# AN ENDLESS HIERARCHY OF PROBABILITIES

Jeanne Peijnenburg and David Atkinson

(To be published in: *American Philosophical Quarterly.*)

ABSTRACT
According to radical probabilism, all factual claims are merely probabilistic in character. Throughout the centuries this view has been criticized on the grounds that it triggers an infinite regress: if every claim is probabilistic, then the proposition that something is probable is itself only probable, and so on. An endless hierarchy of probabilities seems to emerge and, as a result, the probability of the original proposition can never be determined. This criticism goes back as far as David Hume, and in the twentieth century it was raised also by statisticians such as Leonard J. Savage. Recently Nicholas Rescher ventured a similar critique of radical probabilism. In this paper it is argued that the criticism does not hold water, for an endless hierarchy of probability statements is no obstacle to attaching a definite probability value to the original proposition. Moreover, it is claimed that radical probabilism can reinforce some of Rescher's own main claims.

Keywords: higher order probability, infinite regress, Nicholas Rescher.

## 1 Introduction

Suppose $q$ is some proposition, and let

$$P(q) = v_0 \tag{1}$$

be the proposition that the probability of $q$ is $v_0$.[1] How can one know that (1) is true? One cannot know it for sure, for all that may be asserted is a further probabilistic statement like

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

which states that the probability that (1) is true is $v_1$. But the claim (2) is also subject to some further statement of an even higher probability:

$$P(P(P(q) = v_0) = v_1) = v_2\,, \tag{3}$$

and so on. Thus an infinite regress emerges of probabilities of probabilities, and the question arises as to whether this regress is vicious or harmless.

Radical probabilists would like to claim that it is harmless, but Nicholas Rescher, in his scholarly and very stimulating *Infinite Regress: The Theory and History of Varieties of Change*, argues that it is vicious (Rescher 2010). He believes that an infinite hierarchy of probabilities makes it impossible to know anything about the probability of the original proposition $q$:

> "... 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

[1] We follow Rescher in considering a point determination of a continuous probability; but, as Alan Hájek remarks, "The probability of a continuous random variable $X$, taking a particular value $x$, equals 0, being the integral from $x$ to $x$ of $X$'s density function" (Hájek 2003, p. 286). It would be apposite to replace the point determination by an interval, say $P(q) \in [v_0 - \varepsilon, v_0 + \varepsilon]$. The reader is invited to carry this refinement along in a charitable reading of our text, here and in the sequel.

> 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."[2]

Rescher is by no means the only philosopher who holds that an infinite hierarchy of probability statements does not make sense. David Hume, for example, is of the same opinion, albeit for somewhat different reasons. Rather than maintaining that an infinite hierarchy of probabilities does not say anything about the probability of the original proposition $q$, Hume argues that such a hierarchy implies that the probability in question is zero. In *A Treatise of Human Nature* Hume writes:

> "Having thus found in every probability ... a new uncertainty ... and having adjusted these two together, we are obliged ... 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 weakened 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."[3]

The positions of Rescher and Hume appear to be supported by Leonard J. Savage, who states that "insurmountable difficulties" will arise if one allows second order probabilities and if one 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 $G$" (Savage 1954, p. 58). One of these insurmountable difficulties is that an endless hierarchy of probabilities will occur:

> "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" (ibid.)

In this paper it is shown that the arguments of Rescher, Savage and Hume, plausible and persuasive as they may seem at first sight, are actually misleading. Not only can an endless hierarchy of probabilities be given a clear interpretation, it can also ascertain what the probability value for the original proposition $q$ is. Moreover, *pace* Savage, the hierarchy itself is not so unrealistic as it seems.

The paper is structured as follows. Section 2 contains an analysis of Rescher's example, in which "the probability is .9 that (the probability is .9 that (the probability of $q$ is .9))". Special attention is drawn to the fact that two conditional probabilities are needed in order to say something interesting about such examples. In Section 3 it is explained, contra Rescher, that an endless hierarchy of probabilities is no hindrance to determining the probability of the original proposition, $q$. In Section 4 this is illustrated on the basis of Rescher's own example. The conclusion is that, ultimately, the analysis does not undermine, but rather reinforces some of Rescher's own claims.

[2] Rescher 2010, pp. 36–37. Rescher has $p$ rather than $q$, and $v$ instead of $v_0$. Furthermore, Rescher explicitly conditions all his probabilities with respect to some evidence, $E$, so instead of Eqs.(1)–(2) he has $Pr(p|E) = v$ and $Pr(Pr(p|E) = v|E) = v_1$ (ibid., p. 36 — misprint corrected). In the interests of notational brevity, explicit reference to $E$ is suppressed, but it is to be understood.

[3] Hume 1738/1961 (Book I, Part IV, Section I), p. 178. See also Lehrer 1981.

## 2 Two conditional probabilities

Imagine, following Rescher, that "the probability is .9 that (the probability is .9 that (the probability of $q$ is .9))". Eq.(3) is now instantiated by

$$P(P(P(q) = 0.9) = 0.9) = 0.9\,.^{4} \tag{4}$$

Some philosophers conclude on the basis of (4) that the unconditional probability of $q$ is 0.9, since no matter how many times one iterates, the probability value always stays the same.[5] This conclusion is incorrect, but the question remains as to what *is* the correct conclusion that can be drawn from (4) about the unconditional probability of $q$.

In this connection it is important to realize that the unconditional probability of $q$ can only be calculated if the values of two conditional probabilities are known. The first is the conditional probability that $q$ is true if $P(q) = 0.9$ is true, in symbols

$$P(q|P(q) = 0.9).$$

The second is the conditional probability that $q$ is true if $P(q) = 0.9$ is false, in symbols:

$$P(q|P(q) \neq 0.9).$$

The unconditional probability of $q$ is given by the following instance of the rule of total probability:

$$P(q) = P(q|P(q) = 0.9)P(P(q) = 0.9) + P(q|P(q) \neq 0.9)P(P(q) \neq 0.9) \tag{5}$$

which is a consequence of the axioms of probability theory.[6] The right-hand side of this equation contains four terms: two conditional probabilities and two unconditional ones. The first term is the conditional probability $P(q|P(q) = 0.9)$. According to the rule that Brian Skyrms has dubbed 'Miller's Principle', this conditional probability is equal to 0.9 (Skyrms 1980, p. 112; see also Domotor 1981). In other words the probability of $q$, given that the probability of $q$ is 0.9, is 0.9.[7] The second term is the unconditional probability $P(P(q) = 0.9)$. In Rescher's example, this term is equal to 0.9. The third term is $P(q|P(q) \neq 0.9)$, the conditional probability of $q$, given that the unconditional probability of $q$ is not 0.9; and Rescher's reasoning leaves open what the value of this conditional probability might be. The fourth term is the unconditional probability $P(P(q) \neq 0.9)$, which simply is the complement of the second term. Since the unconditional probability that a proposition is true plus the unconditional probability that the same proposition is false, is one, the value of this term is $1 - 0.9 = 0.1$.

However, it is not yet clear that anything substantive has been attained. After all, one cannot be sure that the second term indeed equals 0.9, in other words that $P(P(q) = 0.9) = 0.9$ is true. For there is still Eq.(4) to reckon with:

$$P(P(P(q) = 0.9) = 0.9) = 0.9.$$

At this point things begin to look very complicated, and also rather forbidding. For not only is there a danger of getting bogged down in long and tedious formulas, it also appears forever

[4]As indicated in endnote 1, one might better replace the point determination $P(q) = 0.9$, by an interval, $P(q) \in [0.9 - \varepsilon, 0.9 + \varepsilon]$.

[5]See for example DeWitt 1985, especially pp. 128–129.

[6]If, along the lines of endnote 4, $P(q) = 0.9$ is to be replaced by $P(q) \in [0.9 - \varepsilon, 0.9 + \varepsilon]$, then $P(q) \neq 0.9$ must of course be replaced by $P(q) \notin [0.9 - \varepsilon, 0.9 + \varepsilon]$.

[7]Miller's Principle should not be confused with the Principal Principle of David Lewis. The two principles look alike, but the former is about a subjective probability of a subjective probability, whereas the latter concerns a subjective probability of an objective probability or chance. The fact that, according to Miller's Principle, the first term in Eq.(5) is equal to 0.9, might have caused the confusion in papers such as DeWitt's (see endnote 5).

impossible to say something meaningful about the unconditional probability of the original proposition. It looks as though Rescher is right that, in the face of a regress of probabilities of probabilities, "you would know effectively nothing about the condition of $q$". However, in the next sections it will be shown that this is not so and that things are not so depressing as they appear. An endless hierarchy of probabilities is no stumbling block to having effective knowledge about the probability that $q$ is true, let alone that it constitutes an "insurmountable difficulty". Quite the contrary. If there is a stumbling block, it resides in the finite, not in the infinite hierarchy. For as will be seen, an infinite hierarchy of probabilities is, in a sense, better equipped to reveal the probability of $q$ than is a finite one. The reason is that, 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.

## 3 Is a categorical factual basis needed?

To make progress possible, the notation first needs to be streamlined. Let

$$\begin{array}{lll} q_0 & \text{be the proposition} & q \\ q_1 & \text{be the proposition} & P(q_0) = v_0 \\ q_2 & \text{be the proposition} & P(q_1) = v_1 \\ q_3 & \text{be the proposition} & P(q_2) = v_2 \,, \end{array}$$

and so on. In other words, from now on $q_0$ will be written instead of $q$, and $q_{n+1}$ instead of the proposition $P(q_n) = v_n$, for all $n$. This means that Eqs. (1), (2) and (3) can now be abbreviated as $q_1$, $q_2$, and $q_3$.

The task is to determine $P(q_0)$, i.e., the unconditional probability of $q_0$. As has been explained in the previous section, this probability is given by:

$$P(q_0) = P(q_0|q_1)P(q_1) + P(q_0|\neg q_1)P(\neg q_1)\,, \tag{6}$$

in which the probability of $q_0$ is conditioned on that of $q_1$. However, to evaluate the unconditional probability of $q_1$, this formula must repeated, with $q_1$ in the place of $q_0$, and $q_2$ in the place of $q_1$, yielding

$$P(q_1) = P(q_1|q_2)P(q_2) + P(q_1|\neg q_2)P(\neg q_2)\,. \tag{7}$$

Similarly, to evaluate $P(q_2)$ the formula must be repeated again:

$$P(q_2) = P(q_2|q_3)P(q_3) + P(q_2|\neg q_3)P(\neg q_3)\,. \tag{8}$$

and so on. In this way one can successively eliminate the unknown quantities $P(q_1)$, $P(q_2)$, $P(q_3)$, and so on, in the calculation of $P(q_0)$.

Now a hierarchy of probabilities may be finite or it may be infinite. Consider first a finite hierarchy in which $q_3$ occupies the highest place. Then, in order to know the value of $P(q_0)$, not only all the conditional probabilities in (6), (7) and (8) need to be known, but also the unconditional probability that $q_3$ is true, $P(q_3)$. As Rescher states it: "... without a categorically established factual basis of some sort, there is no way of assessing probabilities" (Rescher 2010, p. 37). And he is right here: without a categorical factual basis in the form of $q_3$ or $P(q_3)$, there is no way of finding out what the probability of $q_0$ might be.[8]

[8]The phrase "a categorical factual basis in the form of $q_3$ or $P(q_3)$" is employed here in order to allow for a certain ambiguity in Rescher's use of 'categorical'. For in Rescher 2010 the term is used in two ways: firstly in opposition to 'probabilistic' (as in "But if these requisites themselves are never categorical but only probabilistic" on p. 37), secondly in opposition to 'conditional probability' (as in "the thesis that one can never determine categorical probabilities but only conditional ones" on p. 40). In the latter case, an unconditional probability is the same as a categorical probability. Under the first meaning, Rescher's "categorically established factual base of some sort" in the example would be $q_3$, under the second meaning it would be $P(q_3)$.

What happens in the other case, i.e., when the hierarchy is endless? Is it possible to calculate $P(q_0)$ if $n$ goes to infinity? According to Rescher, as has become clear, it is not possible. If the hierarchy is endless one cannot know anything about the probability of $q_0$, for "we are propelled into a vitiating regress of presuppositions" (Rescher 2010, p. 37). The reason is that the alleged categorical basis, $P(q_n)$, is now pushed further and further back, forever escaping any fixation or determination. To work out $P(q_0)$ one would need to know $P(q_1)$, and to work out $P(q_1)$ one would need $P(q_2)$, and so on, ad infinitum. Admittedly, it is still possible to eliminate successively $P(q_1)$, $P(q_2)$, $P(q_3)$, and so on, as unknown quantities, and this elimination procedure may be performed as many times as one likes. But it is clear that this does not solve anything. For after the eliminations have been carried out $n$ times, where $n$ is a very large number, there would still remain $P(q_n)$ as an unknown quantity. And as long as $P(q_n)$ is unknown, it seems impossible to assess $P(q_0)$. The situation looks like a probabilistic analog 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 (Carroll 1895).

However, this similarity is only apparent: between the probabilistic and the nonprobabilistic version there is an essential difference. The nonprobabilistic version of the Tortoise's challenge to Achilles might be hopeless, the probabilistic one is not. It is true that the Tortoise can always ask about an unknown $P(q_n)$ after the weary warrior has taken $n$ steps in his argument. It is also true that the unknown $P(q_n)$ could have any value between zero and one. However, the influence that $P(q_n)$ has on the value of $P(q_0)$ will be smaller as the distance between $q_n$ and $q_0$ gets bigger — even if $P(q_n)$ were to take on the largest allowed value of 1. In the limit that $n$ tends to infinity, the influence of $P(q_n)$ on $P(q_0)$ will peter out completely, leaving the value of $P(q_0)$ as a function of the conditional probabilities alone. Note again that this is not because $P(q_n)$ itself becomes smaller as $n$ becomes larger: indeed, it may not do so. Nor is it simply because the iteration of (6), (7) and (8), etc., leads to a series of terms that is convergent. Rather it is because $P(q_n)$ is multiplied by a factor that goes to zero as $n$ tends to infinity. For when Achilles has taken one more step, and the Tortoise has asked about $P(q_{n+1})$, this worrisome probability is multiplied by an even smaller factor, and after yet another step the Tortoise's $P(q_{n+2})$ is multiplied by a yet smaller factor still, and so on, until the factor has shrunk to zero.

This can all be proven rigorously, but here is not the place to go into the technical details. Instead, the next section will provide an illustration, based on the example in Rescher's book. It will be shown that, contrary to what Rescher believes, the endless regress of probabilities that he describes *can* reveal what the unconditional probability is of the original proposition $q_0$. By being offered an explanation in terms of a concrete example rather than an mathematical argument *in abstracto*, the reader will hopefully gain a better understanding of what exactly is going on.

## 4 Rescher's example again

In Section 2 it was shown that in Rescher's example the probability of $q_0$ is given by:

$$P(q_0) = P(q_0|P(q_0) = 0.9)P(P(q_0) = 0.9) + P(q_0|P(q_0) \neq 0.9)P(P(q_0) \neq 0.9)\,. \tag{9}$$

With the new notation of Section 3, Eq.(9) becomes:

$$P(q_0) = P(q_0|q_1)P(q_1) + P(q_0|\neg q_1)P(\neg q_1)\,, \tag{10}$$

where both the first and the second terms on the right side of the equation, $P(q_0|q_1)$ and $P(q_1)$, are equal to 0.9. The fourth term, $P(\neg q_1)$, equals $1 - 0.9 = 0.1$. In Rescher's example, the third term, $P(q_0|\neg q_1)$, is not specified, but it will be clear that Eq.(10) cannot be evaluated without it: as long as the value of the third term is unknown, one cannot determine $P(q_0)$. For the sake of argument, this term is here set equal to 0.3:

$$P(q_0|\neg q_1) = 0.3\,.$$

It should be noted that no strings are attached to this choice of 0.3. Whatever nonzero value of $P(q_0|\neg q_1)$ is chosen, the same reasoning will work. So the argument is robust, as is further explained in the Appendix, where some technical details are given.

Now (10) can be worked out:

$$P(q_0) = [0.9 \times 0.9] + [0.3 \times 0.1] = 0.84\,. \tag{11}$$

The number 0.84 was arrived at on the provisional assumption that the second term, $P(q_1)$, indeed equals 0.9, which would be correct if it were the case that

$$P(P(q_1) = 0.9) = P(q_2) = 1\,.$$

But that is wrong, for $P(q_2) = 0.9$. This means that $P(q_1)$ should rather be

$$P(q_1) = [0.9 \times 0.9] + [0.3 \times 0.1] = 0.84\,, \tag{12}$$

where, similarly, 0.3 is the value of $P(q_1|\neg q_2)$, and so on. On the basis of this new result, the value of $P(q_0)$ in (11) must be revised, yielding

$$P(q_0) = [0.9 \times 0.84] + [0.3 \times 0.16] = 0.804\,. \tag{13}$$

However, the number 0.804 was arrived at on the fictional assumption that the second term in (12), to wit $P(q_2)$, indeed equals 0.9, and thus that

$$P(P(q_2) = 0.9) = P(q_3) = 1\,.$$

But that is also wrong, for $P(q_3) = 0.9$. This means that $P(q_2)$ should rather be

$$P(q_2) = [0.9 \times 0.9] + [0.3 \times 0.1] = 0.84\,. \tag{14}$$

On the basis of this, $P(q_1)$ is revised to

$$P(q_1) = [0.9 \times 0.84] + [0.3 \times 0.16] = 0.804\,. \tag{15}$$

This new value for $P(q_1)$ implies that $P(q_0)$ must again be revised, generating

$$P(q_0) = [0.9 \times 0.804] + [0.3 \times 0.196] = 0.7824\,, \tag{16}$$

and so on.

Here is an overview of the values that $P(q_0)$ takes after an increasing number of revisions[9]:

| $n$ | 1 | 2 | 3 | 5 | 10 | 18 | 19 | 20 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|
| $P(q_0)$ | 0.84 | 0.804 | 0.7824 | 0.7617 | 0.7509 | 0.750015 | 0.750009 | 0.750005 | $\frac{3}{4}$ |

**Unconditional probability of $q_0$ after $n$ revisions**

[9]Technically, it is not really necessary to talk of successive revisions of $P(q_0)$ in order to compute the value that the probabilistic regress generates. For the value of $P(q_0)$ can be calculated directly. The details of this calculation are to be found in the Appendix.

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 $P(q_0)$, even in a situation such as the one sketched by Rescher, where

$$P(P(P(q_0) = 0.9) = 0.9) = 0.9\,, \tag{17}$$

and so on. With the value that was chosen for $P(q_n|\neg q_{n+1})$, namely 0.3, and after an infinite number of revisions, $P(q_0)$ 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 $P(q_0)$. For as can be seen in the above table, there is only a small difference between the value of $P(q_0)$ after, say, twenty revisions and after an infinite number of them. Now 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 $P(q_n|\neg q_{n+1})$ 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). But this is only a minor issue. The essential point is that there is always *some* finite number of revisions, such that the result scarcely differs from what is obtained with an infinite number of revisions.

The point can be regarded as a quantitative reinforcement of an important 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 of 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" (Rescher 2010, p. 47).

> "...in actual practice we need simply proceed 'far enough'. After a certain point there is simply no need — or point — to going on" (ibid., p. 82).

> "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' (Rescher 2005, p. 104).

> "...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" (ibid., p. 105).

Rescher's point is a good one, and it can be reinforced by the reasoning above — certainly in the case of an endless hierarchy of probabilities. For the previous argument makes it clear that, 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 $P(q_0)$ after, say, fifteen steps, or after an infinite number of them. Of course, it is on the basis of the context that the meaning of 'negligible' is to be understood. For example, if one is happy to know what a particular probability is to within, say, one percent, then it is quite 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 is related to the second one. It is that the influence of $P(q_n)$ on the value of $P(q_0)$ becomes smaller as $n$ becomes larger. The further away $q_n$ is from $q_0$, the smaller is the influence that the former exerts on the latter. In the limit that $n$ goes to infinity, this influence dies out completely. Since that is the case for any possible value of $P(q_n)$, from zero to one, this means that, in the end, the unconditional probabilities do not affect the value of $P(q_0)$ at all. The only values that do matter are those of the conditional probabilities. Contrary to what Rescher suggests, the unconditional probability of $q_0$ can be fully determined on the basis of the conditional probabilities, and of nothing else.

Again, this could be interpreted as a reinforcement 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" (Rescher 2010, p. 46). 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" (p. 74) is in force, according to which the steps in the regress recede into "a minuteness of size" (p. 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*" (Rescher 2010, p. 48).

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" (ibid., p. 75). Quite the contrary: in the limit that $n$ goes to infinity, as has been shown, it is no impediment if $P(q_n)$ tends to the highest possible value, namely 1. Nor is it the case that, in the limit, $P(q_n)$ fades into penumbral obscurity in which its nature becomes unclear — another way in which, according to Rescher, an infinite regress can be harmless (ibid., pp. 52 – 53). For the nature of the infinitely remote $P(q_n)$ may be perfectly clear and well-defined. Nevertheless a law of diminishing returns can still be said to be in force. As the above argument made clear, there is no problem in determining the value of $v_0$, in $P(q_0) = v_0$, given a regress of the form:

$$\begin{array}{lll} q_1 & \text{is the proposition} & P(q_0) = v_0 \\ q_2 & \text{is the proposition} & P(q_1) = v_1 \\ q_3 & \text{is the proposition} & P(q_2) = v_2\,, \end{array}$$

and so on, ad infinitum. The reason is that the influence of $P(q_n)$ on $P(q_0)$ diminishes as the distance between $q_n$ and $q_0$ increases. It is not that the probability $P(q_n)$ shrinks in size, nor that it becomes dim in some sense. Rather, the *contribution* that $P(q_n)$ makes to the value of $P(q_0)$ vanishes in the limit that $n$ goes off to infinity.

## 5 Conclusion

In the previous sections it has been argued that a major objection to radical probabilism does not hold water. According to this objection, radical probabilism triggers an infinite regress: if every proposition is only probable, then the proposition, $q_1$, that $q_0$ is probable is itself only probable, and the proposition, $q_2$, that $q_1$ is probable is itself merely probable, and so on. In his recent book on infinite regresses, Nicholas Rescher maintains that this regress is vicious, since it effectively blocks any knowledge about the condition of the original proposition $q_0$. The present paper has demonstrated that this conclusion is not correct. Contrary to what Rescher claims, the infinite regress precipitated by radical probabilism is in fact quite harmless. For the probability of $q_0$ can be exactly determined, notwithstanding the fact that $q_n$ is taken to the infinite limit. Moreover, it has been pointed out that this result actually strengthens, rather than weakens, Rescher's overall analysis.

However, a classic objection remains. If every proposition is only probable, then what is the status of the proposition that an endless hierarchy of probabilities can yield an exact unconditional probability value for $q_0$? Applied to the example examined in Section 3: what is the status of the conclusion that $P(q_0)$ equals $\frac{3}{4}$ in the limit that $n$ is taken to infinity? Claiming that this conclusion is certain appears to be at loggerheads with radical probabilism itself; but claiming that it is only probable may seem to be contrary to what has been argued above.

To resolve this difficulty one first has to realize that the exact value of $\frac{3}{4}$ in the example was obtained on the basis of two exact values for the relevant conditional probabilities, viz. 0.9 and 0.3. In general it is the case that an unconditional probability is determined exactly only if the conditional probabilities themselves are exactly specified. If the conditional probabilities are not precisely defined, then the value obtained for $P(q_0)$ will likewise be imprecise.

The situation in which exact values for conditional probabilities are known is however quite rare and also somewhat artificial: usually one only disposes of probability distributions, which extend from zero to one and which normally possess a single peak of a particular width. If the conditional probabilities are specified in so far that their probability distributions are given, then the unconditional probability distribution associated with $q_0$ can be deduced; but of course this does not mean that the value of $P(q_0)$ has been precisely specified: rather its mean value can be calculated, together with its standard deviation, which is a measure of its uncertainty. This imprecision is the very essence of probabilism — and it is the heart and sinew of statistics. Despite the higher-order uncertainty, it is not the case that we know nothing.

## 6 Acknowledgements

We would like to thank Nicholas Rescher and an anonymous referee for very helpful comments.

## Appendix

In Section 4 an overview has been given of the values that $P(q_0)$ takes after an increasing number of revisions. The purpose of the present Appendix is to explain the claim in endnote 9, namely that there is a more straightforward way to calculate these values.

The generalization of Eqs.(6)-(8) to the $n$th step of the hierarchy is

$$P(q_n) = v_n P(q_{n+1}) + w_n P(\neg q_{n+1})\,, \tag{18}$$

in which the following conditional probabilities appear:

$$v_n = P(q_n|q_{n+1}) \qquad\qquad w_n = P(q_n|\neg q_{n+1})\,. \tag{19}$$

This iterative relation can be solved to yield a solution for $P(q_0)$ in terms of the $v_n$ and the $w_n$, and general conditions have been given for the existence of a unique solution (Peijnenburg et al. 2008, Atkinson et al. 2010). In the example that was considered in Section 4, it was supposed that the conditional probabilities $v_n$ and $w_n$ are the same from step to step of the iteration, so the subscript, $n$, may be dropped, with the result

$$P(q_n) = vP(q_{n+1}) + wP(\neg q_{n+1})\,. \tag{20}$$

In the example of Rescher, $v = P(q_n|q_{n+1}) = 0.9$; and, as an illustration of the argument, the value $w = P(q_n|\neg q_{n+1}) = 0.3$ was assumed. In this appendix, however, no appeal to such special values will be made: it will now be shown that $P(q_0)$ is well-defined – and nonzero – for any $v$ and $w$ satisfying the constraints

$$0 < w \leq 1 \qquad \text{and} \qquad 0 \leq v \leq 1\,. \tag{21}$$

The demonstration starts by the replacement of $P(\neg q_{n+1})$ in Eq.(20) by $1 - P(q_{n+1})$, after which the terms in $P(q_{n+1})$ are gathered together. The result is

$$P(q_n) = w + (v - w)\,P(q_{n+1})\,.$$

This equation is first used with $n = 0$, then with $n = 1$, then with $n = 2$, and so on:

$$\begin{aligned} P(q_0) &= w + (v-w)\,P(q_1) && (22)\\ &= w + (v-w)w + (v-w)^2\,P(q_2) \\ &= w + (v-w)w + (v-w)^2 w + (v-w)^3\,P(q_3) \\ &= \ldots\ldots && (23)\\ &= w\Big[1 + (v-w) + (v-w)^2 + (v-w)^3 + \ldots (v-w)^n\Big] + (v-w)^{n+1}P(q_{n+1})\,. \end{aligned}$$

The values in Table 1 of Section 4 can be obtained by replacing $P(q_{n+1})$ in the last line by $v$, and then by substituting the special values, $v = 0.9$ and $w = 0.3$.

In view of condition (21), it follows that $|v-w| < 1$, and therefore that the factor $(v-w)^{n+1}$ in the last line of Eq.(23) tends to zero in the limit that $n$ tends to infinity. In this infinite limit,

$$P(q_0) = w\Big[1 + (v-w) + (v-w)^2 + (v-w)^3 + \ldots\Big] = \frac{w}{1 - v + w}\,,$$

which shows that the probability is indeed determined uniquely in terms of the conditional probabilities $v$ and $w$. Note that the value thus determined is nonzero, since $w > 0$ . This result is robust, in the sense that it holds for any values of $v$ and $w$ that satisfy condition (21).

*University of Groningen, The Netherlands.*

# References

Atkinson, David, and Jeanne Peijnenburg. 2010. "The Solvability of Probabilistic Regresses. A Reply to Frederik Herzberg," *Studia Logica*, vol. 94, no. 3, pp. 347-353.

Carroll, Lewis. 1895. "What the Tortoise Said to Achilles," *Mind*, vol. 4, no. 14, pp. 278–280.

DeWitt, Richard. 1985. "Hume's Probability Argument of I, iv, 1," *Hume Studies*, vol. 11, no. 2, pp. 125–140.

Domotor, Zoltan. 1981. "Higher Order Probabilities," *Philosophical Studies*, vol. 40, no. 1, pp. 31–46.

Hájek, Alan. 2003. "What Conditional Probability Could Not Be," *Synthese*, vol. 137, no. 3, pp. 273-323.

Hume, David. 1738/1961. *A Treatise of Human Nature*, in two volumes, Volume I. Introduction by A.D. Lindsay. (London: J.M.Dent).

Lehrer, Keith. 1981. "The Evaluation of Method: A Hierarchy of Probabilities among Probabilities," *Grazer Philosophische Studien*, vol. 12-13, pp. 131–141.

Peijnenburg, Jeanne, and David Atkinson. 2008. "Probabilistic Justification and the Regress Problem," *Studia Logica*, vol. 89, no. 3, pp. 333-341.

Rescher, Nicholas. 2005. "Issues of Infinite Regress," in Nicholas Rescher, *Scholastic Meditations.* (Washington: The Catholic University of America Press), pp. 92–109.

Rescher, Nicholas. 2010. *Infinite Regress. The Theory and History of Varieties of Change.* (New Brunswick: Transaction Publishers).

Savage, Leonard J. 1954. *The Foundations of Statistics.* (New York: Dover, 1972, second revised edition).

Skyrms, Brian. 1980. "Higher Order Degrees of Belief," in D.H. Mellor (ed.), *Prospects for Pragmatism: Essays in Memory of Frank Ramsey.* (Cambridge: Cambridge University Press), pp. 109–137.

# Notes

[1]We follow Rescher in considering a point determination of a continuous probability; but, as Alan Hájek remarks, "The probability of a continuous random variable $X$, taking a particular value $x$, equals 0, being the integral from $x$ to $x$ of $X$'s density function" (Hájek 2003, p. 286). It would be apposite to replace the point determination by an interval, say $P(q) \in [v_0 - \varepsilon, v_0 + \varepsilon]$. The reader is invited to carry this refinement along in a charitable reading of our text, here and in the sequel.

[2]Rescher 2010, pp. 36–37. Rescher has $p$ rather than $q$, and $v$ instead of $v_0$. Furthermore, Rescher explicitly conditions all his probabilities with respect to some evidence, $E$, so instead of Eqs.(1)–(2) he has $Pr(p|E) = v$ and $Pr(Pr(p|E) = v|E) = v_1$ (ibid., p. 36 — misprint corrected). In the interests of notational brevity, explicit reference to $E$ is suppressed, but it is to be understood.

[3]Hume 1738/1961 (Book I, Part IV, Section I), p. 178. See also Lehrer 1981.

[4]As indicated in endnote 1, one might better replace the point determination $P(q) = 0.9$, by an interval, $P(q) \in [0.9 - \varepsilon, 0.9 + \varepsilon]$.

[5]See for example DeWitt 1985, especially pp. 128–129.

[6]If, along the lines of endnote 4, $P(q) = 0.9$ is to be replaced by $P(q) \in [0.9 - \varepsilon, 0.9 + \varepsilon]$, then $P(q) \neq 0.9$ must of course be replaced by $P(q) \notin [0.9 - \varepsilon, 0.9 + \varepsilon]$.

[7]Miller's Principle should not be confused with the Principal Principle of David Lewis. The two principles look alike, but the former is about a subjective probability of a subjective probability, whereas the latter concerns a subjective probability of an objective probability or chance. The fact that, according to Miller's Principle, the first term in Eq.(5) is equal to 0.9, might have caused the confusion in papers such as DeWitt's (see endnote 5).

[8]The phrase "a categorical factual basis in the form of $q_3$ or $P(q_3)$" is employed here in order to allow for a certain ambiguity in Rescher's use of 'categorical'. For in Rescher 2010 the term is used in two ways: firstly in opposition to 'probabilistic' (as in "But if these requisites themselves are never categorical but only probabilistic" on p. 37), secondly in opposition to 'conditional probability' (as in "the thesis that one can never determine categorical probabilities but only conditional ones" on p. 40). In the latter case, an unconditional probability is the same as a categorical probability. Under the first meaning, Rescher's "categorically established factual base of some sort" in the example would be $q_3$, under the second meaning it would be $P(q_3)$.

[9]Technically, it is not really necessary to talk of successive revisions of $P(q_0)$ in order to compute the value that the probabilistic regress generates. For the value of $P(q_0)$ can be calculated directly. The details of this calculation are to be found in the Appendix.