so on. (Pearl, 2000, 78, my emphasis)

The reversal of probabilistic dependencies in this case may be the result of probabilistic dependencies between smoking on the one hand and age and income on the other hand.

The version of Simpson's Paradox that is most relevant here is the situation in which two variables X and Y are probabilistically dependent (respectively, independent) in some population, but independent (respectively, dependent) in any subpopulation obtained by conditioning on the value of some third variable Z. Depending on the population chosen, X may or may not be considered a cause of Y. (Cartwright adopts Suppes' notion of *prima facie* causation. X is a *prima facie* cause of Y if and only if it precedes Y in time and is correlated with Y. Genuine causes, then, are *prima facie* causes that survive the same independence tests as in my axiom **(A6)** – see section 8.4.3. Simpson's Paradox shows that X may be a genuine cause of Y, and nevertheless not be a *prima facie* cause of Y in the subpopulations studied.)

The second problem for the faithfulness condition affects cases where one and the same cause has different influences on an effect, and where these influences cancel each other. Cartwright cites Hesslow's famous example of the birth-control pills.

> The pills are a positive cause of thrombosis. On the other hand, they prevent pregnancy, which is itself a cause of thrombosis. Given the right weights for the three processes, the net effect of the pills on the frequency of thrombosis can be zero. (Cartwright, 2001, 246)

If such cancellation occurs, birth-control pills and thrombosis are probabilistically independent even though the former are a genuine cause of the latter.

[3]Simpson's Paradox is not a paradox in the strict logical sense.

10.3 Markov, faithfulness, and the problem of causal discovery

What should we conclude from the objections against the causal Markov condition and the faithfulness condition?[4] Do they affect the results of the previous chapters? We should distinguish between two issues here. The first issue regards the problem of causal discovery. The second issue is semantic or definitional, as it regards the concept of causality and laws of nature. Let me tackle the problem of causal discovery first.

The causal Markov condition and the faithfulness condition both are indispensable fundamentals underlying the **IC** algorithm, the logic **ALIC**, and many other tools for causal discovery. The objections discussed in the previous sections show that **IC** and **ALIC** are fallible. No one would have expected otherwise! As the objections mentioned above show, the causal Markov condition may induce us to posit causal relations where in fact there are none. And the faithfulness condition may induce us to disregard existing such relations. But does that mean that we have to abandon such algorithms completely? Cartwright seems to suggest we do. By presupposing that the causal structure to be discovered is faithful to our probabilistic knowledge (in cases where we lack the information necessary to confirm this) we run the risk of not getting the causal structure right. She responds with a truism:[5]

> when you don't know, you don't know; and it is often dangerous to speculate. (Cartwright, 2001, 249)

Two questions are in order here. Firstly, how often will we fail to get the causal structure right? Secondly, how dangerous would such failure be?

In any case, failures of the causal Markov condition should be ruled out as much as possible. All background knowledge (if any) suggesting that some probabilistic dependency may not be the result of an underlying causal structure but is rather due to accident, or semantic relations, or logical relations,

[4] I will sidestep another assumption underlying most algorithms for causal discovery, viz. that causal relations are not cyclic. The acyclicity assumption can easily be satisfied even in contexts where the structures to be studied are cyclic (Williamson, 2005, 50).

[5] Cartwright often stresses that she favours the Bayes-nets account of causal discovery, but the truism – if taken seriously – seems to undermine the whole approach. When you *don't know* that the causal structure to be discovered is faithful to your probabilistic knowledge, you *don't know*. In that case, Cartwright claims, it is often dangerous to speculate. But when you *know* that it is faithful, you *know*. That is, you know the causal structure and may compare it with your probabilistic knowledge. In that case you would no longer need the algorithms in question.

etc. should be taken into account. One of the main advantages of **ALIC** is that incorporating such background knowledge really is straightforward.

What about failures of the faithfulness assumption? The faithfulness assumption has been defended by showing that at least in a measure theoretic sense it is reasonable (Spirtes et al., 2000, 41–42; Meek, 1995b). Intuitively, the argument is as follows: take the set $\mathbb{P}^G$ of probability distributions that are causally Markov relative to some *DAG* G. Part of these distributions will also be faithful to G, but not all of them. Suppose one randomly draws a distribution from $\mathbb{P}^G$. Then the probability of drawing an unfaithful distribution is zero (Meek, 1995b, theorem 7). More specifically, with respect to the Lebesgue measure over the space of sets of parameters that characterize the members of $\mathbb{P}^G$, the set of distributions which are unfaithful to G has measure zero. This should show that we would seldom fail to get the causal structure right.

Cartwright (2001, 250) is not impressed by this argument. For one thing, she does not see how Lebesgue measures connect with the way in which parameters are chosen or arise naturally for the causal systems to be studied.[6] So Meek's result seems irrelevant. Moreover, she argues, it is also 'an irresponsible interjection into the discussion.'

> Getting it right about the causal structure of a real system in front of us is often a matter of great importance. It is not appropriate to offer the authority of formalism over serious consideration of what are the best assumptions to make about the structure at hand. (Cartwright, 2001, 250)

I would not go this far. I think we rather have to accept that our methods for causal discovery from non-experimental data are fallible. But from the fallibility of these algorithms (or of inductive methods in general) it should not be concluded that they may never be used. If, in some context, they are the best we have, there may be good reasons to use them. All scientific work is incomplete. But, in the words of A.B. Hill, that does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.

This does not mean, of course, that *any* action may be done on the basis of *any* evidence. Let me quote A.B. Hill once more:

> Finally, *in passing from association to causation I believe in 'real life' we shall have to consider what flows from that decision.* On scientific grounds we should do no such thing. The evidence is

[6]For a further discussion of the probability of exceptions to the faithfulness condition and of Cartwright's arguments, see Steel (2006).

> there to be judged on its merits and the judgement (in that sense) should be utterly independent of what hangs upon it – or who hangs because of it. *But in another and more practical sense we may surely ask what is involved in our decision.* (Hill, 1965, 300, my emphasis)

Hill's paper focusses on occupational medicine, and the phrase 'what is involved in our decision' touches medical, economic, social, political and other considerations. Some decisions may be based on 'relatively slight evidence', such as the decision to restrict the use of a drug for early-morning sickness in pregnant women. Others should be based on 'fair evidence', such as the decision to change from a probably carcinogenic oil to a non-carcinogenic oil in a limited environment. But 'very strong evidence' would be needed before we make people "stop smoking the cigarettes and eating the fats and sugar that they do like." (Hill, 1965, 300)

Algorithms for causal discovery, like any inductive inference procedures, should be applied as carefully as possible. And in any application, it should be backed with as much relevant background knowledge as possible. But we should not rule out from the start applying them in cases where they might fail. How dangerous it is to speculate depends on 'what flows from our decision' and cannot be determined apart from concrete cases.

10.4 Invariance and modularity

Up till now I have discussed two important assumptions: the causal Markov condition and the faithfulness condition. I thereby focussed on the problem of causal discovery. Woodward and others closely tie the causal Markov condition to another central concept, *modularity*, in a way that affects the semantics or the definition of causality. Modularity resembles invariance, but whereas invariance is a feature that applies to single generalizations, modularity applies to sets of generalizations. Both modularity and invariance are important with respect to the concept of causality.[7]

Woodward discusses modularity primarily with respect to causally interpreted structural equation models (i.e. sets of causally interpreted equations). It conveys the assumption that each equation is independently changeable, without disrupting the other equations.

> [...] a system of equations will be modular if it is possible to disrupt or replace (the relationships represented by) any one of

[7] 'Modularity' is applied in many different senses, see Callebaut and Rasskin-Gutman (2005) and Mitchell (2006). I will stick to the concept of Hausman and Woodward here.

> the equations in the system by means of an intervention on (the magnitude corresponding to) the dependent variable in that equation, without disrupting any of the other equations. Modularity is thus a feature that a set of representations of causal relationships (e.g., equations or a directed graph) may (or may not) possess.[8] (Woodward, 2003b, 48)

In Woodward's view, the concept of modularity presupposes that all equations or generalizations in question are to some extent invariant, as is seen from his more precise definition **(MOD)**. Thus modularity and invariance are two distinct concepts, where the former implies or presupposes the latter, but not vice versa.

> **(MOD)** A system of equations is modular if (i) each generalization is [...] invariant under some range of interventions and (ii) for each equation there is a possible intervention on the dependent variable that changes only that equation while the other equations in the system remain unchanged and level-invariant. (Woodward, 2003b, 329)

In the previous chapters, I never discussed sets of structural equations explicitly. But I did discuss conditional probability tables. Conditional probability tables play the same role as Woodward's structural equations. They provide constraints to be satisfied by probability distributions and by causal nets. For example, tables 1–3 in section 4.5 provide constraints for $P \in \mathbb{P}^1$, and hence indirectly for $\mathcal{B} \in \mathfrak{B}^1 = \langle G^1, \mathbb{P}^1 \rangle$. $\mathfrak{B}^1$ is the set of causal models that satisfy tables 1–3 (and are isomorphic to the graph in figure 4.1).

In chapter 5 I showed that the generalizations in tables 1–3 are invariant (at least according to the theory of classical genetics). Suppose that a pea plant has the genotype $GT = ts$ and the phenotype $PT = tall$. Then it is to be expected that if by some ideal intervention its genotype were changed to $GT = ss$, its phenotype would change to $PT = short$.

Modularity applies to systems of generalizations. For example, within each of its theory-elements the principles of classical genetics are (or can be deemed) modular. It could be expected that by intervening on the gametic make-up CG_1 of a paternal plant (or a group of paternal plants) in a genetic crossing, detaching it from the causal influence of GT_1, one may change the value (or the probability distribution over the values) of the genotype GT_3 of its offspring in accordance with the principle of composition. If so, the principle of composition is invariant under this intervention. (One may think

[8] In my view, some clarification is in order here. In Woodward's claim that modularity is a feature of directed graphs, 'directed graphs' should be interpreted syntactically, not semantically.

of such an intervention as a divine intervention, where the make-up of the gametes is suddenly changed by some magical power. Derivatively one may think of the gametes being replaced by different gametes with some known and desired make-up. Barring problems relating to the **CG**-theoreticity of 'gametic make-up', this last procedure was feasible in the beginning of the twentieth century.) But it may also be assumed that this intervention would leave the relations between, e.g., GT_2 and PT_2, or between GC_2 and GT_3 unaltered. This is what Woodward calls "modularity". Disrupting (the relationship described by) the conditional probability table relating GT_1 and GC_1 should leave unaltered the conditional probability tables describing the other local mechanisms.

Woodward considers modularity as a precondition for representational adequacy. Hence the concept of modularity plays a semantic or definitional role in his framework.

> It is natural to suppose that if a system of equations [or a set of conditional probability tables] correctly and fully represents the causal structure of some system, then those equations [or conditional probability tables] should be modular. (Woodward, 2003b, 48)

It remains to be seen whether this claim should be endorsed no matter what.

10.5 Modularity and causal pragmatic laws

Two main motivations underlying the modularity assumption may be distinguished. Firstly, modular systems are well suited for manipulative policy. Secondly, it has been argued that modularity implies the causal Markov condition and that this makes it worth arguing for.

Modular systems of generalizations are well suited for manipulative policy. If an intervention on (the magnitude corresponding to) a variable does only affect (the local mechanism described by) one invariant generalization, while leaving the other generalizations unchanged, the effects of manipulative policy can relatively easily be predicted. This is true, and it provides a good pragmatic reason to search for (or to hope for) modular representations. But by itself it does not allow one to conclude that if a representation is not modular, it does not fully represent the causal structure of some system.

A second motivation underlying the modularity assumption seems to go in that direction. It is supposed to underpin the causal Markov condition (Hausman and Woodward, 1999). This motivation has come under attack.

Nancy Cartwright (2002; 2006) has repeatedly argued that modularity does not imply the causal Markov condition and Hausman and Woodward (2004) have tried to defend their thesis. Whether or not Hausman and Woodward are right in claiming that modularity implies the causal Markov condition is a difficult matter and the discussion does not seem to be closed yet. But to a large extent it is immaterial with regard to the previous chapters. Suppose that modularity does not imply the causal Markov condition. This would not compromise the latter. The causal Markov condition has independent grounding. It is a fallible, but heuristically useful tool, together with the faithfulness assumption, for discovering causal pragmatic regularities. Causal pragmatic regularities are described by causal pragmatic generalizations. These are invariant under some range of interventions. But whereas invariance is closely linked to modularity, they are nevertheless two distinct concepts (cf. supra). My concepts of cP-law and cP-regularity incorporate the former, but not the latter.

Suppose now that modularity does imply the causal Markov condition. Again, this would not compromise the latter. Objections to the modularity assumption would not carry over to the causal Markov condition. On the contrary, modularity implying the causal Markov condition would be interesting from a practical point of view. If modular systems of generalizations are particularly useful for predicting the outcome of manipulative policy (as I think they are) then it would be comforting to know that the variables in these systems satisfy the causal Markov condition. The most interesting systems would be the most 'visible' ones.

Nevertheless, we should not impose our wishes upon reality. The principles of classical genetics are modular.[9] Whether other sets of generalizations describing other causal structures are modular better remains an open question. Firstly, the issue of modularity does not directly affect either invariance or the causal Markov condition. Secondly, if taken as part of the definition of causality, modularity would exclude the existence of plastic or adaptively evolving causal structures. Such structures seem to be abundant in many biological (and social, and economic, ...) domains.[10]

10.6 Conclusion

Laws of nature and causality in the special sciences are well captured by the notions of P-regularity, P-law, cP-regularity and cP-law. These concepts

[9] This I illustrated in section 10.4. I did not establish it in detail.

[10] See Mitchell (2003a) for several examples of dynamically complex biological systems. See also Wagner (1999) for the problem of causal inference in complex (biological) systems.

are related to several others, which may not always be innocent: invariance under interventions, the causal Markov condition, the faithfulness condition, and modularity. The causal Markov condition and faithfulness play a central role in causal discovery, but we should always keep in mind they are fallible assumptions. The notion of invariance I took as central to the definition of causal pragmatic laws and regularities. To my opinion this was fruitful, but it implies that we rule out causes that can never be used to manipulate their effects – even in ideal cases. That is a consequence I am ready to bear. But I am reticent to treat modularity on a par. There are good reasons to prefer modular representations to non-modular ones. But from this it does not follow that modularity should be incorporated in our definition of causality or causal pragmatic laws.

Bibliography

Antonucci, A. and Zaffalon, M. (LSCN). Locally specified credal networks. Unpublished manuscript.

Armstrong, D. (1983). *What is a Law of Nature?* Cambridge University Press, Cambridge.

Balzer, W. and Dawe, C. (1986a). Structure and comparison of genetic theories I. *The British Journal for the Philosophy of Science*, 37(1):55–69.

Balzer, W. and Dawe, C. (1986b). Structure and comparison of genetic theories II. *The British Journal for the Philosophy of Science*, 37(2):177–191.

Balzer, W. and Dawe, C. (1997). *Models for Genetics*. Peter Lang, Frankfurt.

Balzer, W. and Lorenzano, P. (2000). The logical structure of classical genetics. *Journal for General Philosophy of Science*, 31(2):243–266.

Balzer, W., Moulines, C. U., and Sneed, J. D. (1987). *An Architectonic for Science. The Structuralist Program*. Reidel, Dordrecht.

Batens, D. (2000). Towards the unification of inconsistency handling mechanisms. *Logic and Logical Philosophy*, 8:5–31.

Batens, D. (2001). A general characterization of adaptive logics. *Logique et Analyse*, 173–175:45–68.

Batens, D. (2004). The need for adaptive logics in epistemology. In Gabbay, D., Rahman, S., Symons, J., and Van Bendegem, J.-P., editors, *Logic, Epistemology and the Unity of Science*, pages 459–485. Kluwer, Dordrecht.

Batens, D. (2007). A universal logic approach to adaptive logics. *Logica Universalis*, 1:221–242.

Bateson, W. (1900). Problems of heredity as a subject for horticultural investigation. *Journal for the Royal Horticultural Society*, 25:54–61. Electronically reprinted by the Electronic Scolarly Publishing Project (www.esp.org; page numbers ranging 1–10).

Bateson, W. (1902). *Mendel's Principles of Heredity. A Defence*. Cambridge University Press, London.

Beatty, J. (1995). The evolutionary contingency thesis. In Wolters, G. and Lennox, J., editors, *Concepts, Theories, and Rationality in the Biological Sciences.*, pages 45–81. University of Pittsburgh Press, Pittsburgh.

Beatty, J. (1997). Why do biologists argue like they do? *Philosophy of Science*, 64(4, supplement):S432–S443.

Bechtel, W. and Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36:421–441.

Bechtel, W. and Richardson, R. C. (1993). *Discovering Complexity. Decomposition and Localization as Strategies in Scientific Research.* Princeton University Press, Princeton.

Beckers, T., Houwer, J. D., Miller, R. R., and Urushihara, K. (2006). Reasoning rats: Forward blocking in pavlovian animal conditioning is sensitive to constraints of causal inference. *Journal of Experimental Psychology: General*, 135(1):92–102.

Beed, C. and Beed, C. (2000). Is the case for social science laws strengthening? *Journal for the Theory of Social Behaviour*, 30(2):131–153.

Bigelow, J. and Pargetter, R. (1987). Functions. *The Journal of Philosophy*, 84(4):181–196.

Billingsley, P. (1979). *Probability and Measure*. John Wiley and Sons, New York.

Bogen, J. (2005). Regularities and causality; generalizations and causal explanations. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36:397–420.

Bogen, J. and Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, XCVII(3):303–352.

Boolos, G. S. and Jeffrey, R. J. (2002). *Computability and Logic*. Cambridge University Press, Cambridge, 4th edition.

Brandon, R. N. (1997). Does biology have laws? The experimental evidence. *Philosophy of Science*, 64(4, supplement):S444–S457.

Bulmer, M. (1998). Galton's Law of Ancestral Heredity. *Heredity*, 81:579–585.

Bulmer, M. (2003). *Francis Galton. Pioneer of Heredity and Biometry*. The John Hopkins University Press, Baltimore, London.

Callebaut, W. and Rasskin-Gutman, D., editors (2005). *Modularity. Understanding the Development and Evolution of Natural Complex Systems*. MIT press, Cambridge (Mass.).

Carroll, J. W. (1994). *Laws of Nature*. Cambridge University Press, Cambridge.

Carroll, J. W., editor (2004). *Readings on Laws of Nature*. University of Pittsburgh Press, Pittsburgh.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Clarendon Press, Oxford.

Cartwright, N. (1994). *Nature's Capacities and Their Measurement*. Oxford University Press, New York.

Cartwright, N. (2001). What is wrong with Bayes nets? *The Monist*, 84(2):242–264.

Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: Comments on Hausman and Woodward. *The British Journal for the Philosophy of Science*, 53:411–453.

Cartwright, N. (2006). From metaphysics to method: Comments on manipulability and the causal Markov condition. *The British Journal for the Philosophy of Science*, 57:197–218.

Christie, M. (1994). Philosophers versus chemists concerning 'laws of nature'. *Studies in the History and Philosophy of Science*, 25(4):613–29.

Correns, C. (1900). G. Mendel's Laws concerning the behavior of progeny of varietal hybrids. In Stern and Sherwood (1966), pages 119–132.

Correns, C. (1903). Über die dominierenden Merkmale der Bastarde. *Berichten der deutschen botanischen Gesellschaft*, 21:133–147.

Cozman, F. G. (2005). Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39:167–184.

Craig, E., editor (1998). *Routledge Encyclopedia of Philosophy*. Routledge, London.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, 68:53–74.

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, 153:355–376.

Craver, C. F. (2007). *Explaining the Brain : Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press, London.

Craver, C. F. and Darden, L. (2005). Introduction: Mechanisms in biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(1):233–244.

Cuénot, L. (1902). La loi de Mendel et l'hérédité de la pigmentation chez les souris. *Archives de Zoologie Expérimentale et Générale*, 3 séries, notes:xxvii–xxx.

Cuénot, L. (1904). L'hérédité de la pigmentation chez les souris (3 note). *Archives de Zoologie Expérimentale et Générale*, 4 séries, notes et revue:xlv–lvi.

Cuénot, L. (1905). Les races pures et leurs combinaisons chez les souris. *Archives de Zoologie Expérimentale et Générale*, 4 séries, tome III, notes et revue:cxxiii–cxxxii.

Darden, L. (1991). *Theory Change in Science: Strategies from Mendelian Genetics*. Oxford University Press, Oxford.

Darden, L. (2002). Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining. *Philosophy of Science*, 69:S354–S365.

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(1):1–31.

De Regt, H. (2004). Discussion note: making sense of understanding. *Philosophy of Science*, 71:98–109.

De Regt, H. and Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese*, 144:137–170.

de Vries, H. (1900a). The law of segregation of hybrids. [Das Spaltungsgesetz der Bastarde (Preliminary communication)]. In Stern and Sherwood (1966), pages 107–117.

de Vries, H. (1900b). Sur la loi de disjonction des hybrides. *Comptes Rendus Hebdomadaires des Sances de l'Académie des Sciences*, 130:845–847.

de Vries, H. (1950). Concerning the law of segregation of hybrids. *Genetics*, 35(5, pt 2):30–32.

Dewey, J. (1889). Galton's statistical methods. *Publications of the American Statistical Association*, 1(7):331–334.

Earman, J., Glymour, C., and Mitchell, S. (2002). Ceteris paribus lost. *Erkenntnis*, 57(3):281–301.

Fara, M. (Fall 2006). Dispositions. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*.

Feller, W. (1968 (1950)). *An Introduction to Probability Theory and Its Applications*, volume I. 3rd edition.

Fisher, R. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52:399–433.

Fisher, R. (1936). Has Mendel's work been rediscovered? In Stern and Sherwood (1966), pages 139–172.

Franklin, A. (1990). *Experiment, Right or Wrong*. Cambridge University Press, Cambridge.

Franklin, A. (1995). Laws and experiment. In Weinert (1995), pages 191–207.

Franklin, A. (Summer 2003). Experiment in physics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*.

Frigg, R. (2006). Models in science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*.

Galison, P. (1987). *How Experiments End*. University of Chicago Press, Chicago.

Galton, F. (1871). Experiments in Pangenesis, by breeding from rabbits of a pure variety, into whose circulation blood taken from other varieties had previously been largely transfused. *Proceedings of the Royal Society*, 19:393–410.

Galton, F. (1877). Typical laws of heredity. *Nature*, 15:492–5, 512–4, 532–3.

Galton, F. (1885). Section h; anthropology; opening address. *Nature*, 32:507–510.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute*, 15:246–263.

Galton, F. (1889). *Natural Inheritance*. Macmillan and Co., London.

Galton, F. (1892 (1869)). *Hereditary Genius*. Macmillan and Co., London, 2nd edition.

Galton, F. (1897). The average contribution of each several ancestor to the total heritage of the offspring. *Proceedings of the Royal Society*, LXI:401–413.

Galton, F. (1898). A diagram of heredity. *Nature*, 57:293.

Geiger, D. and Pearl, J. (1993). Logical and algorithmic properties of conditional independence and graphical models. *The Annals of Statistics*, 21(4):2001–2021.

Giere, R. N. (1997). *Understanding Scientific Reasoning*. Harcourt Brace college publishers, Fort Worth (Tex.).

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44:49–71.

Glennan, S. S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(3):S342–S353.

Glennan, S. S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36:443–464.

Goodman, N., editor (1973). *Fact, Fiction, and Forecast.* Bobbs-Merrill, Indianapolis (Ind.), 3rd edition.

Gopnik, A. et al. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, 111(1):3–32.

Hacking, I. (1983). *Representing and Intervening. Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, Cambridge.

Hacking, I. (1990). *The Taming of Chance*. Cambridge University Press, Cambridge.

Haenni, R. and Hartmann, S. (2006). *Causality, Uncertainty and Ignorance*. Special issue of *Minds and Machines* (vol. 16, no. 3).

Hanks, S. and Besnard, P., editors (1995). *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, San Francisco.

Hausman, D. M. (1998). *Causal Asymmetries*. Cambridge University Press, Cambridge.

Hausman, D. M. and Woodward, J. (1999). Independence, invariance and the causal Markov condition. *The British Journal for the Philosophy of Science*, 50:521–583.

Hausman, D. M. and Woodward, J. (2004). Modularity and the causal Markov condition: a restatement. *The British Journal for the Philosophy of Science*, 55:147–161.

Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press, New York.

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58:295–300.

Hodges, W. (1998). Model theory. In Craig (1998), pages 436–442 (vol. 6).

IARC (2006). Preamble to the IARC monographs on the evaluation of carcinogenic risks to humans. http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In Kitcher and Salmon (1989), pages 410–505.

Kitcher, P. and Salmon, W. C., editors (1989). *Scientific Explanation.*, volume XIII of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. The Guilford Presss, New York, 2nd edition.

Klug, W. and Cummings, M. (1997). *Concepts of Genetics*. Prentice-Hall, New Jersey, 5th edition.

Klug, W., Cummings, M., and Spencer, C. (2006). *Concepts of Genetics*. Pearson Prentice-Hall, New Jersey, 8th edition.

Korb, K. B. and Nyberg, E. (2006). The power of intervention. In Haenni and Hartmann (2006), pages 289–302. Special issue of *Minds and Machines* (vol. 16, no. 3).

Kuhn, T. S. (1962 (1996)). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 3rd edition.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill International Edition, Singapore, 5th edition.

Legator, M. and Morris, D. (2003). What did Sir Bradford Hill really say? *Archives of Environmental Health*, 58(11):718–720.

Leonelli, S. (2007). Weed for thought. Using Arabidopsis thaliana to understand plant biology. Unpublished PhD thesis.

Leuridan, B. (2007a). Galton's blinding glasses. Modern statistics hiding causal structure in early theories of inheritance. In Russo, F. and Williamson, J., editors, *Causality and Probability in the Sciences*, Texts In Philosophy series, pages 243–262. College Publications, London.

Leuridan, B. (2007b). Supervenience: its logic and its inferential role in classical genetics. *Logique et Analyse*, 198:147–171.

Leuridan, B. (CDPI). Causal discovery and the problem of ignorance. An adaptive logic approach. *Journal of Applied Logic*. To appear.

Leuridan, B., Weber, E., and Van Dyck, M. (POL). The practical value of spurious correlations: Selective versus manipulative policy. *Analysis*. To appear.

Lindenmayer, A. and Simon, N. (1980). The formal structure of genetics and the reduction problem. In Asquith, P. and Giere, R., editors, *PSA 1980*. Philosophy of Science Association, East Lansing/Michigan.

Lloyd, E. E. (1998). Models. In Craig (1998), pages 443–447 (vol. 6).

Lorenzano, P. (1995). *Geschichte und Struktur der klassischen Genetik*. Peter Lang, Frankfurt am Main.

Lorenzano, P. (2007). The influence of genetics on philosophy of science. Classical genetics and the structuralist view of theories. In Fagot-Largeault, A., Rahman, S., and Torres, J. M., editors, *The Influence of Genetics on Contemporary Thinking*, pages 97–113. Springer, Dordrecht.

Machamer, P. (2004). Activities and causation: the metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science*, 18(1):27–39.

Machamer, P., Darden, L., and Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1):1–25.

Mäki, U. (1990). Scientific realism and austrian explanation. *Review of Political Economy*, 2(3):310–344.

Mäki, U. (2001). Explanatory unification : double and doubtful. *Philosophy of the Social Sciences*, 31(4):488–506.

Makinson, D. (2005). *Bridges from Classical to Nonmonotonic Logic*. Texts in Computing series. King's College Publications, London.

Marchionni, C. (2005). Unity, plurality and explanation. the case of geographical economics and its neighbours. Unpublished PhD thesis.

Meek, C. (1995a). Causal inference and causal explanation with background knowledge. In Hanks and Besnard (1995), pages 403–410.

Meek, C. (1995b). Strong completeness and faithfulness in Bayesian networks. In Hanks and Besnard (1995), pages 411–418.

Meheus, J. (2003). Do we need paraconsistency in commonsense reasoning? In Delrieux, C. and Legris, J., editors, *Computer Modeling of Scientific Reasoning*, pages 135–145. Universidad Nacional Del Sur, Bahia Blanca, Argentina.

Meijer, O. G. (1983). The essence of Mendel's discovery. In Orel, V. and Matalová, A., editors, *Gregor Mendel and the Foundation of Genetics*, pages 123–178. The Mendelianum of the Moravian Museum, Brno.

Mendel, G. (1865). Experiments on plant hybrids. In Stern and Sherwood (1966), pages 1–48.

Mendel, G. (1865 (1933)). *Versuche über Pflanzenhybriden.* Ostwald's Klassiker der Exakten Wissenschaften. Akademische Verlagsgesellschaft, Leipzig.

Mendel, G. (1869). On Hieracium-hybrids obtained by artificial fertilisation [Über einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde]. In Stern and Sherwood (1966), pages 49–55.

Mitchell, S. D. (1997). Pragmatic laws. *Philosophy of Science*, 64(4, supplement):S468–S479.

Mitchell, S. D. (2000). Dimensions of scientific law. *Philosophy of Science*, 67(4):242–265.

Mitchell, S. D. (2002). Ceteris paribus – an inadequate representation for biological contingency. *Erkenntnis*, 57(3):329–350.

Mitchell, S. D. (2003a). *Biological Complexity and Integrative Pluralism.* Cambridge University Press, Cambridge.

Mitchell, S. D. (2003b). Contingent generalizations: Lessons from biology. In Mitchell (2003a), pages 147–161.

Mitchell, S. D. (2006). Modularity – more than a buzzword? *Biological Theory*, 1(1):98–101.

Morgan, M. S. (2003). Experiments without material interventions. Model experiments, virtual experiments, and virtually experiments. In Radder (2003a), pages 216–235.

Morgan, M. S. and Morrison, M., editors (1999). *Models as Mediators. Perspectives on Natural and Social Science.* Cambridge University Press, Cambridge.

Morgan, T. H. (1909). What are factors in Mendelian explanations? *American Breeders' Association Reports*, 5:365–368.

Morgan, T. H. (1910). Chromosomes and heredity. *The American Naturalist*, 44:449–496.

Morgan, T. H. (1919). *The Physical Basis of Heredity.* Lippincott, Philadelphia.

Morgan, T. H. (1926 (1928)). *The Theory of the Gene*. Yale University Press, New Haven, revised and enlarged edition.

Morgan, T. H., Sturtevant, A., Muller, H., and Bridges, C. (1915). *The Mechanism of Mendelian Heredity*. Henry Holt and Company, New York.

Morrison, M. and Morgan, M. S. (1999a). Introduction. In Morgan and Morrison (1999), pages 1–9.

Morrison, M. and Morgan, M. S. (1999b). Models as mediating instruments. In Morgan and Morrison (1999), pages 10–37.

Mumford, S. (2004). *Laws in Nature*. Routledge Studies in Twentieth-Century Philosophy. Routledge, Oxfordshire.

Nagel, E. (1961). *The Structure of Science. Problems in the Logic of Scientific Explanation*. Routledge and Kegan Paul, London.

Neapolitan, R. E. (2004). *Learning Bayesian Networks*. Prentice Hall Series in Artificial Intelligence. Pearson Education.

Orel, V. (1996). *Gregor Mendel. The First Geneticist*. Oxford University Press, Oxford.

Paris, J. B. (2005). On filling-in missing conditional probabilities in causal networks. *Fuzziness and Knowledge-Based Systems*, 13(3):263–280.

Paris, J. B. and Vencovská, A. (1997). In defense of the Maximum Entropy Inference Process. *International Journal of Approximate Reasoning*, 17:77–103.

Paris, J. B. and Vencovská, A. (2001). Common sense and stochastic independence. In Corfield, D. and Williamson, J., editors, *Foundations of Bayesianism*, pages 204–240. Kluwer, Dordrecht.

Paz, A. (2006). A property of independency relations induced by probabilistic distributions with binary variables. *Fundamenta Informaticae*, 73:229–236.

Pearl, J. (2000). *Causality. Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Pearson, K. (1900). *The Grammar of Science*. Adam and Charles Black, London, 2nd edition.

Pearson, K. (1957 (1911)). *The Grammar of Science*. Meridian Books, New York, reissue of the 3rd edition.

Phillips, C. and Goodman, K. (2004). The missed lessons of Sir Austin Bradford Hill. *Epidemiological Perspectives and Innovations*, 1(3).

Pietroski, P. and Rey, G. (1995). When other things aren't equal: Saving ceteris paribus laws from vacuity. *The British Journal for the Philosophy of Science*, 46:81–110.

Prior, E. W., Pargetter, R., and Jackson, F. (1982). Three theses about dispositions. *American Philosophical Quarterly*, 19(3):251–257.

Psillos, S. (2002). *Causation and Explanation*. Acumen, Chesham.

Psillos, S. (2004). A glimpse of the secret connexion: Harmonizing mechanisms with counterfactuals. *Perspectives on Science*, 12(3):288–319.

Radder, H., editor (2003a). *The Philosophy of Scientific Experimentation*. University of Pittsburgh Press, Pittsburgh.

Radder, H. (2003b). Technology and theory in experimental science. In Radder (2003a), pages 152–173.

Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science*, 45:206–226.

Railton, P. (1981). Probability, explanation, and information. *Synthese*, 48:233–256.

Rizzotti, M. and Zanardo, A. (1986). Axiomatization of genetics I & II. *Journal of Theoretical Biology*, 118:61–71, 145–152.

Roberts, J. T. (2004). There are no laws of the social sciences. In Hitchcock, C., editor, *Contemporary Debates in Philosophy of Science*, pages 151–167. Blackwell Publishing, Malden MA.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton, New Jersey.

Salmon, W. C. (1989). Four decades of scientific explanation. In Kitcher and Salmon (1989), pages 3–219.

Salmon, W. C. (1998). *Causality and Explanation*. Oxford University Press, New York.

Schaffner, K. F. (1993). *Discovery and Explanation in Biology and Medicine*. The University of Chicago Press, Chicago.

Schiff, F. (1930). Die Vererbungsweise der Faktoren M und N von Landsteiner und Levine. *Journal of Molecular Medicine*, 9(42):1432–1440.

Simon, H. A. (1954). Spurious correlation: a causal interpretation. *Journal of the American Statistical Association*, 49:467–479.

Sloan, P. R. (2000). Mach's phenomenalism and the British reception of mendelism. *Comptes Rendus de l'Académie des Sciences. Série III – Sciences de la Vie – Life Sciences*, 323(12):1069–1079.

Sober, E. (1997). Two outbreaks of lawlessness in recent philosophy of biology. *Philosophy of Science*, 64(4, supplement):S458–S467.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Cambridge, Massachusetts.

Spohn, W. (1994). On the properties of conditional independence. In Humphreys, P., editor, *Patrick Suppes: Scientific Philosopher*, volume I, pages 173–196. Kluwer Academic Publishers, Dordrecht.

Stansfield, W. (1983). *Theory and Problems of Genetics*. Mc Graw-Hill Book Company, New York, 2nd edition.

Steel, D. (2006). Homogeneity, selection, and the faithfulness condition. *Minds and Machines*, 16:303–317.

Stern, C. and Sherwood, E. R., editors (1966). *The Origin of Genetics: A Mendel Source Book*. Freeman, San Francisco.

Stigler, S. M. (1986). *The History of Statistics. The Measurement of Uncertainty before 1900*. The Belknap Press of Harvard University Press, Cambridge (Mass.).

Stigler, S. M. (2002). *Statistics on the Table. The History of Statistical Concepts and Methods*. Harvard University Press, Cambridge (Mass.).

Studený, M. (1992). Conditional independence relations have no finite complete characterization. In Kubík, S. and Vísek, J., editors, *Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Held at Prague 1990*, volume B, pages 377–396. Academia, Prague.

Suppe, F. (1974). The search for philosophical understanding of scientific theories. In Suppe, F., editor, *The Structure of Scientific Theories*, pages 3–241. University of Illinois Press, Urbana.

Suppes, P. (1969a). A comparison of the meaning and uses of models in mathematics and the empirical sciences. In Suppes (1969c), pages 10–23. Reprinted from *Synthese*, 12 (1960), 287–301.

Suppes, P. (1969b). Models of data. In Suppes (1969c), pages 24–35. Reprinted from *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* (ed. by E. Nagel, P. Suppes, and A. Tarski), Stanford University Press, Stanford, California, 1962, pp. 252–261.

Suppes, P., editor (1969c). *Studies in the Methodology and Foundations of Science. Selected Papers from 1951 to 1969*. D. Reidel Publishing Company, Dordrecht – Holland.

Tabery, J. G. (2004a). The "evolutionary synthesis" of George Udny Yule. *Journal of the History of Biology*, 37:73–101.

Tabery, J. G. (2004b). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of Science*, 71(1):1–15.

Tschermak, E. (1950). Concerning artificial crossing in *Pisum sativum*. *Genetics*, 35(5, pt 2):42–47.

Van Dyck, M. (2004). Causal discovery using adaptive logics. Towards a more realistic heuristics for human causal learning. *Logique et Analyse*, 185–188:5–32.

Venn, J. (1889). Critical notices. *Mind*, XIV(55):414–420.

Verma, T. and Pearl, J. (1988). Causal networks: semantics and expressiveness. In Shachter, R., Levitt, T., and Kanal, L., editors, *Uncertainty in Artificial Intelligence*, volume 4, pages 69–76. Elsevier Science Publishers, North-Holland.

Verma, T. and Pearl, J. (1992). An algorithm for deciding if a set of observed independencies has a causal explanation. In Dubois, D., Wellman, M. P., D'Ambrosio, B., and Smets, P., editors, *Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence*, pages 323–330. Morgan Kaufmann, San Francisco.

Wagner, A. (1999). Causality in complex systems. *Biology and Philosophy*, 14:83–101.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Monographs on Statistics and Applied Probability, 42. Chapman and Hall, London.

Walsh, D. (1996). Fitness and function. *The British Journal for the Philosophy of Science*, 47(4):553–574.

Weber, E. (1999). Unification: What is it, how do we reach it and why do we want it? *Synthese*, 118:479–499.

Weber, M. (2006). Causes without mechanisms: Experimental regularities, physical laws, and neuroscientific explanation. Unpublished manuscript from the PSA 2006 Symposia, available at http://philsci-archive.pitt.edu/archive/00003287/.

Weinert, F., editor (1995). *Laws of Nature. Essays on the Philosophical, Scientific and Historical Dimensions*. de Gruyter, Berlin.

Weldon, W. (1902). Mendels laws of alternative inheritance in peas. *Biometrika*, 1:228–254.

Williamson, J. (2005). *Bayesian Nets and Causality. Philosophical and Computational Foundations*. Oxford University Press, New York.

Woodger, J. H. (1952). *Biology and Language. An Introduction to the Methodology of the Biological Sciences including Medicine*. Cambridge University Press, Cambridge.

Woodger, J. H. (1959). Studies in the foundations of genetics. In L. Henkin, P. S. and Tarski, A., editors, *The Axiomatic Method*. North-Holland, Amsterdam.

Woodward, J. (1997). Explanation, invariance, and intervention. *Philosophy of Science*, 64(4, supplement):S26–S41.

Woodward, J. (2002a). There is no such thing as a ceteris paribus law. *Erkenntnis*, 57(3):302–328.

Woodward, J. (2002b). What is a mechanism? A counterfactual account. *Philosophy of Science*, 69(3):S366–S377.

Woodward, J. (2003a). Experimentation, causal inference, and instrumental realism. In Radder (2003a), pages 87–118.

Woodward, J. (2003b). *Making Things Happen. A Theory of Causal Explanation*. Oxford University Press, New York.

Wright, S. (1966). Mendel's ratios. In Stern and Sherwood (1966), pages 173–175.

Yates, F. and Mather, K. (1963). Ronald Aylmer Fisher. 1890–1962. *Biographical Memoirs of Fellows of the Royal Society of London*, (9):91–120.

Index