Chapter 8
Loops and Networks

Abstract
The analysis so far concerned only one-dimensional epistemic chains. In this chapter two extensions are investigated. The first treats loops rather than chains. We show that generally, i.e. in what we have called the usual class, infinite loops yield the same value for the target as do infinite chains; it is only in the exceptional class that the values differ. The second extension involves multi-dimensional networks, where the chains fan out in many different directions. As it turns out, the uniform version of the networks yields the fractal iteration of Mandelbrot. Surprising as it may seem, justificatory systems that mushroom out greatly resemble fractals.

8.1 Tortoises and Serpents

In 1956 Wilfrid Sellars famously diagnosed the malaise of epistemology as an unpalatable either/or:

> One seems forced to choose between the picture of an elephant which rests on a tortoise (What supports the tortoise?) and the picture of a great Hegelian serpent of knowledge with its tail in its mouth (Where does it begin?). Neither will do.[1]

Up to this point our focus has been on finite and infinite chains of propositions. We looked, as it were, at an elephant which rests on a tortoise, which in turn might rest on a second tortoise, and so on, without end. *Pace* Sellars' pessimism, we have seen that such structures are not particularly problematic if one takes seriously that the relation of support is probabilistic.

[1] Sellars 1956, 300.

D. Atkinson, J. Peijnenburg, *Fading Foundations*, Synthese Library 383,
DOI 10.1007/978-3-319-58295-5_8

There are now two ways in which we could extend our investigation and go beyond one-dimensional chains. The first is to keep the one-dimensionality, but to look at loops rather than chains: this would take us to the second horn of Sellars's dilemma, where knowledge is pictured as Kundalini swallowing its own tail. The other way is to give up one-dimensionality altogether and to study multi-dimensional networks. This would take us to the coherentist caucus in epistemology, or rather to an infinitist version of it, in which ultimately the network stretches out indefinitely in infinitely many directions. It might seem that such a version will be especially vulnerable to the standard objection to coherentism, according to which coherentist networks of knowledge hang in the air without making contact with the world. Indeed, as Richard Fumerton noted, if we worry about "the possibility of completing one infinitely long chain of reasoning, [we] should be downright depressed about the possibility of completing an infinite number of infinitely long chains of reasoning".[2]

Remarkably enough however, the opposite is the case. Since the connections between the propositions in the network are probabilistic in character, we are dealing with conditional probabilities. As we explained in Section 4.4, the conditional probabilities together carry the empirical thrust, and this is even more so in a multi-dimensional system than in a structure of only one dimension, for the simple reason that now there are more conditional probabilities that may be linked to the world.

Extending the chains to networks thus enables us to catch it all: to develop a form of coherentism which not only is infinitist, but also acknowledges the foundationalist maxim that a body of knowledge worthy of the name must somehow make contact with the world.[3]

We start in Section 8.2 by discussing one-dimensional loops. We will see that, if justification is interpreted probabilistically, then it is in general un-

[2] Fumerton 1995, 57.

[3] Thus we do not have many quibbles with William Roche when he argues that foundationalism, if suitably generalized, can be reconciled with infinite regresses of probabilistic support (Roche 2016). Much depends on what is meant by foundationalism: as we indicated in Section 4.4, we do not want to become embroiled in a verbal dispute. Some commentators write as if foundationalism were the sole guardian of empirical credibility and connection to the world. Although others might find that position unduly imperialistic, we do not object to being called foundationalists in that sense. We have no issue with a form of foundationalism that takes into account fading foundations and the related concept of trading off as it is applied to doxastic justificatory chains. Our concern is less about the classification of our results than about the results themselves.

problematic to maintain that a target is justified by a loop. In Section 8.3 we turn to multi-dimensional networks, where the tentacles stretch out in many different directions. In Section 8.4 we explain that such a multi-dimensional network takes on a very interesting and intriguing shape when it goes to infinity. Surprising and somewhat strange as it may sound, if epistemic justification is interpreted probabilistically, and if we accept that it can go on without end, then justification is tantamount to constructing a fractal of the sort that Benoît Mandelbrot introduced many years ago.

In the final section we explain what happens when the multi-dimensionality springs from the *connections* in the network rather than from the nodes, i.e. when it originates from the conditional probabilities rather than from the unconditional ones. We shall see that in a generalized sense the Mandelbrot construction is preserved.[4]

8.2 One-Dimensional Loops

Finite loops embody the simplest coherentist system. What about infinite ones? It seems that an infinite loop cannot really be called a loop, since there is no end of the tail that the Hegelian serpent can swallow. A loop after all involves a repeat of the same; it may be long, indeed more than cosmologically long, but it seems that it may not be infinite, on pain of having no repetition at all. Even Henri Poincaré, when he formulated his recurrence theorem, had to assume that the universe is finite in spatial extent and of finite energy.

However, from the fact that a finite loop differs from an infinite 'loop', it does not follow that an infinite loop is in fact an infinite chain. Our investigation in this section will explain that such a conclusion would be unwarranted. In what we have called the usual class, the infinite loop indeed produces the same result as does the corresponding infinite chain; but in the exceptional class infinite loops and infinite chains yield different results, as we shall show.

We saw in Chapter 3 that the probability of the target in a finite linear chain can be written as in (3.20), where we have reinstated q in place of A_0:

$$P(q) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 + \ldots + \gamma_0\gamma_1\ldots\gamma_{m-1}\beta_m + \gamma_0\gamma_1\ldots\gamma_m P(A_{m+1}) \,.$$

[4] Section 8.2 in this chapter, about the loops, is based on Atkinson and Peijnenburg 2010a; Sections 8.3 and 8.4, which deal with networks, are based on Atkinson and Peijnenburg 2012.

The general formulation of a finite loop with $m+1$ propositions has a similar form, except that the $(m+1)$st proposition is q itself. Mathematically, there is no problem if we insert $A_{m+1} = q$ into the above equation to yield

$$P(q) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 + \ldots + \gamma_0\gamma_1 \ldots \gamma_{m-1}\beta_m + \gamma_0\gamma_1 \ldots \gamma_m P(q)\,,$$

for this yields

$$P(q) = \frac{\beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 + \ldots + \gamma_0\gamma_1 \ldots \gamma_{m-1}\beta_m}{1 - \gamma_0\gamma_1 \ldots \gamma_m}\,, \tag{8.1}$$

which is well-defined, on condition that $\gamma_0\gamma_1 \ldots \gamma_m$ is not equal to unity.[5] With that proviso, the solution demonstrates the viability of the coherentist scenario in its simplest form, that of a finite one-dimensional loop.

The fact that a self-supporting finite loop or ring makes good mathematical sense is of course not enough. Does it also make sense elsewhere? Can a loop that closes upon itself occur in reality? A temporal example of such a loop is difficult to come by in the real world, but it can occur in the science fiction of time travel. Let q be a proposition stating that young Biff decides in 1955 to use the 2015 edition of the sports almanac, A_1 a proposition asserting that he continues his successful career as bettor until 2015, and A_2 a proposition explaining how old Biff succeeds in borrowing Doc Brown's time machine in 2015, and returns to 1955 in order to give the almanac to his younger self. $A_3 = q$ would then be a proposition stating that young Biff decides in 1955 to use the 2015 edition of the sports almanac ... and so on.

In fact, the events need not follow one another in time. Consider the following three propositions:

C: "Peter read parts of the *Critique of Pure Reason*".
P: "Peter is a philosopher".
S: "Peter knows that Kant defended the synthetic a priori".

Assuming that all philosophers read at least parts of the *Critique of Pure Reason* as undergraduates, if Peter is a philosopher, then he read parts of the *Critique*. Of course, even if he is not a philosopher, he may still have read Kant's magnum opus. If Peter knows that Kant defended the synthetic a priori, he very likely is a philosopher, whereas if he does not, he is probably not a philosopher, although of course he might be an exceptionally incompetent

[5] If $\gamma_0\gamma_1 \ldots \gamma_m = 1$, it follows that *each* γ_n is equal to one. But then all the α_n are equal to one also, and all the β_n are equal to zero, which is the condition of bi-implication. This already indicates that a loop does not make sense when entailment relations are involved.

one, not having understood anything of Kant or the *Critique*. Finally, if he read the *Critique*, he quite likely knows that Kant defended the synthetic a priori, whereas this is rather less likely if he never opened the book. Here then is a simple finite loop, consisting of a fixed number of links, namely three:

$$\mathrm{C} \longleftarrow \mathrm{P} \longleftarrow \mathrm{S} \longleftarrow \mathrm{C}\,, \tag{8.2}$$

where the arrow indicates that the proposition at the right-hand side probabilistically supports the one at the left.

We can make loop (8.2) nonuniform by investing the three propositions C, P and S with for example the following dissimilar values for the conditional probabilities:

$$\begin{array}{llll}
\mathrm{C:} & \alpha_0 = P(\mathrm{C}\,|\mathrm{P}) = 1\,; & \beta_0 = P(\mathrm{C}\,|\neg\mathrm{P}) = \frac{1}{10}; & \gamma_0 = \alpha_0 - \beta_0 = \frac{9}{10} \\
\mathrm{P:} & \alpha_1 = P(\mathrm{P}\,|\mathrm{S}\,) = \frac{9}{10}\,; & \beta_1 = P(\mathrm{P}\,|\neg\mathrm{S}\,) = \frac{1}{5}; & \gamma_1 = \alpha_1 - \beta_1 = \frac{7}{10} \\
\mathrm{S:} & \alpha_2 = P(\mathrm{S}\,|\mathrm{C}\,) = \frac{4}{5}\,; & \beta_2 = P(\mathrm{S}\,|\neg\mathrm{C}\,) = \frac{2}{5}; & \gamma_2 = \alpha_2 - \beta_2 = \frac{2}{5}.
\end{array}$$

Then the unconditional probabilities[6] are

$$\begin{aligned}
P(\mathrm{C}\,) &= \frac{\beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2}{1-\gamma_0\gamma_1\gamma_2} = 0.711 \\
P(\mathrm{P}\,) &= \frac{\beta_1 + \gamma_1\beta_2 + \gamma_1\gamma_2\beta_0}{1-\gamma_0\gamma_1\gamma_2} = 0.679 \\
P(\mathrm{S}\,) &= \frac{\beta_2 + \gamma_2\beta_0 + \gamma_2\gamma_0\beta_1}{1-\gamma_0\gamma_1\gamma_2} = 0.684\,.
\end{aligned}$$

In the above example the number of links was fixed: there were exactly three propositions. Here is an example in which the number of links, m, can be whatever one likes, showing the cogency of *any* finite loop. Consider again the example (3.21) in Section 3.5:

$$\alpha_n = 1 - \frac{1}{n+2} + \frac{1}{n+3}\,; \qquad \beta_n = \frac{1}{n+3}\,; \qquad \gamma_n = 1 - \frac{1}{n+2}\,.$$

[6] As they must, these numbers satisfy

$$P(\mathrm{C}\,) = \beta_0 + \gamma_0 P(\mathrm{P}\,) \qquad P(\mathrm{P}\,) = \beta_1 + \gamma_1 P(\mathrm{S}\,) \qquad P(\mathrm{S}\,) = \beta_2 + \gamma_2 P(\mathrm{C}\,)\,.$$

Incidentally, there is a good reason for considering a loop of at least three propositions. For in a ‘loop’ of two links only, there are only three independent unconditional probabilities, for example $P(q)$, $P(A_1)$ and $P(q \wedge A_1)$, whereas there are four conditional probabilities around the loop, $P(q|A_1)$, $P(q|\neg A_1)$, $P(A_1|q)$ and $P(A_1|\neg q)$, so there must be a relation between them. This difficulty does not arise for a loop of three links, for in this case there are seven independent unconditional probabilities and only six conditional probabilities around the loop. With more than three links on the loop the difference between the numbers of unconditional and conditional probabilities is even greater.

This example is nonuniform (i.e. the conditional probabilities, α_n and β_n, are not the same for different n), and it is in the usual class. It is shown in (A.18) in Appendix A.5 that Eq.(8.1) reduces to

$$P(q) = \frac{3}{4} - \frac{1}{4(m+3)}\,. \tag{8.3}$$

In Table 8.1 the values of $P(q)$ for the chain are reproduced in the first line, while the corresponding values for the loop, as specified in (8.3), are given in the second line. The difference between the two cases is that, while for the chain we had to specify a value for the probability of the ground, which we put equal to a half, for the loop no such specification is required.

Table 8.1 Probability of q for chain and loop $\qquad P(p) = \frac{1}{2}$ for chain
$\alpha_n = P(A_n|A_{n+1}) = 1 - \frac{1}{n+2} + \frac{1}{n+3} \qquad \beta_n = P(A_n|\neg A_{n+1}) = \frac{1}{n+3}$

Number of A_n	1	2	5	10	25	50	75	100	∞
$P(q)$ with chain	.625	.650	.688	.712	.732	.741	.744	.745	.750
$P(q)$ with loop	.688	.700	.719	.731	.741	.745	.747	.748	.750

The probability of the target rises smoothly as the chain, or the loop, becomes longer, eventually reaching the value of three-quarters for both the infinite chain and the infinite loop. As can be seen, the values of $P(q)$ for the loop converge somewhat more quickly than do those for the chain.

The agreement between the infinite chain and the infinite loop is not limited to this special model, for it is an attribute of any example in the usual class. This can be seen quite easily, for when the product $\gamma_0\gamma_1\ldots\gamma_m$ tends to zero as m goes to infinity, the loop (8.1) yields the infinite, convergent series

$$P(q) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 + \gamma_0\gamma_1\gamma_2\beta_3\ldots\,, \tag{8.4}$$

as for the infinite chain in the usual class.

The uniform case, in which the conditional probabilities are the same from link to link, forms an interesting special case, for then the value of $P(q)$ turns out to be always the same, no matter how many links there are in the loop. This can already be seen without doing the actual calculation. Since the propositions are uniformly connected round and round the loop *ad infinitum*, we can immediately understand why it should make no difference how many links there are: the value of $P(q)$ should be the same as that for an infinite, uniform loop. The actual calculation goes as follows: (8.1) becomes

$$P(q)=\frac{\beta(1+\gamma+\gamma^2+\ldots\gamma^m)}{1-\gamma^{m+1}}\,. \tag{8.5}$$

The finite geometrical series $1+\gamma+\gamma^2+\ldots\gamma^m$ is equal to $(1-\gamma^{m+1})/(1-\gamma)$, and on substituting this we see that the factor $(1-\gamma^{m+1})$ cancels, so

$$P(q)=\frac{\beta}{1-\gamma}=\frac{\beta}{1-\alpha+\beta}\,.$$

Indeed this does not depend on m at all, so the number of links may be finite, or infinite, with no change in the value of $P(q)$. It will be recognized that this value is precisely the same as that for the infinite, uniform chain (see Section 3.7).

So much for the usual class. What of the exceptional class, in which the infinite product of the γ's is not zero? As we have seen, here the chain fails, in the infinite limit, to produce a definite answer for the target probability. The infinite loop on the other hand yields a unique value. To illustrate this, consider again the example (3.25):

$$\beta_n=\frac{1}{(n+2)(n+3)} \qquad\qquad \gamma_n=\frac{(n+1)(n+3)}{(n+2)^2}=1-\frac{1}{(n+2)^2}\,.$$

We find now from (8.1) that

$$P(q)=\frac{3}{4}-\frac{1}{4(m+3)}\,, \tag{8.6}$$

as we explain in detail in Appendix A.6, and this has the perfectly definite limit $\frac{3}{4}$. Thus the infinite chain and the infinite loop only differ in the exceptional class. There the infinite chain fails to give a definite answer, but the infinite loop does so.[7]

8.3 Multi-Dimensional Networks

Most systems of epistemic justification are of course much more complicated than the one-dimensional chains and loops that we have considered so far. Certainly modern coherentism envisages many-dimensional nets of interlocking probabilistic relations. The concept of *justification trees* or *J-trees*

[7] The fact that this value of $P(q)$ is the same as that of the loop (8.3), in the usual class, is just a coincidence.

has been introduced as a graphic representation of the relation in such networks.[8] Figure 8.1 is an example of a very simple justification tree. This tree has two branches, with A_1 and A_1' as nodes on the one level, and A_2 and A_2' as nodes on a lower level. It should be read as: proposition q is justified by A_1 and A_1', A_1 is justified by A_2, and A_1' is justified by A_2'. In this section we shall describe what happens when we replace the finite or infinite one-dimensional probabilistic chain by a finite or infinite probabilistic network in two dimensions, along the lines of a justification tree.

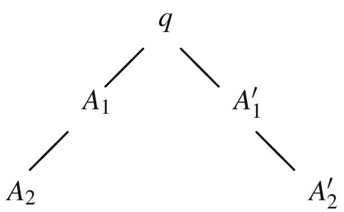

Fig. 8.1 Basic justification tree

We now make the tree more complicated by allowing that A_1 and A_1' are each supported by two, rather than by one proposition, as depicted in Fig. 8.2.

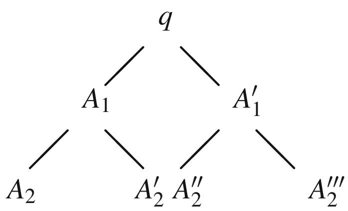

Fig. 8.2 Complex justification tree

Here A_1 is supported by A_2 and A_2'; and A_1' is supported by A_2'' and A_2'''. In their turn, A_2, A_2', A_2'', and A_2''' may each be supported by two propositions.

A complicated tree as in 8.2 could serve as a model for the propagation of genetic traits under sexual reproduction, in which the traits of a child

[8] See for example Sosa 1979; Clark 1988, 374-375; Alston 1989, 19-38; Cortens 2002, 25-26; Aikin 2011, 74.

are related probabilistically to those of both the mother and the father. Let $P(q)$ again be the unconditional probability that Barbara has trait T. This time Barbara is not a bacterium as in Section 3.7, where the reproduction was asexual. Rather she is now an organism with two parents, a father and a mother. For the purpose of fixing ideas it will prove convenient to talk about sexual reproduction and about fathers and mothers, but we should bear in mind that the formalism is of course much more general. Also, although we shall tell the story in terms of events, it should be kept in mind that everything we say applies to justificatory relations between propositions as well.

Since Barbara stems from two parents, the probability that she has T is determined by the characteristics of her mother and of her father. Rather than two reference classes (the mother having or not having T), we now have four: both the mother and the father have T, neither of them has it, the father has T but the mother does not, and the mother has T but the father does not. The analogue of the rule of total probability is

$$P(q) = \alpha_0 P(A_1 \wedge A'_1) + \beta_0 P(\neg A_1 \wedge \neg A'_1) + \gamma_0 P(A_1 \wedge \neg A'_1) + \delta_0 P(\neg A_1 \wedge A'_1)\,, \tag{8.7}$$

where A_1 represents Barbara's mother having T and A'_1 her father having T. Here α_0 means "the probability that Barbara has T, given that her mother and father both have T". The other conditional probabilities are analogously defined: β_0 corresponds to neither parent having T, and γ_0 and δ_0 to the two situations in which one parent does, and the other does not have T.

In the nth generation the corresponding expression is

$$P(A_n) = \alpha_n P(A_{n+1} \wedge A'_{n+1}) + \beta_n P(\neg A_{n+1} \wedge \neg A'_{n+1}) + \gamma_n P(A_{n+1} \wedge \neg A'_{n+1}) + \delta_n P(\neg A_{n+1} \wedge A'_{n+1})\,, \tag{8.8}$$

where A_n stands for one individual in the nth generation, A_{n+1} and A'_{n+1} for that individual's mother and father. The conditional probabilities are

$$\begin{aligned}
\alpha_n &= P(A_n|A_{n+1} \wedge A'_{n+1}) \\
\beta_n &= P(A_n|\neg A_{n+1} \wedge \neg A'_{n+1}) \\
\gamma_n &= P(A_n|A_{n+1} \wedge \neg A'_{n+1}) \\
\delta_n &= P(A_n|\neg A_{n+1} \wedge A'_{n+1})\,.
\end{aligned}$$

In order to iterate the two-dimensional (8.8), much as we did in the one-dimensional case, we now need more complicated relations for the unconditional probabilities. It is no longer sufficient to consider $P(A_1)$ and replace it by $\beta_0 + (\alpha_0 - \beta_0)P(A_2)$, and so on, for now we are dealing with the probability of a conjunction of two parents, A_1 and A'_1. Each of these parents has two parents, so we encounter in fact the probabilities of conjunctions of four individuals. This can be continued further and further, involving more and more progenitors, confronting us with a tree of increasing complexity.

Fortunately, however, we can often make simplifying assumptions. Here we will work under three simplifications:

1. **Independence.** The probabilities for the occurrence of the trait T in females and in males is independent of one another in any of the n generations:
$$P(A_{n+1} \wedge A'_{n+1}) = P(A_{n+1})P(A'_{n+1})\,.$$
This assumption seems reasonable in the genetic context; and it will also apply in many more general epistemological settings.

2. **Gender symmetry.** The probability of the occurrence of the trait T is the same for females and for males in any of the n generations:
$$P(A_n) = P(A'_n)\,.$$
Thus we only consider inheritable traits which are gender-independent, such as having blue eyes or being red-haired, and not, for example, having breast cancer or being taller than two metres. Similarly, in an epistemological context this assumption will sometimes, but not always be satisfied. With this assumption the prime can be dropped on A'_n, and in combination with the first assumption we obtain
$$P(A_{n+1} \wedge A'_{n+1}) = P(A_{n+1})P(A_{n+1}) = P^2(A_{n+1})\,.$$

3. **Uniformity.** The conditional probabilities are the same in any of the n generations. That is, α_n, β_n, γ_n and δ_n are independent of n, so we may drop the suffix.

Together these assumptions enable us to simplify (8.8) to the quadratic function

$$P(A_n) = \alpha P^2(A_{n+1}) + \beta P^2(\neg A_{n+1}) + (\gamma + \delta)P(A_{n+1})P(\neg A_{n+1})\,. \quad (8.9)$$

As we will show in the next section, (8.9) leads to a surprising result, for it generates a structure similar to the Mandelbrot fractal.

8.4 The Mandelbrot Fractal

In 1977 Mandelbrot introduced his celebrated iteration:

$$q_{n+1} = c + q_n^2, \tag{8.10}$$

where c and q are complex numbers.[9] Starting with $q_0 = 0$, the iteration gives us successively

$$\begin{aligned} q_1 &= c \\ q_2 &= c + c^2 \\ q_3 &= c + (c + c^2)^2 \\ q_4 &= c + \left(c + (c + c^2)^2\right)^2, \end{aligned} \tag{8.11}$$

and so on. For many values of c, the iteration will diverge, allowing q_n to grow beyond any bound as n becomes larger and larger. For example, if $c = 1$ we obtain $q_1 = 1$, $q_2 = 2$, $q_3 = 5$ and $q_4 = 26$, and so on.

But if for instance $c = 0.1$, then q_n does not diverge, and in this case actually converges to the number $0.11271\ldots$. Taken together, all the values of c for which the iteration (8.10) does not diverge form the Mandelbrot set, which is reproduced in Figure 8.3.

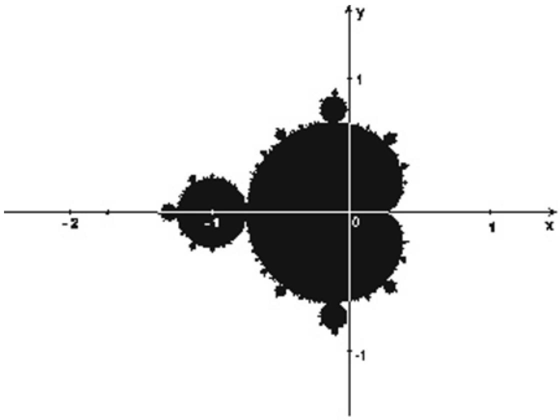

Fig. 8.3 The Mandelbrot fractal is generated by the complex quadratic iteration $q_n = c + q_{n+1}^2$, where $c = x + iy$.

[9] Mandelbrot 1977. The variables q_n in this section should not be confused with q in (8.7), the target proposition of the two-dimensional net.

The black area contains the points that belong to the Mandelbrot set. Each point corresponds to a complex number, c, being the ordered pair of the Cartesian coordinates, (x,y). The edge of the Mandelbrot set forms the boundary between those values of c that are members of the set and those that are not. It is this boundary, the 'Mandelbrot fractal', that has the well-known property of being infinitely structured in a remarkable way: no matter how far you zoom in on it, you will always find a new structure that is similar to, although not completely identical with the Mandelbrot set itself.

Our aim in this section is to demonstrate that, on condition that $\alpha+\beta \neq \gamma+\delta$, the quadratic relation (8.9) is equivalent to the Mandelbrot iteration (8.10). As it turns out, c will be a function of the conditional probabilities α, β, γ and δ alone, and will thus be a known quantity. The q's, on the other hand, will be directly related to the unconditional probabilities; these are unknown and their values are to be determined through the iteration.

It will prove convenient first to define ε as the average of the conditional probabilities γ and δ, that is

$$\varepsilon \stackrel{\text{def}}{=} \tfrac{1}{2}(\gamma+\delta)\,,$$

which is the mean conditional probability that the target — in our case Barbara — has the trait T, given that only one of her parents has T. Eq.(8.9) now becomes

$$P(A_n) = \beta + 2(\varepsilon-\beta)\,P(A_{n+1}) + (\alpha+\beta-2\varepsilon)\,P^2(A_{n+1})\,. \tag{8.12}$$

On the one hand, this iteration may not look very much like the Mandelbrot form (8.10). Firstly, in the latter we go as it were upwards, starting from q_n and then counting to q_{n+1}, whereas in (8.12) we start with $P(A_{n+1})$ and iterate downwards to $P(A_n)$. Secondly, (8.12) is about conditional and unconditional probabilities, and thus about real numbers between zero and one, whereas (8.10) is an uninterpreted formula involving complex numbers. On the other hand, however, we see that there is an important similarity between (8.10) and (8.12). Both are quadratic expressions: the former contains q_n^2 and the latter $P^2(A_{n+1})$. In order to transform (8.12) into (8.10) we introduce a linear mapping that serves to remove from (8.12) the term $2(\varepsilon-\beta)\,P(A_{n+1})$, and also the coefficient $(\alpha+\beta-2\varepsilon)$. The appropriate linear mapping that does the trick, $P(A_n) \to q_n$, is defined by

$$q_n = (\alpha+\beta-2\varepsilon)\,P(A_n) - \beta + \varepsilon\,. \tag{8.13}$$

On substituting (8.12) for $P(A_n)$ in (8.13) we obtain a formula that can be rewritten as

$$q_n = \varepsilon(1-\varepsilon) - \beta(1-\alpha) + q_{n+1}^2 \,. \tag{8.14}$$

The details of this calculation can be found in Appendix D.2.

Now define

$$c = \varepsilon(1-\varepsilon) - \beta(1-\alpha) \,. \tag{8.15}$$

Note that c involves only the conditional probabilities, α, β and ε, and so is an invariant quantity during the execution of the iteration. On the other hand, q_n also contains the unconditional probability, $P(A_n)$, which we seek to evaluate through the iteration. With the definition (8.15), Eq.(8.14) becomes

$$q_n = c + q_{n+1}^2 \,. \tag{8.16}$$

Evidently (8.16) is very similar to the standard Mandelbrot iteration (8.10). There is only the one difference which we have already mentioned: instead of an iteration upwards from $n = 0$, the iteration in (8.16) proceeds from a large n value, corresponding to the primeval parents, down to the target child proposition at $n = 0$. This difference is however only cosmetic and has no significance for the iteration as such.

We are now in a position to take advantage of some of the lore that has accumulated about the Mandelbrot iteration. *Some* but not all, for there is still the second difference that we mentioned: epistemic justification as we discuss it here deals with probabilities, and those are real numbers, rather than complex ones. Hence we must concentrate on the real subset of the complex numbers c in (8.15), namely those for which $c = (x,0)$, corresponding to the x-axis in Figure 8.3. It should be noted that, when c is real, all the q_n are automatically real — compare the explicit expressions for the first few n-values, just after (8.11). It is known that the real interval $-2 \leq c \leq \frac{1}{4}$ lies within the Mandelbrot set, but not all of these values correspond to an iteration that converges to a unique limiting value.

However, let us now impose the condition of probabilistic support, with exclusion of zero and one. Although $0 < \beta < \alpha < 1$ has the same form as the condition of probabilistic support for the one-dimensional chain, it should be realized that α and β do not have quite the same meanings in the two contexts. In the one-dimensional chain, $\alpha > \beta$ means that the probability of the child's having trait T is greater if the mother has it than if the mother does not have it. In the two-dimensional net, however, $\alpha > \beta$ means that the probability of the child's having trait T is greater if *both* of her parents have it than if *neither* of them do.

The essential point is that with $0 < \beta < \alpha < 1$ we can show from (8.15) that $-\frac{1}{4} < c < \frac{1}{4}$ (see again Appendix D.2). In this domain the Mandelbrot iteration is known to converge to a unique limit. Were it not for probabilistic support, convergence would not be guaranteed, indeed a so-called two-cycle, in which q_n flips incessantly between two values, would have been a possibility. Hence the condition of probabilistic support is necessary for convergence in this case.

A fixed point of the mapping (8.16) is a number, q_*, that satisfies

$$q_* = c + q_*^2 . \tag{8.17}$$

In Appendix D it is proved that the solution

$$q_* = \frac{c}{\frac{1}{2} + \sqrt{\frac{1}{4} - c}} , \tag{8.18}$$

is the so-called attracting fixed point of (8.16), meaning that the iteration (8.16) converges to q_*. Independently of the value one takes as the starting point for the iteration (i.e. q_N for some large N), attraction to the same q_* takes place (on condition that the starting point is not too far from q_* — technically, the condition is that it is within the basin of attraction of the fixed point). Under these conditions the starting point or ground has no effect on the final value of the target, q_0. The phenomenon is precisely that of fading foundations, now in the context of a two-dimensional net.

This fixed point (8.18) corresponds to the following fixed point of (8.12):

$$p_* = \frac{\beta}{\beta + \frac{1}{2} - \varepsilon + \sqrt{\beta(1-\alpha) + (\varepsilon - \frac{1}{2})^2}} . \tag{8.19}$$

Note that, if $\varepsilon = \frac{1}{2}(\alpha+\beta)$, which is equivalent to $\alpha+\beta = \gamma+\delta$, p_* reduces to $\beta/(1-\alpha+\beta)$, and this agrees with the sum of the one-dimensional iteration (3.17).

If β tends to zero the solution (8.19) is interesting, for it vanishes only if $\varepsilon \leq \frac{1}{2}$. If $\varepsilon > \frac{1}{2}$ it tends to the nontrivial value $(2\varepsilon - 1)/(2\varepsilon - \alpha)$ — see Appendix D.2. This behaviour is different from that of the one-dimensional case, in which the solution always vanishes when β tends to zero.

The two-dimensional network is generated by the same recursion that produces the Mandelbrot set in the complex plane. True, we have only to do with the real line between $-\frac{1}{4}$ and $\frac{1}{4}$, and not with the complex plane (where the remarkable fractal structure is apparent). But the point is that the algorithm which produces our sequence of probabilities, and that which generates the Mandelbot fractal, are the same.

We have used three simplifying assumptions in proving the above properties, viz. those of independence, probabilistic symmetry between A_{n+1} and

A'_{n+1}, and uniformity. There are however strong indications that essentially similar results also hold when these assumptions are dropped. Imagine a situation in which the probabilities are different for A_{n+1} and A'_{n+1}. Then there will be two coupled quadratic iterations, one for $P(A_n)$ and one for $P(A'_n)$. Each of these is related to $P(A_{n+1})$ as well as $P(A'_{n+1})$. This is however merely a technical complication, for it is still possible to find a domain in which the iterations converge. The relation is in fact a generalized Mandelbrot iteration, and analogous results obtain.

The same applies if we drop the assumption of independence. Clearly, if A_{n+1} and A'_{n+1} are stochastically dependent, we may have to include more distant links in the network, which of course complicates matters considerably. However, in general terms it means nothing more than that the final fixed-point equations will be of higher order. Again a generalized Mandelbrot-style iteration will hold sway, and again domains of convergence will exist.

Furthermore, in many situations the conditional probabilities may not be uniform: they may change from generation to generation. In those cases the iteration will become considerably more involved. We have seen that for the one-dimensional chain it proved possible to write down explicitly the result of concatenating an arbitrary number of steps. It is true that for a two-dimensional net this would be very cumbersome. However, with the use of a fixed-point theorem it is possible to give conditions under which convergence once more occurs.

What will happen when the network has more dimensions than two? In that case the fixed-point equations will be of even higher order, necessitating computer programs for their calculation. The picture itself however remains essentially the same. The probabilities are determined by polynomial recurrent expressions, and there will be a domain in which they are uniquely determined.

We conclude that probabilistic epistemic justification has a structure that gives rise to a generalized Mandelbrot recursion. This still holds when we abandon our three simplifying assumptions, or when we work in more than two dimensions. In short, not only do the algorithms describing ferns, snowflakes and many other patterns in nature generate a fractal, but the same is true for the description of our patterns of reasoning.

8.5 Mushrooming Out

Consider once more our justificatory chain in one dimension

$$q \longleftarrow A_1 \longleftarrow A_2 \longleftarrow A_3 \longleftarrow A_4 \ldots$$

where the arrow is again interpreted as probabilistic support. Above we have constructed multi-dimensional networks by letting new chains spring from the nodes, that is the unconditional probabilities. However chains can also arise from the connections, that is from the arrows. This possibility seems to be have been anticipated by Richard Fumerton.

Fumerton has observed that many examples of sceptical reasoning rely on a principle which he calls the Principle of Inferential Justification. The principle consists of two clauses:

> To be justified in believing one proposition q on the basis of another proposition A_1, one must be (1) justified in believing A_1 and (2) justified in believing that A_1 makes probable q.[10]

He then argues that, ironically, the same principle is used to *reject* scepticism and to support classic foundationalism:

> The foundationalist holds that every justified belief owes its justification ultimately to some belief that is noninferentially justified. ... The principle of inferential justification plays an integral role in the famous regress argument for foundationalism. If all justification were inferential, the argument goes, we would have no justification for believing anything whatsoever. If all justification were inferential, then to be justified in believing some proposition q I would need to infer it from some other proposition A_1. According to the first clause of the principle of inferential justification, I would be justified in believing q on the basis of A_1 only if I were justified in believing A_1. But if all justification were inferential I would be justified in believing A_1 only if I believed it on the basis of something else A_2, which I justifiably believe on the basis of something else A_3, which I justifiably believe on the basis of something else A_4, ..., and so on ad infinitum. Finite minds cannot complete an infinitely long chain of reasoning, so if all justification were inferential we would have no justification for believing anything.[11]

---

[10] Fumerton 1995, 36; 2001, 6. We have substituted q and A_1 for Fumerton's P and E. Fumerton applies the principle in particular to scepticism of what he calls the "strong" and "local" kind (Fumerton 1995, 29-31). Strong scepticism denies that we can have justified or rational belief; it is opposed to weak scepticism, which denies that we can have knowledge. Local scepticism is scepticism with respect to a given class of propositions, whereas global scepticism denies that we can know or rationally believe *all* truth.

[11] Fumerton 1995, 56-57.

We recognize here the finite mind objection to infinite justificatory chains, which we discussed in Chapter 5. This objection, that serves as an argument in support of foundationalism, alludes to the first clause of the Principle of Inferential Justification, and it consitutes the first part of Fumerton's epistemic regress argument for foundationalism.[12] There is however a second part to Fumerton's epistemic regress argument. This part depends on the second clause of the Principle of Inferential Justification, and it has to do with multi-dimensionality arising from chains that spring from connections rather than from nodes. Here again an infinite number of infinite regresses mushroom out in infinitely many directions:

> To be justified in believing q on the basis of A_1, we must be justified in believing A_1. But we must also be justified in believing that A_1 makes probable q. And if all justification is inferential, then we must justifiably infer that A_1 makes probable q from some proposition B_1, which we justifiably infer from some proposition B_2, and so on. We must also justifiably believe that B_1 makes probable that A_1 makes probable q, so we would have to infer that from some proposition C_1, which we justifiably infer from some proposition C_2, and so on. And we would have to infer that C_1 makes probable that B_1 makes probable that A_1 makes probable q ... The infinite regresses are mushrooming out in an infinite number of different directions.[13]

The consequences of this particular mushrooming out seem to be bleak indeed, as Fumerton notes:

> If finite minds should worry about the possibility of completing one infinitely long chain of reasoning, they should be downright depressed about the possibility of completing an infinite number of infinitely long chains of reasoning.[14]

Fortunately, however, things are not as grim as Fumerton suggests. The situation is on the contrary very interesting. For Fumertonian mushrooming out generates a Mandelbrot-like iteration of the sort that we described in the previous section.

Let us explain. In the previous chapters we have thought of the conditional probabilities as somehow being given: they were measured or estimated, for

[12] For Fumerton's distinction between the *epistemic* and the *conceptual* regress argument for foundationalism, see Section 6.1. There we argued that the conceptual regress argument amounts to the no starting point objection to infinite epistemic chains.

[13] Fumerton 1995, 57. B_1, C_1 etc. come in the place of Fumerton's F_1, G_1.

[14] Ibid.

instance in a laboratory, as in our example about the bacteria. With given conditional probabilities, there is of course no Fumertonian mushrooming out: we can iterate the unconditional probabilities in the usual way on the basis of the conditional probabilities as our pragmatic starting point. However, Fumerton is right to intimate that sometimes the conditional probabilities are unknown or at least uncertain; then their values have to be justified by some further proposition, which has to be justified by yet another proposition, and so on, and we are faced with mushrooming in Fumerton's sense. How to deal with this situation?

Again let q be probabilistically supported by A_1:

$$P(q|A_1) > P(q|\neg A_1)\,.$$

Now suppose that these two conditional probabilities are not given. The only thing we know is that "q is probabilistically supported by A_1" is in turn made probable by another proposition, for example by B_1. The way to express this is by writing down the relevant rules of total probability, this time for conditional rather than unconditional probabilities:

$$\begin{aligned} P(q|A_1) &= P(q|A_1 \wedge B_1)P(B_1|A_1) + P(q|A_1 \wedge \neg B_1)P(\neg B_1|A_1) \qquad (8.20)\\ P(q|\neg A_1) &= P(q|\neg A_1 \wedge B_1)P(B_1|\neg A_1) + P(q|\neg A_1 \wedge \neg B_1)P(\neg B_1|\neg A_1)\,. \end{aligned}$$

These rules are clearly more complicated than the simple rule for an unconditional probability, although we already encountered this complicated form in (7.26) of Chapter 7, when we discussed our model for higher-order probabilities.[15]

The unconditional probability $P(q)$ can be written as

$$P(q) = P(q|A_1)P(A_1) + P(q|\neg A_1)P(\neg A_1)\,,$$

and on using (8.20) to evaluate the two conditional probabilities, we find that

$$\begin{aligned} P(q) = &\ \big[P(q|A_1 \wedge B_1)P(B_1|A_1) + P(q|A_1 \wedge \neg B_1)P(\neg B_1|A_1)\big]P(A_1)\\ &+ \big[P(q|\neg A_1 \wedge B_1)P(B_1|\neg A_1) + P(q|\neg A_1 \wedge \neg B_1)P(\neg B_1|\neg A_1)\big]P(\neg A_1)\\ = &\ \alpha_0 P(A_1 \wedge B_1) + \gamma_0 P(A_1 \wedge \neg B_1) + \delta_0 P(\neg A_1 \wedge B_1) + \beta_0 P(\neg A_1 \wedge \neg B_1)\,. \end{aligned}$$

The last line has precisely the structure of (8.7), reading B_1 here for A'_1 there. This shows that a single mushrooming out *à la* Fumerton is isomorphic to the two-dimensional equations of the previous section.

---

[15] An intuitive way of seeing that (8.20) is correct is to realize that, in the reduced probability space in which A_1 is the whole space, all the occurrences of A_1 can be omitted. Then (8.20) reduces to the rule of total probability for an unconditional probability.

We have seen that, where chains spring from the nodes, the two-dimensional equations could be extended to equations in many, and even infinitely many dimensions, yielding a Mandelbrot structure. The same reasoning can be applied here, where chains spring from the connections. If many, or even a denumerable infinity of conditional probabilities are in turn probabilistically supported, then one has to do with the many-dimensional generalization.

Of course we will never deal with all these dimensions in reality. Our result is first and foremost a formal one. Having said this we should not underestimate the relevance of formal results for real life justification. Although it is true that in justifying our beliefs we can handle only short, finite chains, it is thanks to formal reasoning that we can recognize in these chains the manifestation of fading foundations: solely through formal proofs do we know that what we see in real life justification is not a fluctuation or a coincidence.[16]

8.6 Causal Graphs

In the first chapter we briefly referred to the similarities between epistemic and causal chains. Especially at a formal level, as we stressed in Chapter 2, a chain of reasons and a chain of causes are very much alike. Thus the linear chain

$$A_0 \longleftarrow A_1 \longleftarrow A_2 \longleftarrow A_3 \longleftarrow A_4 \longleftarrow \dots \tag{8.21}$$

can be interpreted as a one-dimensional causal series, where A_0 is the fact or event (rather than the proposition) that bacterium Barbara from Chapter 3 has trait T, and A_1 is the fact or event that her mother had T, and so on, backwards in time. The arrows in (8.21) stand for probabilistically causal influences: if a mother has T, it is more likely, but not certain, that her daughter will have T. This is in line with ordinary usage, for example when one says that smoking causes lung cancer, even though one knows that not all smokers contract the affliction, and that some non-smokers succumb to it. To avoid cumbersome language, we shall sometimes say that A_0 stands for Barbara

[16] As the size and complexity of the multi-dimensional networks increase, it will become more and more difficult to have them correspond to empirically based conditional probabilities. A rather wild speculation is that in the end such a world-network might have only one solution. See Atkinson and Peijnenburg 2010c, where we mull over the implications of such a speculation, taking as our starting point Susan Haack's crossword metaphor for 'foundherentism' (Haack 1993, Chapter 4).

(rather than for the fact that Barbara has T), that A_1 stands for her mother (rather than for the fact that her mother has T), and so on.

In the language of Directed Acyclic Graphs (DAGs) one would say that (8.21) is a DAG just in case the Markov condition holds.[17] This means in particular that A_1 screens off A_0 from A_2 in the sense of Reichenbach, that A_2 screens off A_1 from A_3, and so on.[18] However, the Markov condition is much stronger than a screening-off constraint that involves only three successive events. The idea is that the 'parent event' of a 'child event' screens off the child from any and all 'ancestor events', or combinations thereof. For the chain of (8.21), the condition is formally as follows:

$$P(A_n|A_{n+1} \wedge Z) = P(A_n|A_{n+1})$$
$$P(A_n|\neg A_{n+1} \wedge Z) = P(A_n|\neg A_{n+1}),$$

for all $n \geq 0$. Here Z stands for any event, A_m, in the chain, apart from the descendents of A_n, i.e. for any $m \geq n+2$, or for any conjunction of such events, or their negations. This can be written succinctly as

$$P(A_n|\pm A_{n+1} \wedge Z) = P(A_n|\pm A_{n+1}),$$

where it is understood that $+A_{n+1}$ simply means A_{n+1}, and $-A_{n+1}$ means $\neg A_{n+1}$. The idea, informally, is that the Markov condition ensures that the causal influences which probabilistically circumscribe Barbara's genetic condition are determined by her mother alone, and that one can forget about all her ancestors except for her mother.

It should be stressed that our analysis of the probabilistic regress in no way requires the imposition of the Markov condition: fading foundations and the emergence of justification in the case of a justificatory regress work just as well with, as without the Markov condition. The causal influence of the primal ancestor fades away as the distance between Barbara and the ancestor increases, and Barbara's probabilistic tendency to have T emerges from the causal regress, whether or not the Markov condition holds.

It is certainly possible, in a particular causal chain, that fact A_2 could have a causal influence on A_0 directly, apart from its indirect influence through A_1. Hesslow has given an example.[19] Birth control pills, A_2, directly increase the probability of thrombosis, A_0, but indirectly reduce it in sexually active women by reducing the probability of pregnancy, A_1, which itself constitutes

[17] Spirtes, Glymour and Scheines 1993; Pearl 2000; Hitchcock 2012.
[18] Reichenbach 1956.
[19] Hesslow 1976.

a thrombosis risk. Then the Markov condition, as we have stated it for (8.21), would break down, and one would have to add a direct causal link between A_0 and A_2, as shown in Figure 8.4. In this case a modified Markov condition could still be in force: now both A_1 and A_2 count as parent events of A_0, and they together might screen off A_0 from the rest of the chain (depending on the details of the case, of course).

$$\begin{tikzcd} A_0 & A_1 \arrow[l] & A_2 \arrow[l] \arrow[ll, bend right] & A_3 \arrow[l] & A_4 \arrow[l] & \dots \arrow[l] \end{tikzcd}$$

Fig. 8.4 Modified causal chain

An advantage of the above considerations concerning the Markov condition is that they facilitate a demonstration of the consistency of our probabilistic regress.[20] This works just as well for the regress of justification as it does for the regress of causes. The idea is that, with the Markov condition in place, one can work out the probabilities of the conjunction of any of the A_n in terms of the usual conditional probabilities and the unconditional probabilities of the A_n, which, as we know, can be calculated from the conditional probabilities alone (on condition of course that the latter are in the usual class). For example, as shown in Appendix A.8,

$$P(A_1 \wedge \neg A_3 \wedge A_4) = (\beta_1 + \gamma_1\beta_2)(1 - \alpha_3)P(A_4)\,.$$

So there is a probability distribution over all the conjunctions of events (or propositions), and thus the probabilistic regress is consistent in this sense. If the Markov constraint is not imposed, on the other hand, so that the chain may not be a genuine DAG, then there are in general many ways to distribute probabilities over the various conjunctions; but we are sure that there is at least one way, thanks to Markov, that is consistent.

Let us now progress from one to two dimensions. Consider the tree 8.2 of Section 8.3, but now reinterpreted as a causal net:
Note that, while the direction of epistemic support in Figure 8.2 is from the bottom of the figure to the top, the direction of causal influence in Figure 8.5 is from top to bottom. Thus event q probabilistically causes events A_1 and A_1', and A_1 in turn causes A_2 and A_2', while A_1' causes A_2'' and A_2'''. For example, q could stand for Barbara's grandmother — more accurately, for the event that Barbara's grandmother had T. Through binary fission this grandmother

[20] Herzberg 2013.

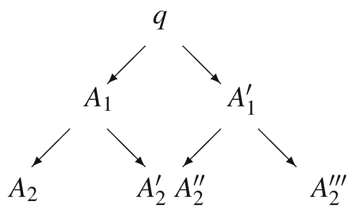

Fig. 8.5 Two-dimensional causal net with common causes

would split into two daughter cells, which would probably, but not certainly, have T. Then A_1 could stand for Barbara's mother, and finally A_2 for Barbara herself, A_2' for her sister bacterium. The eventualities A_1', A_2'' and A_2''' would have analogous meanings in respect of Barbara's aunt and her cousins.

One would expect the following Markov condition to hold, namely that A_1 screens off A_2 and A_2' from all the other events in the net. Thus

$$P(A_2|\pm A_1 \wedge Z) = P(A_2|\pm A_1)$$
$$P(A_2'|\pm A_1 \wedge Z) = P(A_2'|\pm A_1)\,,$$

where Z can be any of q, A_1', A_2'' or A_2''', or their negations, or any conjunctions of the same. Similarly, A_1' screens off A_2'' and A_2''' from q, A_1, A_2 and A_2'. One would also expect A_2 and A_2' to be positively correlated, so

$$P(A_2 \wedge A_2') > P(A_2)P(A_2')\,,$$

although they are conditionally independent in the sense that

$$P(A_2 \wedge A_2'|\pm A_1) = P(A_2|\pm A_1)P(A_2'|\pm A_1)\,.$$

This equation is in fact a consequence of the Markov condition. Following Reichenbach, we say that A_1 is the common cause of A_2 and A_2', and that event A_1 has brought it about that A_2 is more likely to occur if A_2' occurs, and *vice versa*.

A different kind of causal net is shown in Figure 8.6. Here the causal arrows go from bottom to top, which is the same as the direction of epistemic support in Figure 8.2. In Figure 8.6, A_2 could stand for a mother (i.e. for the event that a mother carries a particular trait, for example having blue eyes), A_2' could stand for her husband, and A_1 could stand for their daughter. Assuming that mother and father were not related, A_2 and A_2' are unconditionally independent,

$$P(A_2 \wedge A_2') = P(A_2)P(A_2')\,,$$

but they become correlated on conditionalization by A_1,

$$P(A_2 \wedge A_2' | \pm A_1) \neq P(A_2 | \pm A_1) P(A_2' | \pm A_1).$$

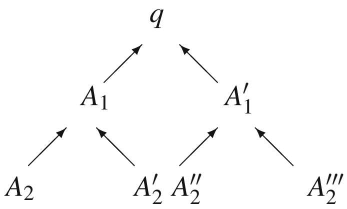

Fig. 8.6 Two-dimensional causal net with unshielded colliders

The subgraph involving A_2, A_2' and A_1 is a so-called unshielded collider. The behaviour of this collider, insofar as conditional and unconditional dependencies are concerned, is just the opposite of the behaviour of the common cause. Clearly Figure 8.6 is more like the two-dimensional justification tree of 8.2 than is the common cause graph of Figure 8.5. In the justification tree, proposition A_1 is probabilistically supported by A_2 and A_2': moeities of justification accrue to A_1 from A_2 and A_2', and from the conditional probabilities. In the causal collider, A_1 is probabilistically caused by A_2 and A_2'. Similarly, parents A_2'' and A_2''' cause A_1', the event that their son carries the trait in question. And finally A_1 and A_1' can cause the event that a child in the third generation has blue eyes.

Strictly speaking, Figure 8.6 is inaccurate, or at least ambiguous. The point is that A_1 would not be caused at all by A_2 in the absence of A_2'. We should replace Figure 8.6 by Figure 8.7, in which the joint nature of the causal influences is explicitly represented.

Mathematically, such a picture is called a directed hypergraph; and its properties have been studied by Selim Berker in the context of justificatory trees rather than causal trees.[21] Berker makes the point that such hypergraphs offer coherentists and infinitists a way of attaching a justification tree of beliefs or propositions to empirical facts. This is done without thereby making them foundational trees in which the facts constitute grounds in the sense of the foundationalist, that is as regress stoppers. For example, suppose now that A_2 in Figure 8.7 is an agent's experience that the sun is shining, and that

[21] Berker 2015.

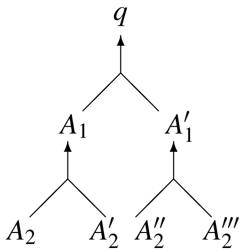

Fig. 8.7 Two-dimensional hypergraph

A'_2 is her belief that her eyes and visual cortex are functioning normally. Then A_1 could be the belief that the sun is indeed shining. The crux of the matter is that the fact A_2 does not by itself justify A_1, but does so only together with A'_2.

Berker claims that a coherentist (or infinitist) account of justification cannot consistently be based on probabilistic considerations. His reasoning is that the probabilistic coherence of a set of beliefs and experiences is the same as that of a similar set in which however all the experiences have been replaced by corresponding beliefs. He argues that the first set, the one including experiences, should be accorded a higher degree of justification than the second, which lacks experiences and is nothing but a collection of beliefs.

Berker's idea seems to hinge on a Humean view in which experiences outweigh beliefs. More importantly in the present context, it only bears on models in which probabilistic coherence is a sufficient determinant of justification. For models like ours, in which probabilistic coherence is only necessary, it is not apposite. And of course the phenomenon of fading foundations is not restricted to propositions or beliefs: it manifests itself also in the domain of experiences.

Just as the ground's share in the epistemic justification lessens, so the measure of the ground's causal influence vanishes in the end. In general, whether a regress is epistemic or causal, or whether it is in one or in many dimensions, justification and causation will progressively emerge and foundations will gradually fade away.