Reasoning About the Future: Doom and Beauty

Dennis Dieks
History and Foundations of Science, Utrecht University
P.O.Box 80.000, 3508 TA Utrecht

October 14, 2007

Abstract

According to the Doomsday Argument we have to rethink the probabilities we assign to a soon or not so soon extinction of mankind when we realize that we are living now, rather early in the history of mankind. Sleeping Beauty finds herself in a similar predicament: on learning the date of her first awakening, she is asked to re-evaluate the probabilities of her two possible future scenarios.

In connection with Doom, I argue that it is wrong to assume that our ordinary probability judgements do not already reflect our place in history: we justify the predictive use we make of the probabilities yielded by science (or other sources of information) by our knowledge of the fact that we live now, a certain time before the possible occurrence of the events the probabilities refer to. Our degrees of belief should change drastically when we forget the date—importantly, this follows without invoking the "Self Indication Assumption". Subsequent conditionalization on information about which year it is cancels this probability shift again. The Doomsday Argument is about such probability *shifts*, but tells us nothing about the concrete values of the probabilities—for these, experience provides the only basis. Essentially the same analysis applies to the Sleeping Beauty problem. I argue that Sleeping Beauty "thirders" should be committed to thinking that the Doomsday Argument is ineffective; whereas "halfers" should agree that doom is imminent—but they are wrong.

1 The Doomsday Argument

There are dangers threatening the future of mankind. Using information about the political situation in the world, the risks of nuclear technology, the possibility of epidemics, and so on, or perhaps relying more on our intuition or even astrology, we may attempt to become precise about how serious we think these dangers are. We can do so by assigning a probability to the hypothesis that mankind will become extinct soon, within the present century, say. Let us assume that there is only one alternative hypothesis: that the human civilization will flourish for a very long time, with population figures that grow and grow as time goes on. Perhaps the human race will in this case colonize other planets, or even galaxies, to find room for all humans alive in the future.[1]

Suppose I have determined probabilities that measure my degrees of belief in these two hypotheses. Then the following thought occurs to me. In the evaluation of my probabilities I have not taken into account that I live now, in the beginning of the third millennium AD. Of course, I have used information about the present state of the world; but I have not paid attention to the fact that *I myself* happen to live at this particular point in time. This is, however, relevant additional information. For according to the Doom Soon hypothesis, my place in history makes me a fairly typical human being: a significant proportion of all humans ever alive will live in this century if the Doom Soon scenario corresponds to the actual fate of the world. By contrast, the number of people living in the far future will be overwhelmingly great if the Doom Late hypothesis is true. So on the latter hypothesis the situation in which I actually find myself is a very unlikely one. As a general methodological principle, I should prefer the hypothesis that endows the available evidence with the highest probability. Taking into account that I live now should therefore move my odds appreciably in favor of the Doom Soon scenario.

This is the Doomsday Argument, in the form defended by John Leslie. It has been around in the literature for more than fifteen years ([8, 9]) and

[1] Assuming that there are only two scenarios is of course unrealistic. A more natural assumption would be that there are many hypotheses to consider, each one with its own prediction for when mankind will die out. However, we could collect these hypotheses in two groups: those predicting doom before a certain date, and those predicting doom later. We would thus effectively end up with two possibilities, and the argument in the text would apply. Nothing essential therefore hinges on the restriction to two rival scenarios.

continues to be discussed as a controversial, but basically unrefuted line of reasoning [1, 2, 3, 14, 4, 13, 15]. In his book, *The End of the World: The Ethics and Science of Human Extinction* ([10]), Leslie draws the conclusion that we should do our best to counteract the just-described probability shift by making the Earth a safer place.

The Doomsday Argument can be regarded as an example of Bayesian reasoning. The proponent of the Argument asks us to consider two hypotheses, H_1 and H_2, Doom Soon and Doom Late, respectively. These are assigned prior probabilities p_1 and p_2. When new evidence E is taken into account, the probabilities shift in accordance with Bayes's rule:

$$P(H_i/E) = \frac{p_i.P(E/H_i)}{p_1.P(E/H_1) + p_2.P(E/H_2)}, \quad i = 1, 2. \tag{1}$$

Here $P(E/H_i)$ is the probability that E will occur if hypothesis H_i is true. $P(H_i/E)$ is the a posteriori probability that we should assign to H_i once we know that E actually occurred.

To see the seriousness of the doomsayer's conclusions, suppose that on the basis of our ordinary evidence we think that Doom Late has a probability of 99%, so $p_1 = 1/100$ and $p_2 = 99/100$. Let us denote by N_1 the number of human beings ever alive according to H_1; and let the corresponding number according to H_2 be denoted by N_2. The proponent of the Argument wants us to consider the case in which N_2 is very much larger than N_1, $N_2 = 10^6 N_1$, say. (The Argument's conclusions become the more pessimistic the more optimistic the scenario H_2 is.) Finally, denote by n the number of people alive in the present year. (We have to assume that both hypotheses, H_1 and H_2, predict the same population figures at least up to and including the present year—otherwise we could falsify at least one of the hypotheses directly by inspection of past or present population numbers.)

The doomsayer now reasons as follows. I am an arbitrary member of mankind, and can consider myself as a random selection from all people ever alive. Therefore, according to H_i the probability of actually finding myself in the present year is n/N_i. I can treat the fact that I live now as new evidence E, because I did not reflect on my place in history when I first estimated how probable H_1 and H_2 were. So I can use Bayes's formula to update my probability estimates, with $P(E/H_i) = n/N_i$. For the posterior probability

of Doom Soon, I thus find

$$P(H_1/E) = \frac{(1/100).(n/N_1)}{(1/100).(n/N_1) + (99/100).\{n/(10^6.N_1)\}}$$
$$= \frac{1/100}{1/100 + 99/10^8} = 0.9999\,.$$

Therefore, it becomes virtually certain to me that the extinction of mankind is imminent.

2 The essential weakness in the Argument

In ([6]) it was pointed out that the doomsayer should be careful to formulate his hypotheses H_1 and H_2 in a non-indexical way: if H_1 says "Doom will strike *soon*", that would presuppose my existence now, because "soon" means "a short time from now". But if the hypotheses already contain information about my existence in the present year, subsequent conditionalization on that information will not lead to a shift in probabilities (because $P(E/H_i) = 1$ in that case). So the Argument can get off the ground only if the hypotheses have a form like: "Doom will strike soon/late *after 2005*", say. Moreover, the Argument needs the supposition that both hypotheses leave it completely open when in human history I am living: I may turn out to be any human in the history of mankind, and all these possibilities are equiprobable. This is behind taking n/N_i as the probability, on hypothesis H_i, that I live in the year I actually live in.

It might initially be doubted whether it makes sense at all to consider myself a random member of the human race: my existence here and now may well be the outcome of a deterministic process. But this is not the weak point of the Argument. What is at stake are credences, probabilities that measure the strength of my belief. A situation in which I am completely uncertain about my place in human history is certainly logically possible.

The initial probabilities in the Argument should therefore refer to a situation in which I consider two possible scenarios about the fate of mankind after 2005, say, without having any clue about the present date. When I subsequently add information about when I live, this leads via Bayes's rule to a probability shift. Whether the resulting posterior probabilities are different from my usual ones depends, of course, on the concrete values of the

prior probabilities I started with. What the proponent of the Argument has to assume in order to arrive at his alarming conclusion is that we should use our usual probability values as priors, and that the posterior probabilities that we end up with after the Bayesian shift are the ones that really apply to our actual situation. The doomsayer justifies this assumption about the priors by pointing out that in making our actual probability evaluations we never reflect explicitly on the fact that we live now—it appears that we make our actual calculations while leaving our place in history open.

As I will explain in detail in the following sections, this claim about the situation to which our actual probabilities refer constitutes the essential weakness in the Argument. To see the basic difficulty already now, think a bit more about the situation the prior probabilities in the Argument pertain to. It is a situation in which we are asked to consider the hypothesis "the human race will become extinct soon after 2005", and assign a probability to it. We are allowed to use evidence available to humans living in 2005 (the dangers of technological developments in 2005, wars going on in that year, etc.), while we must also assume that we could subsequently get the information that we ourselves are cavemen living very, very long before 2005; or humans drastically changed by evolution, living millions of years later. This is a bizarre situation, very different from our actual one. It is not immediately obvious that our probability assignments in this strange predicament should be equal to our usual ones. In particular, we have to take into account that we could be living *after* the date of Doom Soon; but in this case our ordinary probability judgements (the ones we use now, in our actual situation) are irrelevant for our credence. The Doom Late scenario gets probability 1 if we know that it is later than Doom Soon. We should factor this in in our probability considerations—as we will do below—with the foreseeable result that the probabilities will become different from our actual ones. The resulting difference will turn out to be exactly such as to balance the Bayesian shift that ensues when we subsequently conditionalize on our place in history. Consequently, the final net results are the probabilities that we entertained before this whole issue of the Doomsday Argument arose.

There is a line of argument in the literature (starting with [6], see [3, p. 122] for further references) that similarly says that the prior probabilities in the Doomsday Argument should be different from our actual ones, in precisely such a way that we are led back to our usual probabilities after conditionalization on information about the date. This line of argument

is based on what has become known as the "Self Indication Assumption" (SIA): knowledge that I exist favors hypotheses that predict populous human civilizations over hypotheses according to which there will not be so many humans [3, p. 66]. According to SIA the probabilities of the various hypotheses should be taken as proportional to the numbers of observers the hypotheses predict.

As Bostrom ([3, pp. 122-126]) has correctly observed, however, SIA is untenable when interpreted as a general *a priori* principle that would apply even if there is no subsequent balancing by a Doomsday reasoning: Such a principle would entail, e.g., the unpalatable conclusion that armchair philosophizing would suffice for deciding between cosmological models that predict vastly different chances for the development of human civilization. The infinity of the universe would become certain *a priori*.

The argument of this paper does not make use of such a general Self Indication Assumption about priors. Although it is true that we will arrive at the same prior probabilities as SIA would give us when applied to the Doomsday Argument, we will base our conclusions exclusively on a consideration of the consequences of *a lack of information about the date*, together with an analysis of the status of our ordinary probability estimates. The objection that SIA can lead to absurd results when used as a general independent rule does not touch our argument, as we will discuss below.

3 The relevance of our knowledge of the date

That the lack of information about the date, characteristic of the initial situation in the Doomsday Argument, is very relevant to the prior probabilities to be assigned can easily be seen.

Suppose that I have been hypnotized and do not know when in human history I am living. I firmly believe that I may turn out to be any one of all individuals ever alive, be it Napoleon, a prehistoric troglodyte, or someone living in 130000 AD (if human civilization will last that long); I think that every possibility is equally likely. Suppose further that I have been told that physical theory predicts a probability of 1/2 that the world will end in early 2006, given the state of the world in 2005; and an equally large probability that civilization will go on for a very long time after 2005. It could be, for example, that a quantum mechanical spin measurement is performed

in the last minutes of 2005, with possible outcomes +1 and −1. If the outcome +1 is realized, a destructive device is automatically activated that annihilates the world within one hour; if −1 is the result, physical conditions then obtaining will make it certain that the human race will prosper for many thousands years. The best available physical theory, quantum mechanics, says that for each of the two possibilities the chance of being realized is 1/2, given the details of the experimental set-up. I now want to determine prior probabilities for my two hypotheses; probabilities that I can update when information about who I am will come in later. What probability should I assign to the possibility that the world ends in 2006 (Doom Soon)?

Let us consider what happens if I decide to follow common-sense strategy, and set my credence equal to the quantum chance[2]. I thus take 1/2 as my probability of Doom Soon. When I subsequently learn that it is 2005 AD, and update my probability with the help of Eqn. 1 in the way explained before, doom in 2006 becomes a virtual certainty to me.

So I appear to have justified the doomsayer's conclusion. However, a problem occurs to me. I have been assuming that it may now be any year in the history of mankind. But that implies that it may be a year far after 2006, the date of Doom Soon. However, if the actual date is some year *after* the date of Doom Soon, the outcome of the chance experiment must have been −1. In this case, the probability to be assigned to the outcome +1 would certainly not be 1/2; it would be 0. I have to make a more sophisticated calculation, taking the just-mentioned complication into account.

I denote the probabilities whose values I am going to determine by p_1 and p_2, pertaining to doom in 2006 and doom later, respectively. There are two possibilities for me to consider. First, I may be living before the date of Doom Soon, in which case the experiment is still going to take place and I can use my information about the physical chances to gauge my belief (if I accept the aforementioned common-sense strategy). In this case I will judge Doom Soon and Doom Late as equiprobable. Second, I may be living after the date of Doom Soon, in which case the physical chance is irrelevant for my belief in the truth of the hypotheses. In the latter case the probability that I will assign to hypothesis H_1 will be zero. My probability p_1 is therefore

[2]Quantum mechanics is mentioned only for illustrative purposes; the same argument would apply if the probabilities were derived from the prediction of a fortune teller. We shall return to the justification of equating subjective probabilities to physical chances in section 5.

equal to

$$P(H_1|\,\text{I live before Doom Soon}).\,P(\text{I live before Doom Soon})$$
$$+P(H_1|\,\text{I live after Doom Soon}).\,P(\text{I live after Doom Soon}).$$

In other words:

$$p_1 = P(\text{I live before the date of Doom Soon}).\,1/2$$
$$+P(\text{I live after the date of Doom Soon}).\,0. \qquad (2)$$

The probability that I live before Doom Soon is itself the sum of two probabilities, pertaining to two mutually exclusive cases: either H_1 is true, in which case I will certainly live before Doom Soon, or H_2 is true, in which case the probability that I, a human randomly selected from my long-lasting civilization, live before the date of Doom Soon is N_1/N_2.

Things are simple now. It follows from what has just been said that the belief I should attach to the possibility of finding myself in some year before the date of Doom Soon is represented by $p_1 + p_2.N_1/N_2$. Inserting this in Eq. 2, I obtain:

$$p_1 = (p_1 + p_2.N_1/N_2).1/2,$$

or, equivalently, $p_1 = p_2.N_1/N_2$. Therefore, the assumption that I lack all information about my place in history, together with consistency, compel me to adopt a value of the a priori probability of Doom Soon that is very much smaller than the a priori probability of Doom Late!

Substituting the result in Bayes's formula, I find for the *a posteriori* probability of Doom Soon, after I have taken the new information that I live in 2005 into account,

$$P(H_1/E) = \frac{p_1.n/N_1}{p_1.n/N_1 + (N_2/N_1).p_1.(n/N_2)} = 1/2.$$

So when I finally have the same information as I would have had without the intervention of the mesmerist, I think that H_1 and H_2 are equiprobable—just as I would have thought if I had not been hypnotized in the first place.

Let me briefly rehearse the point for the general case. Consider the situation in which I decide, in full awareness that it is now 2005, and on the basis of whatever further considerations (be they scientific or not), to assign probabilities q_1 to H_1 (Doom Soon) and q_2 to H_2 (Doom Late), respectively. If

my memory is partly erased, so that I no longer know the date, or if I decide to discard my information about the date, I have to revise my probabilities. In the new situation q_1 and q_2 are still the probabilities I *would* assign to the possible fates of the world if I knew it were 2005 now; but I do not know what year it is. I might be living after the date of Doom Soon, in which case the probability I have to assign to Doom Soon is not q_1 but 0.

Making a calculation in the same way as before, I find that the probability I should assign to the Doom Soon hypothesis in the new situation, p_1, should satisfy

$$\begin{aligned} p_1 = {} & P(\text{I live before the date of Doom Soon}).\, q_1 \\ & + P(\text{I live after the date of Doom Soon}).\, 0. \end{aligned} \tag{3}$$

Furthermore, $P(\text{I live before the date of Doom Soon}) = p_1 + p_2.N_1/N_2$. Using this in Eqn. 3 I find that my new probabilities for the two hypotheses should be:

$$p_i = \frac{q_i.N_i}{q_1.N_1 + q_2.N_2}\,,\ i = 1,\, 2. \tag{4}$$

Subsequent conditionalization on the evidence about which year it is now, by means of Bayes's rule with p_1 and p_2 as priors, leads to q_1 and q_2 as the posterior probabilities. These are the same probabilities I had assigned directly, knowing it is 2005.

The main point is that it is inconsistent to assign the same probabilities to hypotheses about what is going to happen after a certain date *both* in the situation in which I do not know my place in history *and* in my actual situation, in which I know that the events in question have not yet occurred. Probability theory specifies a definite relation between the probabilities in the two cases. This relation is such that the probabilities in my actual situation can be recovered from the probabilities that apply if I do not know which year it is by Bayesian updating with evidence about the present year. This point is quite general, and does not depend on the basis of my assignment of probability values—whether I use scientific results, astrology, or intuition does not matter because what is at stake is the consistency of my probabilistic reasoning rather than the concrete values of my probabilities.

The Self Indication Assumption, in the general form discussed by Bostrom [3], has not been used in the above. SIA says that the probabilities of hypotheses should always be taken as proportional to how many observers they

say that exist (given the fact that I myself am an observer). It is true that in our example we found that $p_2/p_1 = N_2/N_1$ in the initial situation of the Doomsday argument; but this conclusion was not drawn on the basis of SIA, but on the basis of a consideration of the relevance of knowledge of the date. Our analysis does not lead to SIA as a general principle, but only says that a hypothesis that leaves my position in time entirely open, and which I consider without any knowledge of the date, should be assigned a probability that differs from my ordinary one by a factor that is proportional to the population size the hypothesis predicts. If we consider, in our actual situation, theories that predict different population numbers, this result does not provide any reasons for assigning probabilities that differ from our ordinary degrees of belief. The objection that SIA would enable presumptuous philosophers to decide about the credibility of cosmological models by just looking at the numbers of humans they predict is therefore not effective against our argument.

4 The justification of our ordinary probability estimates

The analysis of the previous sections emphasized the inconsistency of using the same probability values both in our actual situation and in the situation in which we lack information about when we are living. If we start with our actual probability judgements, correct them to obtain probabilities for the situation in which we lack knowledge about the year we are living in, and finally apply Bayes's rule to add the missing information again, we end up with the probability we started with—exactly as it should be. Now, the point remains that if we became convinced that our usual probabilities do not really pertain to our actual circumstances but rather to the situation in which we lack information about the date, Bayesian upgrading would lead to a probability of Doom Soon in our actual situation that is much larger than our usual one. So the force of the Doomsday Argument depends essentially on the question of whether we are justified in thinking that our usual probability judgements pertain to our actual situation. The doomsayer's basic claim is (or should be) that they do not; if he is right about this, it is a matter of course that we should urgently revise our probabilities.

We already heard one argument for the doomsayer's basic claim—namely that in arriving at our usual probability assignments we have made no use of our knowledge that we live now—and showed that it is fallacious. We *do* use the information that we live now. If we had thought that we could be living anytime in human history, we should have taken into account that we could be living *after* the date of Doom Soon. In the latter case the scientific, astrological, etc., data (of the form: if the situation at t is this, then the situation at $t + \triangle$ will be that) that we actually use have no relevance for our credence; in this case we should calculate revised probability values, in the way explained. We do not make such calculations, and this demonstrates that we make use of our knowledge about the year we are living in.

Still, we might be mistaken in the use we make of the data and the doomsayer might challenge us to provide a fundamental justification of our ordinary probability estimates. After all, he himself thinks it is rational to believe in a much higher probability of Doom Soon than the value provided by science [2]. The first thing we have to observe in this connection is that as long as purely *subjective* probabilities are at stake, there can be no question of disproving the doomsayer's probability values. The doomsayer has the right to fancy any subjective belief he likes. The dispute about the justification of our actual probability values becomes interesting only if the probabilities are not purely individual.

The obvious way of making our probability judgements objective is by relating them to observed relative frequencies or to probabilities predicted by science. The hypothesis that there is a correlation between our own rank in series of events involving humans and the total number of humans in such series can be tested empirically; as argued by Sober, there are no indications that available empirical evidence supports the existence of such correlations [15]. The following additional example illustrates the situation. Suppose that there is a country in which families have either one child or two children; this has been an empirical fact since times immemorial. I am my mother's first child and have lost contact with my family soon after my birth. I ask myself: what is the probability that I have a brother or sister? I look at the birth statistics and find out that exactly half of all families have one child, and half have two children. I conclude that the probability I am looking for is 1/2. I feel quite confident about this, because of all families I could be a member of 50% have one child and 50% have two. But now someone asks me: "Have you fully considered the special status you have?

You are a first-born, and this makes it more probable that you are alone—the chance of being a first-borne is only 1/2 if you have a brother or sister, and 1 if you are alone." Swayed by this, I use Bayes's formula to update my belief and conclude that the probability that I have a brother or sister is only 1/3. Now, I am entitled to believing whatever I want; but if I am going to base predictions on this new probability I will place myself in a bad position. If all first-borns use this value, a radically wrong prediction of the total number of children will be only one of the consequences. What went wrong is that I overlooked that my original $p = 1/2$ judgement already took into account that I am a first-born child. Indeed, if I forget about this piece of information my initial probability should be different: on average, two out of every three children have a brother or sister, so my probability of not being alone becomes 2/3. Needless to say, when I use Bayes to update this probability with the evidence that I am a first-born child, I return to my original $p = 1/2$. The case closely parallels that of the Doomsday Argument, and also that of the Sleeping Beauty problem of the next section. The moral is that what the right probability estimate is, in an objective probability dispute, is determined by empirical data. The relative frequency of first-borns who possess a brother or sister is 1/2, and this provides me with the justification of my credence.

Probability statements coming from scientific theories function in the same way. Such scientific probability judgements are relational: they state a relation between the state of the world at a certain instant and the probability of an event at a later time. The example described in section 3 illustrates the general point. There it was supposed that our best fundamental physical theory supplied a probability value of 1/2 of Doom Soon (in 2006), given the state of the world in 2005. If we know that we are living in or before 2005, we can use this value to guide our credence. More generally, the scientific probability of something happening in the year $t + \triangle$ given the situation in year t, can be used for gauging our belief that the event in question will take place if we know that we live in the year t.

The doomsayer objects to this procedure. What reasons does he have? As we have seen, the putative fact that we have not used our information about our position in time does not afford a valid ground. But the doomsayer will be hard-pressed to supply another reason. Acceptance of a probabilistic physical theory entails the belief that the predicted probabilities will materialize as relative frequencies in the long run. If we *accept* a physical theory that

leads to probability statements, and are intent on assigning probabilities in an objective way, without making claims that would cost us money if we based bets on them, it would therefore be inconsistent not to use the physical probabilities as the values for our credence function. This is Lewis's Principal Principle [11], about which more will be said in the next section.

The upshot of all this is that there are good empirical reasons for relying on the usual experimental and scientific probability values. By contrast, all the doomsayer offers to convince us of the errors of our ways is the incorrect idea that in our actual probability assignments we have not fully used our information about our own place in history, and the Doomsday reasoning, which is irrelevant because it is only about the *relation* between prior and posterior probabilities and does not touch upon the question of what the concrete values of our probabilities should be.

5 Sleeping Beauty

The most interesting aspect of the doomsday argument is that it casts light on the role played by information about the date in probabilistic reasoning about possible scenarios. It shares this feature with the Sleeping Beauty problem, which has been attracting much attention recently. Indeed, I will argue that the two problems are structurally the same and that the analysis of one can be carried over to the other.

In the Sleeping Beauty problem we are asked to assign probabilities to the possible outcomes of a fair coin toss, in a situation in which we do not know the date; subsequently, information about the date is supplied and we are asked to update our probabilities. The story is as follows. Sleeping Beauty is put into deep sleep late Sunday night. She is awakened on Monday; after some time she will be told that it is Monday (at first, when she awakes, she has no clue about the date). After this she is put into sleep again, and her memories of the awakening are erased. Then, depending on the outcome of the toss of a fair coin, she will *either* be awakened again on Tuesday (when everything looks to her the same as on Monday), put into sleep again and sleep until Wednesday when the experiment ends (Tails), *or* will not be wakened a second time and sleep on until the end of the experiment (Heads). Beauty knows about this protocol before the experiment starts, on Sunday.

Immediately after waking up on Monday, what should be Beauty's proba-

bilities for the hypotheses that the coin lands heads or tails (H or T), respectively? And what should these probabilities become when she learns that it is Monday?

If Beauty decides to calibrate her credence on the basis of the physical data available to her, she will assign probabilities 1/2 to H and T on Sunday. This can be regarded as an application of the Principal Principle: Beauty has only admissible information about the future toss of a fair coin. So when she falls asleep, $p(H) = p(T) = 1/2$. Now, one is perhaps inclined to argue that during her sleep until her first awakening, and also in this first awakening, Beauty does not receive any new information: she knew for certain that she would be awakened all along. So upon waking up on Monday her probabilities of H and T should still be 1/2 (this position was defended by Lewis, [12]).

However, something *did* change when Beauty fell asleep: she lost information about the date, like the mesmerized protagonist of the story in section 3. When Beauty wakes up, she does not know whether it is Monday or Tuesday. When she subsequently learns that it is in fact Monday, this constitutes new and relevant information. Her odds should shift accordingly, namely in favor of H: On hypothesis H, Monday is the only day on which an awakening can take place, whereas on T the awakening could have been on Tuesday as well. So H makes the fact that the awakening is on Monday more "typical" than T. We can calculate the associated probability shift by Bayesian updating—but obviously, the values of the resulting probabilities depend on the values of the priors we use. If we follow Lewis's approach, and assume that Beauty's probabilities are still the same on Monday, when she awakes, as what they were on Sunday—in other words, if we take $p(H) = p(T) = 1/2$ as priors—we find $p(H|M) = 2/3$ and $p(T|M) = 1/3$, where M stands for the evidence that it is Monday. So Beauty is going to believe that the future toss of the fair coin will more probably than not result in heads.

We may consider a variation on the Sleeping Beauty problem in which the experiment takes many days and in which Beauty is awakened day after day if the coin lands tails, whereas she remains asleep after the awakening on Monday if the coin lands heads. The necessary precautions are taken to ensure that Beauty does not remember anything about previous awakenings. Let n be the number of days the experiment lasts. Adapting the above argument to the new experiment, we conclude that it becomes practically certain to Beauty that H is true when she learns that it is the first Monday after the Sunday the experiment started. This follows because according to

T the probability that her awakening is on this Monday becomes smaller and smaller when n grows. Bayes's formula yields $p(H|M) = n/(n+1)$, if we take $p(H) = p(T) = 1/2$ as priors again; this probability approaches 1 when n increases.

The analogy to the Doomsday Argument is obvious. Saying that it should not make a difference to Beauty's probabilities whether it is Sunday or whether it is Monday immediately after her awakening (when she does not know the date) is tantamount to saying that I should use my usual probabilities (think of the 1/2 probabilities in the quantum example) even when I do not know the date in the initial situation of the Doomsday Argument. Just as Beauty becomes virtually certain about Heads when she hears it is Monday, I become sure of Doom Soon when I take the present date into account. Defenders of the 1/2 line in the Sleeping Beauty problem (in the original version) should therefore also subscribe to the doomsayer's conclusion that Doom Soon is imminent.

Repetition of the experiment, which can be done in the case of Sleeping Beauty, unlike in the case of Doom, will result in approximately 50% of heads and 50% of tails after Beauty, having heard it is Monday, expresses her credence—the coin is fair. This is also true, of course, in the variant of the experiment with n days. In the latter case Beauty will practically never be right in her predictions if she adheres to the policy explained above: she will expect tails in only 1 out of $n+1$ cases. Also in the original version of the experiment her credences will systematically deviate from the actual relative frequencies. So if the probabilities assigned by Beauty should reflect what will happen in repetitions of the experiment, something has gone seriously wrong.

Our preceding discussion of the Doomsday Argument contains the diagnosis of the problem. Taking the probabilities as 1/2 on awakening on Monday means thinking that there was no relevant change in information since Sunday afternoon. But this is wrong. On Sunday Beauty knew that it was Sunday, and that a fair coin would be tossed in the future. She was justified in applying the Principal Principle then. By contrast, when she awakes on Monday she no longer knows the date; it may be either Monday or Tuesday. If it is Tuesday, however, the coin toss has already happened and certainly had the result tails. In other words, if Beauty knew it is Tuesday, she would possess *inadmissible information* and would not be allowed to use the Principal Principle; she should assign probability 1 to T. So her going

to sleep and the resulting loss of information about the day it is do involve a significant change in her knowledge situation. Because it may be Tuesday she is no longer allowed to use the Principal Principle to determine the odds of T and H in her new situation; she has to factor in her uncertainty about the date. Taking into account this uncertainty in the way described in section 3, she finds 1/3 for her probability of H and 2/3 for her probability of T. When she subsequently learns that it is in fact Monday, and updates her credences using that information, she returns to the probabilities of H and T she entertained on Sunday, before the experiment started.

On Sunday Beauty is already sure that she will be going to assign new probabilities to H and T upon awakening on Monday. She also sees very well that the situation on Monday will be drastically different from her present one, in which she knows the date. Although she knows that she will inevitably end up in that future situation, there is consequently no reason for her to adopt the new probabilities already now, on Sunday.

As pointed out by Elga [5], the probabilities 1/3 and 2/3 have counterparts in relative frequencies that are realized in repetitions of the experiment. Of all awakenings, 1/3 will be awakenings that go together with heads, and 2/3 that go together with tails. The probabilities 1/2 that result from updating on the date will also be borne out by empirical data: of all coin tosses after Beauty's predictions on Monday, roughly half will result in heads. An investigation of which betting quotients will not lead to a systematic loss of money, when confronted with the empirical frequencies, not surprisingly also leads to the conclusion that Beauty should take 1/3 and 2/3 as her probabilities upon waking up on Monday [7][3].

6 Conclusion

Both in the Doomsday Argument and the Sleeping Beauty problem one is asked to assign probabilities to two rival scenarios about the history of the

[3]Hitchcock objects to Elga's simple argument for 1/3, based on the relative frequency of Head awakenings, because the two Tails awakenings are not statistically independent. However, this seems irrelevant in the present context: what counts is that in each trial Beauty cannot distinguish between the three possible awakening instances and considers her own awakening a random selection from all possibilities—in analogy with the randomness assumption in the Doomsday case.

world, while being ignorant about one's own place in history. In both cases one is provided with the missing information and uses this to updates one's probabilities. The question is whether the resulting probabilities are different from the usual ones—the probabilities one uses directly, in full awareness of the date. The answer depends, evidently, on the values of the priors. The proponent of the Doomsday Argument maintains that these priors should be equal to our usual ones; similarly, in the context of the Sleeping Beauty problem the "halfer" contends that it should make no difference for her probabilities that Beauty loses information about the date when she falls asleep. Both claims are wrong, I have argued. Whoever thinks that Beauty should believe that heads and tails are equiprobable in a future toss of a fair coin after she has learnt it is Monday, should not be impressed by the Doomsday Argument; who thinks that at that point her correct degree of belief in Heads is 2/3 should also think Doom is near.

Acknowledgement

My thanks go to Luc Bovens for suggesting the analogy between the Doomsday Argument and the Sleeping Beauty problem, and to Andrea Lubberdink for comments on earlier versions of this paper and stimulating discussions.

References

[1] Bostrom, N.: 1999, 'The Doomsday Argument is Alive and Kicking', *Mind* **108**, 539-550.

[2] Bostrom, N.: 2001, 'The Doomsday Argument, Adam & Eve, UN^{++}, and Quantum Joe', *Synthese* **127**, 359-387.

[3] Bostrom, N.: 2002, *Anthropic Bias: Observation Selections Effects in Science and Philosophy*, Routledge, London.

[4] Bostrom, N.: 2003, 'The Doomsday Argument and the Self-Indication Assumption: Reply to Olum', *The Philosophical Quarterly* **53**, 83-91.

[5] Elga, A.: 2000, 'Self-Locating Belief and the Sleeping Beauty Problem', *Analysis* **60**, 143-147.

[6] Dieks, D.: 1992, 'Doomsday—Or: The Dangers of Statistics', *The Philosophical Quarterly* **42**, 78-84.

[7] Hitchcock, C.: 2004, 'Beauty and the Bets', *Synthese* **139**, 405-420.

[8] Leslie, J.: 1989, *Universes*, Routledge, London.

[9] Leslie, J.: 1993, 'Doom and Probabilities', *Mind* **102**, 489-491.

[10] Leslie, J.: 1996, *The End of the World: The Ethics and Science of Human Extinction*, Routledge, London.

[11] Lewis, D.: 1986, 'A Subjectivist's Guide to Objective Chance', in: *Philosophical Papers, Volume II*, Oxford University Press, Oxford, 83-132.

[12] Lewis, D.: 2001, 'Sleeping Beauty: Reply to Elga', *Analysis* **61**, 171-176.

[13] Monton, B.: 2003, 'The Doomsday Argument Without Knowledge of Birth Rank', *The Philosophical Quarterly* **53**, 79-82.

[14] Olum, K.D.: 2002, 'The Doomsday Argument and the Number of Possible Observers', *The Philosophical Quarterly* **52**, 164-184.

[15] Sober, E.: 2003, 'An Empirical Critique of Two Versions of the Doomsday Argument—Gott's Line and Leslie's Wedge', *Synthese* **135**, 415-430.