Chapter 7
Higher-Order Probabilities

Abstract
At first sight, a hierarchical regress formed by probability statements about probability statements appears to be different from the probabilistic regress of the previous chapters. After all, the former involves higher and higher-order probabilities, whereas the latter is an epistemic chain in which one proposition or belief probabilistically supports another. Closer examination, however, teaches us that the two regresses are in fact isomorphic. A model based on coin-making machines demonstrates that the hierarchical regress is consistent.

7.1 Two Probabilistic Regresses

We have extensively discussed chains of propositions which probabilistically support one another. But in Chapter 3 we did mention that Lewis, and independently Russell, seemed sometimes to be talking about higher-order probability statements rather than about straightforward chains of propositions.[1] The ambiguity is understandable enough. As we have seen, both Lewis and Russell took the view that probability statements like 'q is probable' or 'the probability of q is x' only make sense if one assumes that something else is

---

[1] Section 3.2, footnote 8, and Section 3.3, footnote 22. *Cf.* Reichenbach 1952, 151, where mention is also made of a probability of a probability. Roderick Chisholm has taken issue with Reichenbach's idea (especially as it is expressed in Reichenbach 1938), but in turn received criticism from Bruce Aune (Chisholm 1966, 22 *ff*; Aune 1972).

D. Atkinson, J. Peijnenburg, *Fading Foundations*, Synthese Library 383,
DOI 10.1007/978-3-319-58295-5_7

certain. The question then arises what exactly this 'something else' could be, and two answers appear to be natural.

According to the first, the 'something else' is the *reference class* on the basis of which the unconditional probability of q is determined. Lewis, we recall, argued that 'the probability of q is x' is in fact elliptical for 'the probability of q is x, on condition that A_1'. In many cases A_1 will be assumed to be certain, and thus to have probability unity. If this assumption is not made, then one has to assume that A_2 is certain in 'the probability of A_1 is x, on condition that A_2'. This reasoning forms the background to Lewis's conclusion that a regress of probability statements only makes sense if it is rooted in a certainty. According to the second answer, however, it is the *entire probability statement* that is taken to be certain. In asserting 'the probability of q is x', one usually presupposes that this assertion itself has probability unity. If one does not, then one might assume that the probability that the probability of this assertion is y (with y smaller than one) is certain. In other words, with the abbreviation of 'the probability of q is x' as A_1, one way in which A_1 could fail to be certain is if the assertion 'the probability of the probability that A_1 is y' (call this assertion A_2) is one.

These two answers lead to two different readings of a probabilistic regress. According to the first, the regress states (with v_n standing for the unconditional probability values):

the probability of q, on condition that A_1 is true, is v_0 ;
the probability of A_1, on condition that A_2 is true, is v_1 ;
the probability of A_2, on condition that A_3 is true, is v_2 ;
and so on.

According to the second reading, the regress amounts to:

A_1: the probability of q is v_0 ;
A_2: the probability of A_1 is v_1 ;
A_3: the probability of A_2 is v_2 ;
and so on.

In the first kind of regress every A_n represents a condition on the probability of q or on that of A_{n-1}. In the presence of such a regress, as we have seen, we generally are able to determine the unconditional probability of q via an infinite iteration of the rule of total probability. In fact, as we have explained, the iteration need not even be infinite in order for us to compute the unconditional probability of q to an acceptable approximation. However, in the second regress every A_n names a statement about a probability. It thus involves infinitely many statements about ever and ever higher-order probabilities, whereas the first regress refers to an infinite number of conditions.

Up to this point we have concentrated on the first kind of regress. In this chapter we shall focus on probabilistic regresses of the second kind, culminating in infinite series of probability statements about probability statements. We start in Section 7.2 by discussing probability statements of second and higher order. We will see that, although second-order probabilities do not pose any particular problem, many philosophers have objected to probability statements of a higher than second order. Especially the indefinite accumulation of probabilities to infinity has been generally regarded as not making sense.

In Section 7.3 we discuss an objection that Nicholas Rescher made to infinite-order probabilities. Our analysis of Rescher's argument will reveal that the above mentioned two readings of a probabilistic regress are in fact isomorphic, and in 7.4 this isomorphy will be demonstrated in a more formal way. Since regresses under the first reading are coherent, the isomorphy tells us that those under the second reading are too. Thus the properties of regresses under the first reading, such as those of fading foundations and emerging justification, are also properties of regresses under the second reading. In Section 7.5 we make the concept of infinite-order probability statements explicit by describing an executable model.

7.2 Second- and Higher-Order Probabilities

Suppose that the probability of the target proposition q is v_0:

$$P(q) = v_0 \,. \tag{7.1}$$

If we know that (7.1) is true, then there is no more to be said; but what if we lack this knowledge? In that case, we may only be in a position to assert a further probabilistic statement like

$$P(P(q) = v_0) = v_1 \,, \tag{7.2}$$

which is a second order probability statement, saying that the probability that (7.1) is true is v_1. Does (7.2) make sense? It can be argued that it does not. For if one supposes that (7.1) implies that $P(q) = v_0$ is true, then $P(P(q) = v_0) = 1$, and so, unless $v_1 = 1$, (7.2) would be inconsistent with (7.1). A way to avoid such an inconsistency would be to introduce two different probability functions instead of one, viz. $P^{(1)}$ and $P^{(2)}$. For evidently the intention is that (7.2) should *adjust* the initial bald statement (7.1). Thus we need to replace (7.1) and (7.2) by

$$P^{(1)}(q) = v_0$$
$$P^{(2)}(P^{(1)}(q) = v_0) = v_1\,, \tag{7.3}$$

where $P^{(2)}$ is a second-order probability function.

However, objections have been raised against second-order functions like $P^{(2)}$, based on the contention that it is unclear what they mean. David Miller even argued that they lead to an absurdity.[2] In his view the only way second-order probability statements could make sense, if at all, would be if the second-order probability of q, given that the first-probability of q is v_0, is itself v_0:

$$P^{(2)}\big(q|P^{(1)}(q) = v_0\big) = v_0\,. \tag{7.4}$$

He then goes on to argue that (7.4) leads to an unacceptable conclusion. For if we replace v_0 in (7.4) by $P^{(1)}(\neg q)$, we obtain

$$P^{(2)}\big(q|P^{(1)}(q) = P^{(1)}(\neg q)\big) = P^{(1)}(\neg q)\,,$$

which is the same thing as

$$P^{(2)}\big(q|P^{(1)}(q) = \tfrac{1}{2})\big) = P^{(1)}(\neg q)\,.$$

However, if instead we put $\frac{1}{2}$ for v_0 in (7.4), we find $P^{(2)}\big(q|P^{(1)}(q) = \frac{1}{2}\big) = \frac{1}{2}$. Therefore $P^{(1)}(\neg q) = \frac{1}{2}$, and thus $P^{(1)}(q) = \frac{1}{2}$. So if (7.4) were unrestrictedly valid, we could prove that the probability of an arbitrary proposition q is equal to one-half, which is absurd. This is known as the Miller paradox.

Brian Skyrms has argued against Miller's reasoning. Although Skyrms maintains that (7.4) is perfectly acceptable, playfully dubbing it 'Miller's Principle', he points out that Miller's further reasoning is fallacious, since it "rests on a simple *de re–de dicto* confusion".[3] As Skyrms explains, one and the same expression is used both referentially and attributively, so that a number (here v_0) is wrongly put on a par with a random variable, here $P^{(1)}(\neg q)$, that takes on a range of possible values.[4] So long as we recognize this confusion and keep the two levels apart, the notion of a second-order probability is harmless, and the Miller paradox disappears. We agree with

[2] Miller 1966.

[3] Skyrms 1980, 111.

[4] See Howson and Urbach 1993, 399-400, who make a similar observation.

Skyrms that Miller's Principle as such is harmless, but in what follows we will not need it: our reasoning goes through without the principle.

In addition to parrying Miller's argument, Skyrms warded off another objection to second-order probability statements, namely one that can be discerned in de Finetti's work. As is well known, de Finetti held that probability judgements are expressions of attitudes that lack truth values. Skyrms however pointed out that de Finetti's work is not particularly hostile to a theory of second-order probabilities:

> For a given person and time there must be, after all, a proposition to the effect that that person then has the degree of belief that he might evince by uttering a certain probability attribution.
>
> De Finetti grants as much:
>
> > The situation is different of course, if we are concerned not with the assertion itself but with whether 'someone holds or expresses such an opinion or acts according to it,' for this is a real event or proposition. (de Finetti 1972, 189)
>
> With this, de Finetti grants the existence of propositions on which a theory of higher-order personal probabilities can be built, but never follows up this possibility.[5]

De Finetti and Skyrms are not alone in having taken the view that second-order probabilities need not pose any particular problem. Several other authors recognize that, when the relevant distinctions are taken into account, second-order probabilities can be shown to be formally consistent.[6] This is not to say that such probabilities are mandatory. As Pearl has explained, second-order probabilities, although consistent, can be dispensed with, for one can always express them by using a richer first-order probability space.[7]

Once we accept the cogency of second-order probabilities, there is no impediment to constructing probabilities to any finite order.[8] We could continue the sequence (7.3) and introduce a hierarchy of higher-order probability statements:

$$\begin{aligned} P^{(1)}(q) &= v_0 \\ P^{(2)}(P^{(1)}(q) = v_0) &= v_1 \\ P^{(3)}(P^{(2)}(P^{(1)}(q) = v_0) = v_1) &= v_2\,, \end{aligned} \tag{7.5}$$

[5] Skyrms 1980, 113-114.

[6] Uchii 1973; Lewis 1980; Domotor 1981; Kyburg 1987; Gaifmann 1988.

[7] Pearl 2000.

[8] See Atkinson and Peijnenburg 2013.

and so on, with $P^{(m)}$ being the mth-order probability. In the previous section we introduced A_1, A_2 and A_3 as names of probability statements. Here we shall specify the probabilities in question more fully by stipulating their orders:

$$
\begin{aligned}
&A_1 \text{ is the proposition } P^{(1)}(q) = v_0\\
&A_2 \text{ is the proposition } P^{(2)}(A_1) = v_1\\
&A_3 \text{ is the proposition } P^{(3)}(A_2) = v_2\,,
\end{aligned}
$$

and so on. With these definitions, (7.5) can be written as

$$
\begin{aligned}
P^{(1)}(q) &= v_0\\
P^{(2)}(A_1) &= v_1\\
P^{(3)}(A_2) &= v_2\,,
\end{aligned}
\tag{7.6}
$$

and so on.

However, the key question is of course not whether any finite series of higher-order probability statements is cogent, but whether the notion of infinite-order probabilities make sense. Is it coherent to continue the above sequence *ad infinitum*, in the limit defining a probability, $P^{(\infty)}(q)$, of infinite order? Leonard Savage has answered this question in the negative. For him, the mere fact that second-order probabilities provoke the introduction of probability statements of infinite order was enough to discard them altogether:

> Once second order probabilities are introduced, the introduction of an endless hierarchy seems inescapable. Such a hierarchy seems very difficult to interpret, and it seems at best to make the theory less realistic, not more.[9]

His conclusion is that “insurmountable difficulties” will arise if one opens the door to second-order probabilities and starts using such phrases as “the probability that B is more probable than C is greater than the probability that F is more probable than G”.[10]

Savage was mainly talking about statistics, but in philosophy too it has been argued that an infinite order of probabilities of probabilities leads to problems that are insuperable. Thus David Hume argued in *A Treatise of Human Nature* that an infinite hierarchy implies that the probability of the target will always be zero:

[9] Savage 1954, 58.
[10] Ibid.

> Having thus found in every probability ... a new uncertainty ... and having adjusted these two together, we are oblig'd ... to add a new doubt .... This is a doubt ... of which ... we cannot avoid giving a decision. But this decision, ... being founded only on probability, must weaken still further our first evidence, and must itself be weaken'd by a fourth doubt of the same kind, and so on in infinitum: till at last there remain nothing of the original probability, however great we may suppose it to have been, and however small the diminution by every new uncertainty.[11]

Nicholas Rescher, in his book *Infinite Regress: The Theory and History of Varieties of Change*, also argued against an infinite hierarchy of probabilities. As he sees it, the problem with such a hierarchy is not that the probability of the target q will always be zero, but rather that it becomes impossible to know what that probability is:

> ... unless some claims are going to be categorically validated and not just adjudged probabilistically, the radically probabilistic epistemology envisioned here is going to be beyond the prospect of implementation. ... If you can indeed be certain of nothing, then how can you be sure of your probability assessments. If all you ever have is a nonterminatingly regressive claim of the format ... the probability is .9 that (the probability is .9 that (the probability of q is .9)) then in the face of such a regress, you would know effectively nothing about the condition of q. After all, without a categorically established factual basis of some sort, there is no way of assessing probabilities. But if these requisites themselves are never categorical but only probabilistic, then we are propelled into a vitiating regress of presuppositions.[12]

---

[11] Hume 1738/1961, Book I, Part IV, Section I. See also Lehrer 1981, for similar reasoning. As we noted in Section 3.3, Quine states in his lectures on Hume that this Humean argument is incorrect, since an infinite product of factors, all less than one, can be convergent, yielding a non-zero probability for the target (Quine 2008). Quine is right to point out this possibility, but note that it corresponds to what happens in our exceptional class, not in the usual class. Moreover, the possibility can only serve as a critique of Hume if one forgets about the second term in the rule of total probability, i.e. if all the $\beta_n = P(A_n|\neg A_{n+1})$ are zero. As we have seen, Hume does indeed leave out that term, as would Lewis and Russell many years later. Hume's argument is therefore not generally valid. Thus Quine's analysis of Hume is based on two unwarranted assumptions: first he assumes that all the conditional probabilities α_n belong to the exceptional class, i.e. the class of quasi-bi-implication, and second he supposes that all the conditional probabilities β_n are zero. What diminishes as the chain lengthens is not the probability of the target, as Hume and Quine thought, but rather the incremental changes that distant links bring about.

[12] Rescher 2010, 36-37. Rescher has p rather than q. Furthermore, Rescher explicitly conditions all his probabilities with respect to some evidence, E, and therefore

The argument of Rescher may seem plausible and persuasive. Yet we shall argue in the next section that an endless hierarchy of probabilities is in fact no stumbling block to having effective knowledge about the probability that q is true, let alone that it constitutes "an unsurmountable difficulty", as Savage would have it. In a sense the opposite is the case. If there is a stumbling block, it resides in the finite, not in the infinite hierarchy. For an infinite hierarchy of probabilities is, to a certain extent, better equipped to reveal the probability of q than is a finite one. The reason is reminiscent of the reason why a probabilistic regress of the sort that we have investigated in the previous chapters is cogent: in order to compute an infinite sequence, only the conditional probabilities need be known, whereas the computation of a finite sequence requires also knowledge of an unconditional probability.

7.3 Rescher's Argument

In this section we shall examine Rescher's claim: "If all you ever have is a nonterminatingly regressive claim of the format ... the probability is .9 that (the probability is .9 that (the probability of q is .9)) then in the face of such a regress, you would know effectively nothing about the condition of q", which amounts to putting v_0, v_1 and v_2 in (7.6) all equal to 0.9. We will show in this section that Rescher's assertion is in fact ill-founded.

Imagine, following Rescher, that we have a probability statement of the third order:

$$P^{(3)}(P^{(2)}(P^{(1)}(q) = 0.9) = 0.9) = 0.9\,. \tag{7.7}$$

Some philosophers conclude on the basis of (7.7) that the unconditional probability of q is 0.9, since no matter how many times one iterates, the probability value always stays the same.[13] This conclusion is also incorrect, but the question remains as to what *is* the correct conclusion that can be drawn from (7.7) about the unconditional probability of q.

Consider the definitions

A_1: The first-order probability $P^{(1)}$ of q is 0.9 ,
A_2: The second-order probability $P^{(2)}$ of A_1 is 0.9 ,
A_3: The third-order probability $P^{(3)}$ of A_2 is 0.9 ,

instead of Eqs.(7.1)–(7.2) he has $Pr(p|E) = v_0$ and $Pr(Pr(p|E) = v_0|E) = v_1$ (ibid., 36 — misprint corrected). In the interest of notational brevity, explicit reference to E will be suppressed.

[13] See for example DeWitt 1985, 128.

and so on. In the rest of this section we will temporarily suppress the orders of the probabilities to facilitate an intuitive grasp of the course of the reasoning. So we have:

A_1: The probability of q is 0.9,
A_2: The probability of A_1 is 0.9,
A_3: The probability of A_2 is 0.9,

and so on. We will now successively revise $P(q)$; in the next section, when we approach the matter more formally, we will reinstate the higher orders.

We call on the rule of total probability,

$$P(q) = P(q|A_1)P(A_1) + P(q|\neg A_1)P(\neg A_1)\,, \tag{7.8}$$

in which the probability of q is conditioned on that of A_1. In order to evaluate the unconditional probability of A_1, this formula must be repeated in the familiar way, with A_1 in the place of q, and A_2 in the place of A_1,

$$P(A_1) = P(A_1|A_2)P(A_2) + P(A_1|\neg A_2)P(\neg A_2)\,, \tag{7.9}$$

and so on. Is it possible to calculate $P(q)$ if the format goes on to infinity? Rescher thinks not. If the hierarchy is endless one cannot know anything about the probability of q, for "we are propelled into a vitiating regress of presuppositions"[14]. The situation looks like a probabilistic analogue of the Tortoise's interminable query to Achilles, where the latter successively satisfies the former *pro tem* in higher and higher-order querulousness without end.[15]

However, this similarity is only apparent. Between the probabilistic and the nonprobabilistic version of the Tortoise's challenge to Achilles there is an essential difference: the latter might be hopeless, the former is not. It is true that the Tortoise can always ask about an unknown $P(A_n)$ after the weary warrior has taken n steps in his argument. It is also true that the unknown $P(A_n)$ could have any value between zero and one. However, the influence that $P(A_n)$ has on the value of $P(q)$ will be smaller as the distance between A_n and q gets bigger — even if $P(A_n)$ were to take on the largest allowed value of 1, see Section 4.3. As we know now, in the limit that n tends to infinity, the influence of $P(A_n)$ on $P(q)$ will peter out completely, leaving the value of $P(q)$ as a function of the conditional probabilities alone. Note again that this is not because $P(A_n)$ itself becomes smaller as n becomes larger: indeed,

[14] Rescher 2010, 37.
[15] Carroll 1895.

it may not do so. Nor is it simply because the iteration of (7.8) and (7.9), etc. leads to a series of terms that is convergent. Rather it is because $P(A_n)$ is multiplied by a factor that goes to zero as n tends to infinity. Each time Achilles has taken one more step, and the Tortoise has asked about $P(A_{n+1})$, this worrisome probability is multiplied by an even smaller factor, and after yet another step the Tortoise's $P(A_{n+2})$ is multiplied by a yet smaller factor still, and so on, until the factor has shrunk to zero.

Referring back to (7.8), we know from Miller's Principle that the term $P(q|A_1)$ is equal to 0.9. In Rescher's example, the third term, $P(q|\neg A_1)$, is not specified, but it will be clear that (7.8) cannot be evaluated without it: as long as the value of the third term is unknown, one cannot determine $P(q)$. For the sake of argument, we shall set this term equal to 0.3. It should be noted that no strings are attached to this choice of 0.3, since the argument is robust: whatever nonzero value of $P(q|\neg A_1)$ is chosen, so long as it is less than $P(q|A_1)$, the same reasoning will work.

Now (7.8) can be worked out:

$$P(q) = [0.9 \times 0.9] + [0.3 \times 0.1] = 0.84\,. \tag{7.10}$$

The number 0.84 was arrived at on the provisional assumption that the second term, $P(A_1)$, indeed equals 0.9, which would be correct if it were the case that

$$P(P(A_1) = 0.9) = P(A_2) = 1\,.$$

But that is wrong, for $P(A_2) = 0.9$. This means that $P(A_1)$ should rather be

$$P(A_1) = [0.9 \times 0.9] + [0.3 \times 0.1] = 0.84\,, \tag{7.11}$$

where, similarly, 0.3 is taken to be the value of $P(A_1|\neg A_2)$, and so on. On the basis of this new result, the value of $P(q)$ in (7.10) must be revised, yielding

$$P(q) = [0.9 \times 0.84] + [0.3 \times 0.16] = 0.804\,. \tag{7.12}$$

However, the number 0.804 was arrived at on the fictional assumption that the second term in (7.11), to wit $P(A_2)$, indeed equals 0.9, and thus that

$$P(P(A_2) = 0.9) = P(A_3) = 1\,.$$

But that is also wrong, for $P(A_3) = 0.9$. This means that $P(A_2)$ should rather be

$$P(A_2) = [0.9 \times 0.9] + [0.3 \times 0.1] = 0.84\,. \tag{7.13}$$

On the basis of this, $P(A_1)$ is revised to

$$P(A_1) = [0.9 \times 0.9] + [0.3 \times 0.1] = 0.804\,. \tag{7.14}$$

This new value for $P(A_1)$ implies that $P(q)$ must again be revised, generating

$$P(q) = [0.9 \times 0.804] + [0.3 \times 0.196] = 0.7824\,, \tag{7.15}$$

and so on. It should be noted that these 'revisions' of the value of $P(q)$ are really higher and higher-order probabilities of q. We have suppressed the specification of the orders for greater readability: in the next section the technique will be explained with more care and with greater generality.

Here is an overview of the values that $P(q)$ takes after an increasing number of revisions:

Table 7.1 Unconditional probability of q after n revisions

n	1	2	3	5	10	15	20	∞
$P(q)$	0.84	0.804	0.7824	0.7617	0.7509	0.75007	0.750005	$\frac{3}{4}$

There are three important lessons to be drawn from these seemingly tedious calculations.

The first is that an endless hierarchy of probabilities can indeed determine what the probability of the original proposition is — contrary to what Rescher and many others have claimed. For it is possible to calculate the value of $P(q)$, even in a situation such as the one sketched by Rescher, where

$$P(P(P(q) = 0.9) = 0.9) = 0.9\,, \tag{7.16}$$

and so on. With the value that was chosen for $P(A_n|\neg A_{n+1})$, namely 0.3, and after an infinite number of revisions, $P(q)$ is exactly equal to $\frac{3}{4}$.

The second lesson is that an infinite number of revisions is not needed to come very close to the actual value of $P(q)$. For, as can be seen in Table 7.1, there is only a small difference between the value of $P(q)$ after, say, twenty revisions and after an infinite number of them. Of course, the size of the difference will depend on the numbers that are chosen for the conditional and unconditional probabilities in the equations: had the values of the first two terms been, for example, 0.8 rather than 0.9, and had $P(A_n|\neg A_{n+1})$ been 0.4 rather than 0.3, then not even twenty steps would have been needed to come as close to the limit value (which would have been $\frac{2}{3}$ in that case). There is always *some* finite number of revisions, such that the result scarcely differs from what is obtained with an infinite number of them.

The point can be regarded as a quantitative reinforcement of a claim that Rescher makes in qualitative terms. Partly in the wake of Kant and Peirce, Rescher stresses several times that some infinite regresses should be approached in a pragmatic way, in which it is acknowledged that contextual factors play an important role and that, at a certain point, “enough is enough”:

> ...in any given context of deliberation the regress of reasons ultimately runs out into ‘perfectly clear’ considerations which are (contextually) so plain that there just is no point in going further. It is not that the regress of validation ends, but rather that we stop tracking it because in the circumstances there is no worthwhile benefit to be gained by going on. We have rendered a state [or] situation by coming to the end not of what is possible but of what is sensible — not of what is feasible but of what is needed. Enough is enough.[16]

> ...in actual practice we need simply proceed ‘far enough’. After a certain point there is simply no need — or point — to going on.[17]

> Our explanations, interpretations, evidentiations, and substantiations can always be extended. But when we carry out these processes adequately, then after a while ‘enough is enough’. The process is ended not because it has to terminate as such, but simply because there is no point in going further. A point of sufficiency has been reached. The explanation is ‘sufficiently clear’, the interpretation is ‘adequately cogent’, the evidentiation is ‘sufficiently convincing’. ... [T]ermination is not a matter of necessity but of sufficiency — of sensible practice rather than of inexorable principle. ... What counts is doing enough ‘for practical purposes’.[18]

> ...regressive viciousness in explanation can be averted ... by the consideration that the practical needs of the situation rather than considerations of general principle serve to resolve our problems here. ... [I]n the end, what matters for rational substantiation is not theoretical completeness but pragmatic sufficiency.[19]

Rescher’s point is a good one, and it can be buttressed by the reasoning above — certainly in the case of an endless hierarchy of probabilities. Beside practical reasons for deciding that ‘enough is enough’, principled considerations can be used to determine when there is a negligible difference between the value of $P(q)$ after, say, fifteen steps, or after an infinite number of them.

[16] Rescher 2010, 47.
[17] Ibid., 82.
[18] Rescher 2005, 104.
[19] Ibid., 105.

Of course, it is on the basis of the context that the meaning of 'negligible' is to be understood. If one is happy to know what a particular probability is to within, say, one percent, then it is easy to work out, for given conditional probabilities, at what point the regress can be terminated, such that the error which is thereby committed is less than the desired one percent.

The third lesson, finally, must by now sound familiar: the further away A_n is from q, the smaller is the influence that the former exerts on the latter, until in the limit it dies out completely. In the end, the unconditional probabilities do not affect the value of $P(q)$ at all, only the conditional probabilities matter. Contrary to what Rescher suggests, the unconditional probability of q can be fully determined on the basis of the conditional probabilities, and of nothing else.

Again, this could be interpreted as a strengthening rather than a critique of Rescher's claims. At several places in his book Rescher explains that one of the ways in which an infinite regress can be harmless is when it is subject to "compressive convergence".[20] As he phrases it: "compressive convergence can enter in to save the day for infinite regression" (ibid.). In regresses governed by compressibility, "a law of diminishing returns" (ibid., 74) is in force, according to which the steps in the regress recede into "a minuteness of size" (ibid., 52):

> An infinite regress can thus become harmless when the regressive steps become vanishingly small in size so that the transit of regression becomes convergent. An ongoing approximation to a fixed result is then achieved, and the regress, while indeed proceeding *in infinitum*, does not reach *ad infinitum*.[21]

In the same vein, a law of diminishing returns can be said to be operating in the endless hierarchy of probabilities discussed above. Granted, it is not the case that in such a hierarchy the successive steps become smaller, let alone that they recede into "imperceptible minuteness".[22] Quite the contrary: in the limit that n goes to infinity, as has been shown, it is no impediment if $P(A_n)$ tends to the highest possible value, namely 1. Nor is it the case that, in the limit, $P(A_n)$ fades into penumbral obscurity in which its nature becomes unclear — another way in which, according to Rescher, an infinite regress can be harmless.[23] For the nature of the infinitely remote $P(A_n)$ may be perfectly clear and well-defined. Nevertheless a law of diminishing returns can still be said to be in force. Although the probability $P(A_n)$ does not shrink in

[20] Rescher 2010, 46.
[21] Ibid., 48.
[22] Ibid., 75.
[23] Ibid., 52.

size, nor becomes dim or otherwise unclear, the influence of $P(A_n)$ on $P(q)$, and thus the contribution that $P(A_n)$ makes to the value of $P(q)$, diminishes as the distance between A_n and q increases. This is because the hierarchical regress is isomorphic to the probabilistic regress, as we shall now prove.

7.4 The Two Regresses Are Isomorphic

In this section we will show that the regress of higher-order probabilities is strictly equivalent to the familiar probabilistic regress of propositions. Consider again (7.6). What is the second-order probability of q? It can be obtained from an instantiation of the rule of total probability at the second level:

$$\begin{aligned} P^{(2)}(q) &= P^{(2)}(q|A_1)P^{(2)}(A_1) + P^{(2)}(q|\neg A_1)P^{(2)}(\neg A_1) \\ &= \alpha_0^{(2)} v_1 + \beta_0^{(2)}(1 - v_1) \\ &= \beta_0^{(2)} + \gamma_0^{(2)} v_1 \,, \end{aligned} \tag{7.17}$$

where v_1 was defined in (7.6), and

$$\alpha_0^{(2)} = P^{(2)}(q|A_1)\,; \quad \beta_0^{(2)} = P^{(2)}(q|\neg A_1)\,; \quad \gamma_0^{(2)} = \alpha_0^{(2)} - \beta_0^{(2)}\,. \tag{7.18}$$

According to Miller's Principle in the form (7.4), $\alpha_0^{(2)}$ is equal to v_0; but since we do not need to call on this principle for our purposes, we will let $\alpha_0^{(2)}$ stand.

The third-order probability of q is given by

$$P^{(3)}(q) = P^{(3)}(q|A_1)P^{(3)}(A_1) + P^{(3)}(q|\neg A_1)P^{(3)}(\neg A_1)\,; \tag{7.19}$$

but the probability of A_1 at third order is no longer v_1, as it was at second order. Instead

$$\begin{aligned} P^{(3)}(A_1) &= P^{(3)}(A_1|A_2)P^{(3)}(A_2) + P^{(3)}(A_1|\neg A_2)P^{(3)}(\neg A_2) \\ &= \alpha_1^{(3)} v_2 + \beta_1^{(3)}(1 - v_2) \\ &= \beta_1^{(3)} + \gamma_1^{(3)} v_2 \,, \end{aligned} \tag{7.20}$$

where v_2 was defined in (7.6), and

$$\alpha_1^{(3)} = P^{(3)}(A_1|A_2)\,; \qquad \beta_1^{(3)} = P^{(3)}(A_1|\neg A_2)\,; \qquad \gamma_1^{(3)} = \alpha_1^{(3)} - \beta_1^{(3)}\,.$$

On substituting (7.20) into Eq.(7.19) we obtain

$$\begin{aligned} P^{(3)}(q) &= \alpha_0^{(3)}(\beta_1^{(3)} + \gamma_1^{(3)} v_2) + \beta_0^{(3)}(1 - \beta_1^{(3)} - \gamma_1^{(3)} v_2) \\ &= \beta_0^{(3)} + \gamma_0^{(3)}\beta_1^{(3)} + \gamma_0^{(3)}\gamma_1^{(3)} v_2 \,, \end{aligned}$$

with

$$\alpha_0^{(3)} = P^{(3)}(q|A_1)\,, \quad \beta_0^{(3)} = P^{(3)}(q|\neg A_1)\,, \quad \gamma_0^{(3)} = \alpha_0^{(3)} - \beta_0^{(3)}\,, \quad (7.21)$$

which is like Eq.(7.18), except that the conditional probabilities are now at third order.

The pattern should by now be obvious. The $(m+2)$nd-order probability of q is

$$P^{(m+2)}(q) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 + \ldots + \gamma_0\gamma_1\ldots\gamma_{m-1}\beta_m + \gamma_0\gamma_1\ldots\gamma_m v_{m+1}\,, \quad (7.22)$$

where we have suppressed the superscript $(m+2)$ on the conditional probabilities, for reasons of legibility, but they are to be understood.

Within the usual class we obtain, in the limit that m goes to infinity,

$$P^{(\infty)}(q) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 + \gamma_0\gamma_1\gamma_2\beta_3 + \ldots\,, \quad (7.23)$$

in which, with A_0 doing duty for the target proposition, q,

$$\alpha_n = P^{(\infty)}(A_n|A_{n+1})\,, \quad \beta_n = P^{(\infty)}(A_n|\neg A_{n+1})\,, \quad \gamma_n = \alpha_n - \beta_n\,, \quad (7.24)$$

for $n = 0, 1, 2, \ldots$.

It will be clear that the above argumentation on the basis of the rule of probability is formally the same as our reasoning in Chapter 3. Indeed, (7.23) has the same shape as (3.24), so an infinite series of higher-order probability statements makes sense. Like our regress of propositions that probabilistically justify one another, the regress of higher-order probabilities is subject to fading foundations and to justification that gradually emerges as we go to probability statements of higher and higher level.

Let us take stock. We have seen that higher-order probability statements are not as unintelligible as has often been thought. From Brian Skyrms and others we already learned that probabilities of the second order are not particularly problematic; but we have now seen that the same applies to probability statements of any finite order, and even that infinite-order probabilities turn out to be coherent. The two regresses, the one from the previous chapters and the hierarchical one, are formally equivalent.

However, formal equivalence is not yet equivalence in a very strict sense. We have shown that both regresses have the same form, not that there is a bijection between the two. The latter we will prove now, for the really conscientious reader.

We start straightforwardly, with the simplest form of Lewis's claim 'if something is probable, something else must be certain', i.e. the form where the series consists of only one step, namely from q to A_1. Here the two interpretations of Lewis's claim can be symbolized as follows:[24]

(1) If $P(q|A_1) = \alpha$, and $P(q|\neg A_1) < \alpha$, then A_1 is certain, i.e. $P(A_1) = 1$.
(2) It is certain that the probability of q is α, i.e. $P^{(2)}\big(P^{(1)}(q) = \alpha\big) = 1$.

It is not difficult to see that (1) entails (2). If $P(q|A_1) = \alpha$, and $P(A_1) = 1$, then

$$\begin{aligned} P(q) &= P(q|A_1)P(A_1) + P(q|\neg A_1)P(\neg A_1) \\ &= \alpha \times 1 + P(q|\neg p) \times 0 \\ &= \alpha\,; \end{aligned}$$

and if $P(q) = \alpha$, which we should write more explicitly as $P^{(1)}(q) = \alpha$, then the probability that this is so is one, i.e. $P^{(2)}\big(P^{(1)}(q) = \alpha\big) = 1$.

It is a little trickier to show that (2) entails (1). The first thing we have to do is to demonstrate that $P^{(2)}\big(P^{(1)}(q) = \alpha\big) = 1$ entails $P^{(1)}(q) = \alpha$. The difficulty is that $P(A) = 1$ does not imply A in an infinite probability space. On the other hand A does entail $P(A) = 1$, so if we substitute the proposition '$P^{(1)}(q) \neq \alpha$' for A, we obtain

$$P^{(1)}(q) \neq \alpha \quad \text{entails} \quad P^{(2)}\big(P^{(1)}(q) \neq \alpha\big) = 1\,.$$

By contraposition it follows that

$$\neg\big[P^{(2)}\big(P^{(1)}(q) \neq \alpha\big) = 1\big] \quad \text{entails} \quad \neg[P^{(1)}(q) \neq \alpha],$$

or in other words that

$$P^{(2)}\big(P^{(1)}(q) \neq \alpha\big) \neq 1 \;\; \text{entails} \;\; P^{(1)}(q) = \alpha\,. \tag{7.25}$$

However,

$$P^{(2)}\big(P^{(1)}(q) = \alpha\big) = 1 \quad \text{implies that} \quad P^{(2)}\big(P^{(1)}(q) \neq \alpha\big) = 0,$$

[24] Recall that Lewis does not mention $P(q|\neg A_1)$, but we specifically include the condition of probabilistic support.

and this trivially means that $P^{(2)}\big(P^{(1)}(q) \neq \alpha\big) \neq 1$. On combining this result with (7.25), we conclude that $P^{(2)}\big(P^{(1)}(q) = \alpha\big) = 1$ entails $P^{(1)}(q) = \alpha$.[25]

The rest of the demonstration employs only first-order probabilities, so we will drop the superscript. So with $P(q) = \alpha$, and the rule of total probability,

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

we see that, if A_1 is such that $P(q|A_1) = \alpha$ and $P(q|\neg A_1) < \alpha$, then

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

Therefore $P(A_1) = 1$, and so we have shown that (2) entails (1).

The above shows that the two interpretations of Lewis's claim are equivalent when the series consists only of q and A_1. However, the interesting question is whether the generalization still holds when the series is longer, and especially when it is of infinite length.

The generalization of (1) and (2) above to any finite series is given by:

$(1')$ If $P(A_n|A_{n+1}) = \alpha_n$ and $P(A_n|\neg A_{n+1}) = \beta_n$, with $\alpha_n > \beta_n$, for $n = 0, 1, \ldots m$, then it must be that A_{m+1} is certain, i.e. $P(A_{m+1}) = 1$.

$(2')$ It is certain that the mth-order probability of q is v_m, i.e. $P^{(m+1)}\big(P^{(m)}\big(\ldots\big(P^{(2)}\big(P^{(1)}(q) = v_0\big) = v_1\big)\ldots\big) = v_m\big) = 1$.

We have incorporated Reichenbach's correction of Lewis's position by including β_n, i.e. the second term in the rule of total probability. The condition of probabilistic support has also been included in order to exclude multiple solutions.

We will now show that $(1')$ and $(2')$ are equivalent. The right-hand side of (7.22) matches that of (3.20) in Chapter 3, excepting only that v_{m+1} in the former replaces $P(A_{m+1})$ in the latter. But v_{m+1} is just the value of $P^{(m+2)}(A_{m+1})$, so the two equations have the same form, term for term. Going from $(1')$ and $(2')$ is immediate, whereas in the opposite direction we must first demonstrate that $P^{(m+1)}(A_m) = 1$ entails A_m. But A_m is a probability statement, so the demonstration is just the same rigmarole as the one we detailed above in going from (2) to (1). Thus the finite chains are isomorphic; and therefore, if the conditional probabilities belong to the usual class, the infinite chains have the same form too. Infinite-order probabilities are not

[25] A shorter, intuitive 'proof' of this result is to say that $P(B) = 1$ entails B almost everywhere, and if B is a measure, namely the proposition $P^{(1)}(q) = \alpha$, then the restriction 'almost everywhere' loses its bite.

only cogent, but they also exhibit the phenomena that we have been talking about, in particular those of fading foundations and emerging justification.

As an example of an infinite-order probability, we take as conditional probabilities

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

These are the same as the ones we had in Eq.(3.21) of Chapter 3; but the interpretation is now different. Here they refer to infinite-order conditional probabilities. However, the equations have the same structure as those in Chapter 3; and we can read off the infinite-order probability of q by letting m go to infinity in Eq.(3.22), obtaining $P^{(\infty)}(q) = \frac{3}{4}$.

7.5 Making Coins

We have formally proved that an infinite series of higher-order probability statements is strictly equivalent to an infinite justificatory chain of the probabilistic kind. However, we might still have qualms: how can we understand the matter in an intuitive way? Being able to check all the steps in an algebraical proof is one thing, it is quite another thing to 'see through' the series, as it were, and to appreciate what is actually going on.

In this section we will try to allay these worries by offering a model that is intended to make the above abstract considerations concrete. The model is completely implementable; it comprises a procedure in which every step is specified. The model gives us a probability distribution over all the propositions as well as over their conjunctions. It satisfies the Markov condition in a very natural way, and we do not have to assume this condition as an external condition.[26]

Imagine two machines which produce trick coins. Machine V_0 produces coins each of which has bias α_0, by which we mean that each has probability α_0 of falling heads when tossed; machine W_0, on the other hand, makes coins each of which has bias β_0. We define the propositions q and A_1 as follows:

q is the proposition 'this coin will fall heads'

A_1 is the proposition 'this coin comes from machine V_0'.

[26] In Herzberg 2014 the Markov condition is imposed as an extra constraint. See also our discussion in Appendix A.8.

We shall use the symbol $P_1(q)$ for the probability of a head when A_1 is true; evidently it is the conditional probability of q, given A_1:

$$\begin{aligned}P_1(q) &\stackrel{\text{def}}{=} P(q|A_1)\\ &= P\big(\text{'this coin will land heads'}\,|\,\text{'this coin comes from machine } V_0\text{'}\big)\,.\end{aligned}$$

We know that $P_1(q) = \alpha_0$, for if the coin comes from machine V_0, the probability of a head is indeed α_0, for that is the bias produced by machine V_0. Note that P_1 is conceptually not the same as $P^{(1)}$. The former is a conditional probability, in this case the probability of q given A_1; the latter is a first-order unconditional probability.

An assistant is instructed to take many coins from *both* machines, and to mix them thoroughly in a large pile. The numbers of coins that she must add to the pile from machines V_0 and W_0 are determined by the properties of two new machines: V_1, which produces trick coins with bias α_1, and W_1, which produces trick coins with bias β_1. A supervisor has told the assistant that the relative number of coins that she should take from her machine V_0 should be equal to the probability, α_1, that a coin from V_1 would fall heads when tossed. So if α_1 is for example $\frac{1}{4}$, then one quarter of the total number of coins that the assistant takes from V_0 and W_0 are from V_0; the rest from W_0.[27]

The assistant takes one coin at random from her pile and she tosses it. Understanding q now to refer to this coin, we can deduce the probability of q in the new situation. Indeed, if A_2 is the proposition:

$$\begin{aligned}A_2 = \;&\text{'the relative number of } V_0 \text{ coins in the assistant's pile is determined}\\ &\text{by the bias towards heads of the } V_1 \text{ coins'},\end{aligned}$$

then we can ask what the probability is that the assistant's coin falls heads, given that A_2 is true. We use the symbol $P_2(q)$ for this probability. It is equal to the conditional probability of q, given A_2, which can be calculated from the following variation of the rule of total probability:[28]

[27] For the sake of this story, we limit α_1 to be a rational number, so it makes sense to say that the number of coins to be taken from V_0 is equal to α_1 times the total number taken from V_0 and W_0. Similarly, in the subsequent discussion, the biases should all be considered to be rational numbers. Since the rationals are dense in the reals, this is not an essential limitation.

[28] The proof of Eq.(7.26) goes as follows:

$$\begin{aligned}P(q\wedge A_2) &= P(q\wedge A_1\wedge A_2) + P(q\wedge\neg A_1\wedge A_2)\\ &= P(q|A_1\wedge A_2)P(A_1\wedge A_2) + P(q|\neg A_1\wedge A_2)P(\neg A_1\wedge A_2)\,.\end{aligned}$$

On dividing both sides of this equation by $P(A_2)$ we obtain (7.26).

$$P_2(q) \stackrel{\text{def}}{=} P(q|A_2) = P(q|A_1 \wedge A_2)P(A_1|A_2) + P(q|\neg A_1 \wedge A_2)P(\neg A_1|A_2)\,. \tag{7.26}$$

By definition, $P(q|A_1 \wedge A_2)$ is the probability that the assistant's coin will fall heads, on condition that this coin has come from machine V_0, *and* that the number of V_0 coins in the pile is subject to the condition specified by A_2. Similarly $P(q|\neg A_1 \wedge A_2)$ is the probability that the assistant's coin will fall heads, on condition that this same coin has *not* come from machine V_0, *and* that A_2 is true.

This series of procedures gives rise to a Markov chain. For the condition that the assistant's coin has come from machine V_0 is already enough to ensure that the probability that this coin will fall heads is α_0; and that situation is not affected by the condition that A_2 is true, so $P(q|A_1 \wedge A_2) = P(q|A_1) = \alpha_0$. Likewise, the condition that the assistant's coin has not come from machine V_0 guarantees that it has come from machine W_0, and therefore ensures that the probability of a head is β_0; again, that is not affected by A_2, so $P(q|\neg A_1 \wedge A_2) = P(q|\neg A_1) = \beta_0$. In Reichenbach's locution, A_1 is said to screen off q from A_2.[29] The screening-off or Markov condition will turn out to be an essential part of our model. We shall show that the model, as well as the abstract system of which it is an interpretation, are consistent, even if the abstract system does not itself satisfy the Markov condition.

The Markov condition enables us to simplify (7.26) as follows:

$$\begin{aligned} P_2(q) = P(q|A_2) &= P(q|A_1)P(A_1|A_2) + P(q|\neg A_1)P(\neg A_1|A_2) \\ &= \alpha_0\alpha_1 + \beta_0(1-\alpha_1)\,, \end{aligned} \tag{7.27}$$

where, as usual, we employ β_0 as shorthand for $P(q|\neg A_1)$. We conclude that, if the assistant repeats the procedure of tossing a coin from her pile many times (with replacement and randomization), the resulting relative frequency of heads would be approximately equal to $P_2(q)$, as given by (7.27). The approximation would get better and better as the number of tosses increases — more carefully: the probability that the relative number of heads will differ by less than any assigned $\varepsilon > 0$ from $\alpha_0\alpha_1 + \beta_0(1-\alpha_1)$ will tend to unity as the number of tosses tends to infinity.

It is important to understand that $P_2(q)$ is not simply a correction to $P_1(q)$. It is rather that they refer to two different operations. In the first operation it is certain that the assistant takes a coin from machine V_0. In the second operation something else is certain, namely that the number of V_0 coins in the pile consisting of V_0 coins and W_0 coins is determined by the bias towards heads of a coin from machine V_1. The consequence of this difference is substantial,

[29] Reichenbach 1956, 159-167.

for in the second operation it is no longer sure that the assistant takes a coin that comes from V_0. Instead of being only a correction, $P_2(q)$ is the result of a longer, and more sophisticated procedure than is $P_1(q)$.

So much for the description of the model of the first iteration of the regress, constrained by the veridicality of A_2. In the next iteration, the supervisor receives instructions from an artificial intelligence that simulates the working of yet another duo of machines, V_2 and W_2, which produce simulated coins with biases α_2 and β_2, respectively. The supervisor makes a large pile of coins from his machines V_1 and W_1; and he adjusts the relative number of coins that he takes from V_1 to be equal to the probability that a simulated coin from V_2 would fall heads when tossed. So if α_2 is for example $\frac{1}{2}$, then equal numbers of coins will be taken from each of the machines V_1 and W_1.

Let A_3 be the proposition:

$A_3 =$ 'the relative number of V_1 coins in the supervisor's pile is determined by the bias towards heads of the V_2 coins',

If A_3 is true, then the probability of A_2 is equal to α_2, that is to say $P(A_2|A_3) = \alpha_2$. Again, screening off is essential here: A_2 screens off A_1 from A_3. So we may write

$$\begin{aligned} P(A_1|A_3) &= P(A_1|A_2 \wedge A_3)P(A_2|A_3) + P(A_1|\neg A_2 \wedge A_3)P(\neg A_2|A_3) \\ &= P(A_1|A_2)P(A_2|A_3) + P(A_1|\neg A_2)P(\neg A_2|A_3) \\ &= \alpha_1\alpha_2 + \beta_1(1-\alpha_2)\,. \end{aligned} \tag{7.28}$$

This value of $P(A_1|A_3)$ is handed down to the assistant, and she reruns her procedure, but with $P(A_1|A_3)$ in place of $P(A_1|A_2)$. Since A_1 screens off q from A_3 (and from all the higher A_n), we calculate

$$\begin{aligned} P_3(q) \stackrel{\text{def}}{=} P(q|A_3) &= P(q|A_1 \wedge A_3)P(A_1|A_3) + P(q|\neg A_1 \wedge A_3)P(\neg A_1|A_3) \\ &= P(q|A_1)P(A_1|A_3) + P(q|\neg A_1)P(\neg A_1|A_3) \\ &= \alpha_0 P(A_1|A_3) + \beta_0[1 - P(A_1|A_3)]\,, \end{aligned} \tag{7.29}$$

in which we are to replace $P(A_1|A_3)$ by $\alpha_1\alpha_2 + \beta_1(1-\alpha_2)$, in accordance with Eq.(7.28). This yields

$$P_3(q) = P(q|A_3) = \beta_0 + (\alpha_0 - \beta_0)\beta_1 + (\alpha_0 - \beta_0)(\alpha_1 - \beta_1)\alpha_2\,. \tag{7.30}$$

The relative frequency of heads that the assistant would observe will be approximately equal to $P_3(q)$, as given by (7.30) — with the usual probabilistic proviso. The above constitutes a model of the second iteration of the regress, constrained by the condition that the simulated coin of the artificial intelligence comes from the simulated machine V_2, that is by the veridicality of A_3.

This procedure must be repeated *ad infinitum*. A subprogram encodes the working of yet another duo of virtual machines, V_3 and W_3, which simulate the production of coins with biases α_3 and β_3, and so on, all under the assumption that A_n is the proposition:

$A_n =$ 'the relative number of V_{n-2} coins in the relevant pile is determined by the bias towards heads of the V_{n-1} coins'.

From this it follows that at the $(m+2)$nd step of the iteration one finds

$$P_{m+2}(q) \stackrel{\text{def}}{=} P(q|A_{m+2}) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 \ldots + \gamma_0\gamma_1 \ldots \gamma_{m-1}\beta_m + \gamma_0\gamma_1 \ldots \gamma_m\alpha_{m+1}\,, \quad (7.31)$$

where we have introduced the customary abbreviation $\gamma_n = \alpha_n - \beta_n$. Under the requirement that the conditional probabilities belong to the usual class, the sequence $P_1(q)$, $P_2(q)$, $P_3(q)\ldots$ converges to a limit, $P_\infty(q)$, that is well-defined. Moreover, under the same condition the last term in (7.31), namely $\gamma_0\gamma_1 \ldots \gamma_m\alpha_{m+1}$, tends to zero as m tends to infinity, so finally

$$P_\infty(q) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 + \gamma_0\gamma_1\gamma_2\beta_3 \ldots \quad (7.32)$$

This has the same form as (7.23).

In this way we have designed a set of procedures that is clear-cut in the sense that it could in principle be performed to any finite number of steps, where the successive results for the probability that the assistant throws a head get closer and closer to a limiting value that can be calculated. To be precise, for any $\varepsilon > 0$, and for any set of conditional probabilities that belongs to the usual class, one can calculate an integer, N, such that $|P_N(q) - P_\infty(q)| < \varepsilon$, and one could actually carry out the procedures to determine $P_N(q)$. That is, one can get as close to the limit of the infinite regress of probabilities as one likes.

The probabilities in this model are objective, but that is not the essential point. What is essential is that the structure to be described is a genuine model, which implies that two desiderata have been met. First, the model is well-defined and free from contradictions. Second, it maps into the infinite hierarchy of probabilities. The model has the same form as the probabilistic regress of Chapter 3, for which we have already given a proof of convergence. It also matches the series for the infinite-order probability of Eq.(7.23), thereby providing a model for the abstract system of Section 7.4.