

Sleeping Beauty and the Problem of World Reduction

preprint

Paul Franceschi
University of Corsica

revised January 2005

p.franceschi@univ-corse.fr
<http://www.univ-corse.fr/~franceschi>

ABSTRACT. I describe in this paper a solution to the Sleeping Beauty problem. I begin with the consensual emerald case and discuss then Bostrom's Incubator Gedanken. I address then the Sleeping Beauty problem. I argue that the root cause of the flaw in the argument for 1/3 is an erroneous assimilation with a repeated experiment. I show that the same type of analysis also applies to Elga's version of the argument for 1/3. Lastly, I show that the core of the Sleeping Beauty problem is related to the problem of world reduction.

1. The emerald case

Consider, to begin with, the following experiment:

Experiment 1: an urn contains p red balls and q green balls. You draw a ball at random from the urn. You try to evaluate the probability of drawing a red or a green ball. Let $P(R)$ and $P(G)$ denote respectively the probability of drawing a red or a green ball. According to reasoning (I), you conclude then that $P(R) = p / (p + q)$ and $P(G) = q / (p + q)$.

Let us turn now to the situation corresponding to the *emerald case*, described by John Leslie (1996, p. 20):

The emerald case: 'At some point in time, three humans would each be given an emerald. Several centuries afterwards, when a completely different set of humans was alive, five thousands humans would again each be given an emerald in the experiment. You have no knowledge, however, of whether your century is the earlier century in which just three people were to be in this situation, or the later century in which five thousand were to be in it'. What is then your credence that your emerald originates from the set of three humans?

Let us identify now the first set of three emeralds with red balls and the second set of five thousands emeralds with green balls. The situation is now equivalent to an urn that contains three red balls and five thousands green balls. At this stage, it should be clear that the Emerald case is structurally analogous to experiment 1, with $p = 3$ and $q = 5000$. We get then accordingly: $P(R) = 3 / (3 + 5000) = 3/5003$ and $P(G) = 5000 / (3 + 5000) = 5000/5003$. The resulting probability that your emerald comes out from the set of three humans equals $3/5003$; and the probability that it originates from the set of five thousand humans equals $5000/5003$.

I take it that the above reasoning should be consensual. However at this step, agreement stops. Let us then proceed a bit further by reviewing some other experiments and situations.

2. The Incubator

Let us consider now the following experiment:

Experiment 2: The content of an urn depends on the flipping of a fair coin. If Heads, the urn contains one red ball; if Tails, it contains one red ball and one green ball. You try to evaluate the probability of drawing a red or a green ball. A first line of reasoning (I) goes as follows. Consider, to begin with, the probability of drawing a red ball. If the coin has landed Heads then the probability of drawing a red ball is 1. Now if the coin has landed Tails then this latter probability equals $1/2$. The probability of Heads and Tails being $1/2$, we get accordingly: $P(R) = 1 \times 1/2 + 1/2 \times 1/2 = 3/4$. It is worth mentioning in passing that in the Tails case, the situation is in all respects analogous to experiment 1 with an urn that contains one red ball and one green ball, except that the probability of the corresponding situation is $1/2$. Let us turn now to the probability of drawing a green ball. If the coin has landed Heads then this latter probability equals 0. By contrast, if the coin has landed Tails then the probability of drawing a green ball is $1/2$. Hence $P(G) = 0 \times 1/2 + 1/2 \times 1/2 = 1/4$. It is worth noting that the Tails case is analogous to experiment 1 with an urn that contains one red ball and one green ball, except that the probability of the corresponding situation is $1/2$. To sum up, according to reasoning (I): $P(R) = 3/4$ and $P(G) = 1/4$.

However, an alternative reasoning (II) goes as follows. Let us term *iterated experiment 2*, experiment 2 repeated n times. If experiment 2 is repeated n times, say 1000 times, then there will be in total 1000 red balls ($1 \times 1000 \times 1/2 + 1 \times 1000 \times 1/2$) and 500 green balls ($0 \times 1000 \times 1/2 + 1 \times 1000 \times 1/2$). According to reasoning (II) this experiment is equivalent, in the long run, to a type 1 experiment with an urn that contains 1500 balls from whose 1000 red balls and 500 green balls. Hence $P(R) = 1000/1500 = 2/3$ and $P(G) = 500/1500 = 1/3$.

I shall argue that reasoning (II) in experiment 2 is fallacious. Reasoning (II) rests on the fact that experiment 2 can be repeated and the corresponding situation is then analogous to a type 1 experiment with 1500 balls from whose 1000 red and 500 green balls. In what follows, my concern will be with showing that *iterated experiment 2* is not structurally analogous to experiment 1.

For the sake of clarity, let us draw first a distinction between red-Heads (red balls created after the coin has landed Heads), red-Tails (red balls created in the Tails case) and green-Tails (green balls created after the coin has landed Tails) balls. In this context, it should be clear that there only exists red-Heads, red-Tails and green-Tails balls in experiment 2.

The intuition underlying reasoning (II) in experiment 2 is that one is entitled to add red and green balls to compute frequencies. However, I shall argue that this intuition is misleading. With our terminology, it means that one feels intuitively entitled to add red-Heads, red-Tails and green-Tails balls to compute frequencies. Let us begin with red-Heads balls. In the current context, red-Heads balls can be considered properly as single objects. You are then entitled to envisage drawing isolately red-Heads balls and these latter can acceptably be seen as single objects. By contrast, it appears that red-Tails balls are quite undissociable from green-Tails balls. For you cannot draw a red-Tails ball without drawing the associated green-Tails ball. And conversely, you cannot pick a green-Tails ball without picking the associated red-Tails ball. From this viewpoint, it is mistaken to consider red-Tails and green-Tails balls as separate objects. The correct intuition is that the association of a red-Tails and a green-Tails ball constitute one single object, in the same sense as red-Heads balls constitute single objects. And red-Tails and green-Tails balls are best seen intuitively as constituents and mere parts of one single object. In other words, red-Heads balls and, on the other hand, red-Tails and green-Tails balls, cannot be considered as objects of the same type for frequency probability purposes. And this situation justifies the fact that one is not entitled to add red-Heads, red-Tails and green-Tails balls to compute probability frequencies. For in both cases, you add objects of intrinsically different types, i.e. you add one single object with the mere part of another single object. The correct intuition, however, is that red-Heads balls can be seen as single objects, while on the other hand, red-Tails and green-Tails balls must be considered properly as mere parts of single objects which are on a par with red-Heads objects. Now the analogy with experiment 1 proves to be ungrounded, since in this latter case, red and green balls can legitimately be put on a par and considered as objects of the same type. Now this invalidates the analogy of iterated experiment 2 with experiment 1. It follows that reasoning (II) in experiment 2 is incorrect. This leaves us with reasoning (I). As we have seen, the whole idea of

reasoning as if experiment 2 were repeated is related to the frequentist interpretation of probabilities (Hájek 2002). The upshot, however, is that this latter interpretation of probabilities should not be adopted unrestricted. In particular, frequencies should not be computed by adding objects of intrinsically different types.

At this step, it is worth considering the situation corresponding to the *Incubator* (Bostrom 2002, p. 64):¹

The Incubator: ‘Stage (a) In an otherwise empty world, a machine called ‘the incubator’ kicks into action. It starts by tossing a fair coin. If the coin falls [heads] then it creates one room and a man with a black beard inside it. If the coin falls [tails] then it creates two rooms, one with a black-bearded man and one with a white-bearded man. As the rooms are completely dark, nobody knows his beard color. Everybody who's been created is informed about all of the above. You find yourself inside one of the rooms.’ Question: What should be your credence that you have a black or a white beard?²

Now it appears that the line of reasoning related to experiment 2 can be applied straightforwardly to the *Incubator*. It suffices to identify a black-bearded man with a red ball and a white-bearded man with a green ball. The resulting situation is a machine that creates one red ball on Heads and one red ball and one green ball on Tails. This has the effect of rendering the situation corresponding to the *Incubator* analogous to experiment 2. Now given that reasoning (I) applies to experiment 2, it follows that reasoning (I) also applies to the *Incubator*.

3. The Sleeping Beauty problem: a solution

Let us envisage now the following experiment:

Experiment 3: an urn contains one red ball and one green ball. If Heads then due to a filtering effect, you cannot see nor feel green balls and you can only see and feel one red ball. If Tails then there is no filter effect and you can see and feel one red ball and one green ball. Your task is to evaluate the probability of drawing a red or a green ball.

At this stage, it appears that experiment 3 is in all respects similar to experiment 2, except for what concerns the Heads case. In this latter case, in experiment 2 there is only one red ball in the urn, which is devoid of green balls. By contrast, in experiment 3, there is one red ball and one green ball in the urn, but you cannot see nor feel the green ball, due to a selection effect (Leslie 1989, Bostrom 2002). But the outcome is that this precludes your from drawing the green ball in the Heads case, thus rendering the situation equivalent—from a probability standpoint—to experiment 2. As a consequence, reasoning (I) also applies to experiment 3.

At this step, it is worth considering the *Sleeping Beauty problem* (Elga 2000, p. 143):

The Sleeping Beauty problem: ‘Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you back to sleep with a drug that makes you forget that waking’. Once awakened, what should Sleeping Beauty's credence be that (i) it is a Monday waking; and (ii) the coin has landed Heads?³

¹ The Heads and Tails cases are reverted here, with regard to Bostrom's original description. The extensive version of the incubator also includes a later stage: ‘Stage (b): a little later, the lights are switched on, and you discover that you have a black beard. Question: What should your credence in Tails be now?’.

² Bostrom's original question: ‘What should be your credence that the coin fell tails?’.

³ Adapted from Elga (2000). Elga's original text: ‘When you are *first* [my emphasis] awakened, to what degree ought you believe that the outcome of the coin toss is Heads?’. Considering here *any* waking (Heads-waking on Monday, Tails-waking on Monday or Tails-waking on Tuesday) is more general and equally allowed by the formulation of the problem, since all wakings are indistinguishable.

'First answer: 1/2, of course! Initially you were certain that the coin was fair, and so initially your credence in the coin's landing Heads was 1/2. Upon being awakened, you receive no new information (...). So your credence in the coin's landing Heads ought to remain 1/2. *Second answer:* 1/3, of course! Imagine the experiment repeated many times. Then in the long run, about 1/3 of the wakings would be Heads-wakings (...). So on any particular waking, you should have credence 1/3 that that waking is Heads-waking, and hence have credence 1/3 in the coin's landing Heads on that trial'.

I shall argue here that the situation corresponding to the *Sleeping Beauty problem* is structurally analogous to experiment 3. In the Sleeping Beauty experiment, the time variable includes two temporal locations: Monday and Tuesday. Moreover, it appears that Beauty faces a selection effect in the case where the coin lands Heads, for in this latter case, Beauty is not awakened on Tuesday. By contrast, Beauty faces no selection effect in the Tails case, since she is awakened on both Monday and Tuesday. In short, in the Heads case, Sleeping Beauty perceives the first time location (Monday) but is unable to perceive the second temporal location (Tuesday). However, in the Tails case, she is able to perceive both time locations (Monday and Tuesday). Let us identify now Monday with a red ball and Tuesday with a green ball. Now a Monday-waking can be assimilated with drawing a red ball and a Tuesday-waking can be identified to drawing a green ball. Furthermore, being not awoken on Tuesday in the Heads case can be assimilated with being incapable of seeing nor feeling the green ball due to a filtering effect. At this stage, it should be clear that the situation corresponding to the Sleeping Beauty problem parallels the urn analogy from experiment 3. It follows that reasoning (I) also applies to the situation corresponding to the Sleeping Beauty problem.

At this stage, we are in a position to diagnose precisely the flaw in the argument for 1/3 in the Sleeping Beauty problem. For this purpose, it is worth stating more accurately the argument for 1/3. It begins informally with the transposition of reasoning (II) in experiment 3. The argument for 1/3 rests crucially on the fact that if the Sleeping Beauty experiment is repeated, it can be assimilated to a type 1 experiment. The corresponding line of reasoning runs as follows: if the Sleeping Beauty experiment is repeated n times, say 1000 times, then there will be in total 1000 ($1 \times 1000 \times 1/2 + 1 \times 1000 \times 1/2$) Monday-wakings and 500 ($0 \times 1000 \times 1/2 + 1 \times 1000 \times 1/2$) Tuesday-wakings. This experiment is equivalent in the long run, the argument goes, to a type 1 experiment with an urn that contains 1500 balls from whose 1000 red and 500 green balls. The respective probabilities of a Monday-waking and of a Tuesday-waking are computed accordingly: $P(\text{Monday-waking}) = 1000/1500 = 2/3$ and $P(\text{Tuesday-waking}) = 500/1500 = 1/3$.

At this point, it is worth mentioning some additional steps which are specific to the argument for 1/3 and which are targeted at computing $P(\text{Heads})$ and $P(\text{Tails})$. Given that the total number of Heads-wakings on Monday and of Tails-waking on Monday amounts respectively to $1 \times 1000 \times 1/2 = 500$, it follows that the probability of a Heads-waking on Monday and of a Tails-waking on Monday equals $500/1500 = 1/3$. Now given that the probability of Heads equals the probability of a Heads-waking on Monday, it follows that $P(\text{Heads}) = P(\text{Heads-waking on Monday}) = 1/3$. In parallel, the probability of Tails equals the probability of a Tails-waking on Monday plus the probability of a Tails-waking on Tuesday. Hence, $P(\text{Tails}) = P(\text{Tails-waking on Monday}) + P(\text{Tuesday-waking}) = 1/3 + 1/3 = 2/3$.

Now the flaw in the thirder's line of reasoning can be accurately diagnosed. The erroneous step is the *analogy step*, namely the consideration that if the experiment is repeated n times, it will be equivalent to a type 1 experiment. And the diagnosis of the fallacy in the argument for 1/3 now parallels the flaw in reasoning (II) in experiments 2 and 3. What is at the origin of the problem is the misleading intuition that each waking is intuitively considered as one single event. And the apparent plausibility of the argument for 1/3 emerges from the fact that one feels pre-theoretically entitled to add Monday wakings and Tuesday wakings to compute frequencies. However, as underlined above, one must draw first a distinction between Monday-Heads, Monday-Tails and Tuesday-Tails wakings. It follows then that Monday-Heads wakings and, on the other hand, Monday-Tails and Tuesday-Tails wakings cannot be considered properly as objects as the same type. For Monday-Tails wakings are undissociable from Tuesday-Tails wakings. And this finally prohibits adding (i) Monday-Heads and Monday-Tails wakings and (ii) Monday-Heads and Tuesday-Tails wakings to compute frequencies. This renders reasoning (II) fallacious and finally does justice to reasoning (I).

4. Elga's variation of the Argument for 1/3

Although Elga's initial motivation of the argument for 1/3 bears on a repeated experiment,⁴ he further presents his own version of the argument in a somewhat more sophisticated way, which doesn't make use of the repeated experiment. Nevertheless, I shall argue that the preceding line of reasoning applies all the same to Elga's variation of the argument for 1/3. Let us scrutinize then Elga's two-stage version of the argument. Elga argues first, in virtue of a very restricted principle of indifference, that the probability of a Tails-waking on Monday equals the probability of a Tails-waking on Tuesday. Lewis (2001) agrees with that and the present account also accepts this conclusion, given that Monday-Tails and Tuesday-Tails wakings can be considered properly as objects of the same type. To put it equivalently in terms of centered worlds (Quine, *Propositional objects* 1969, Lewis 1983): a Monday-Tails centered world can legitimately be considered as analogous to a Tuesday-Tails centered world, thus providing an adequate model for applying a restricted principle of indifference.

Elga proceeds then to demonstrate, second, that the probability of a Heads-waking on Monday equals the probability of a Tails-waking on Monday. He argues as follows:⁵

Now: if (upon awakening) you were to learn that it is Monday, that would amount to your learning that you are in either H_1 or T_1 . Your credence that you are in H_1 would then be your credence that a fair coin, soon to be tossed, will land Heads. It is irrelevant that you will be awakened on the following day if and only if the coin lands Tails—in this circumstance, your credence that the coin will land Heads ought to be 1/2. But your credence that the coin will land Heads (after learning that it is Monday) ought to be the same as the conditional credence $P(H_1|H_1 \text{ or } T_1)$. So $P(H_1|H_1 \text{ or } T_1) = 1/2$, and hence $P(H_1) = P(T_1)$.

This can be stated step-by-step as follows:

- (E1) $P(\text{Monday-Heads} \mid \text{Monday-Heads or Monday-Tails}) = P(\text{Heads})$
- (E2) $P(\text{Heads}) = 1/2$
- (E3) $P(\text{Monday-Heads} \mid \text{Monday-Heads or Monday-Tails}) = 1/2$ from (E1),(E11)
- (E4) $\therefore P(\text{Monday-Heads}) = P(\text{Monday-Tails})$ from (E12)

It follows that the probability of a Heads-waking on Monday equals the probability of a Tails-waking on Monday. Combining then the two preceding stages, Elga observes that $P(\text{Monday-Heads}) = P(\text{Monday-Tails}) = P(\text{Tuesday-Tails})$. Given that these probabilities sum to 1, he further concludes that $P(\text{Monday-Heads}) = 1/3$.

Now the above analysis can be applied straightforwardly to Elga's second stage of the argument.⁶ In effect, from the present standpoint, Heads-waking on Monday and Tails-waking on Monday cannot be considered legitimately as objects of the same type. For a Tails-waking on Monday is undissociable from a Tails-waking on Tuesday, whereas a Tails-waking on Monday lacks such specific feature. To put it into the framework of centered worlds: a Monday-Heads centered world cannot be considered properly as analogous to a Monday-Tails centered world. For the Monday-Tails centered world has a distinctive feature from the Monday-Heads centered world, namely the fact that the Monday-Tails centered world is undissociable from the Tuesday-Tails centered world. And this precludes us from regarding the Monday-Heads and the Monday-Tails centered worlds as analogous, thus blocking the application of the principal principle to the relevant situation.

⁴ Cf. Elga 2000, p. 143: 'Imagine the experiment repeated many times. Then in the long run, about 1/3 of the wakings would be Heads-wakings (...)'.

⁵ Cf. Elga 2000, p. 145. In Elga's notation, H_1 = Heads-waking on Monday, T_1 = Tails-waking on Monday, T_2 = Tails-waking on Tuesday. Elga considers here a variation of the experiment where the coin is tossed *after* Beauty's waking on Monday. But the same applies equivalently if the coin is tossed *before*.

⁶ I don't address here the fact that step (E11) assumes what is at issue in the Sleeping Beauty problem, namely the computation of $P(\text{Heads})$. But this would be a red herring, I think. In effect, I take it that the core of the Sleeping Beauty problem could be reduced to the correct evaluation of the probability of a Heads-waking on Monday, of a Tails-waking on Monday and of a Tails-waking on Tuesday. And from this viewpoint, Elga's argument is accurately motivated.

5. The problem of world reduction

What precedes casts light on the central topic which is at issue here. The problem is related to the *reduction* of a world to another, a problem notably hinted at by Quine (1969)⁷ and Goodman (1978, p. 99-102). To put it otherwise: how can the objects of a given world be transferred properly into another structurally different world? As we have seen, the flaw in the argument for 1/3 resides in the fact that objects of different types (to say it otherwise, of different worlds) are added in order to compute probability frequencies. To put it equivalently, the fallacy arises from the fact that the principal principle cannot be applied properly to dissimilar centered worlds. However, this is only the negative aspect of the issue, which appears to be two-faceted. For the positive aspect of the issue consists of settling how to render two *prima facie* dissimilar centered worlds compatible, in order to properly compute probability frequencies. To solve this reduction problem, we need to clarify the issue of how to transfer the objects of a given centered world into another, while preserving their internal structure.

In the present context, two different instances of the problem emerge: (i) how can the objects of the Heads-world be transferred into the Tails-world, without loss of content? (ii) how can the objects of the Tails-world be transferred into the Heads-world? Before addressing these issues, let us pause a moment to point out the differences between the Heads-world and the Tails-world. Consider then the objects of the Incubator. Recall that the Heads-world of the Incubator contains one red ball, while on the other hand, the Tails-world contains one composite object made up of two undissociable parts which consist of a red ball and a green ball (a RG object).



Figure 1. *The Incubator in the Heads world*

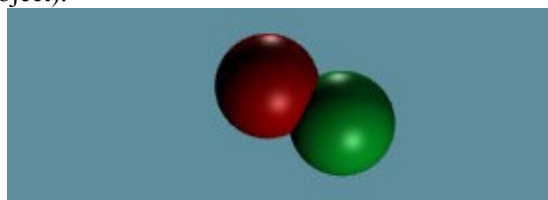


Figure 2. *The Incubator in the Tails world*

The Heads-world and the Tails-world contain respectively only one single object. For present purposes, what we need is a slight extension of these two worlds which are capable of accepting *several* objects of the same type. Call them respectively Heads-world* and Tails-world*. On the one hand, the Heads-world which contains one red ball seems very similar to our familiar world. Now what we need is an extension of the Heads-world that is capable of accepting several objects like the red ball. A Heads-world* which can contain several red balls (and also eventually some balls of a different color, say green) will do the job. To sum up now: the Heads-world* is capable of including several red or green balls.

On the other hand, the Tails-world contains one RG object. *Prima facie*, the objects of the Tails-world are no less different than the objects of the Heads-world. But this intuition is mistaken, as a more in depth analysis will reveal. For we should bear in mind that the RG object of the Tails-world is made up of two undissociable parts. And the point is that there is no way to separate the red ball from the green one. No force in the Tails-world is capable of dissociating the two balls. Thus, you cannot pick a red (resp. green) ball without picking the associated green (resp. red) ball. Such dispositional property is very different from the usual properties of our familiar world and from the Heads-world*, where two balls of different colors are fully separated. Now what we need is an extension of the Tails-world that is capable of accepting multiple objects of the RG sort. A Tails-world* which can contain several objects made up of two undissociable balls (red or green) will do the job. To sum up: the Tails-world* is capable of containing several RG,⁸ RR or GG objects.

⁷ Cf. Quine (1969, *Ontological Relativity*, p. 55): 'A usual occasion for ontological talk is reduction, where it is shown how the universe of some theory can by a reinterpretation be dispensed with in favor of some other universe, perhaps a proper part of the first'.

⁸ For present purposes, there is no need to differentiate RG from GR objects.

At this step, we are in a position to handle (i) the first instance of the above reduction problem: how can the objects of the Heads-world be transferred into the Tails-world? From the above, it results that the problem should be more accurately termed: how can the objects of the Heads-world be transferred properly into the Tails-world*? Recall that the Heads-world of the Incubator originally contains one red ball. What becomes this red ball once transferred into the Tails-world*, where there only exists composite objects? Given the above, the answer is that the red ball is transformed into a RR object, i.e. a composite object made up of two undissociable red balls, once plugged into the Tails-world*.

We are now in a position to handle a related issue: what is the equivalent of the Incubator in the Tails-world*, i.e. what is the equivalent of the objects of the Heads-world *plus* the objects of the Tails-world, once transferred into the Tails-world*? Given the above, in total, the objects of the Incubator (i.e. the red ball of the Heads-world and the RG object of the Tails-world) are replaced by one RR and one RG object in the Tails-world*. We observe then that the relative proportions of the balls (1:2), i.e. one ball (Heads-world) for one 2-composite RG object (Tails-world) have been preserved, since we have exchanged one ball (Heads-world) for one composite RR object (Tails-world*).

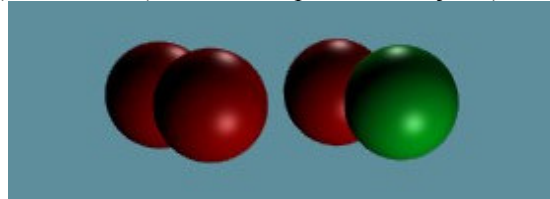


Figure 3. *The Incubator in the Tails* world*

Let us turn now (ii) to the second instance of our transference problem: how can the objects of the Tails-world be transferred properly into the Heads-world? From the above, it results that the problem should be more accurately expressed: how can the objects of the Tails-world be transferred into the Heads-world*? Recall that the Tails-world contains one RG object. What becomes this RG object once transferred into the Heads-world*, where there only exists red or green balls? Given the above, the answer is that this RG object is transformed into two balls, respectively red and green, once plugged into the Tails-world*.

Let us tackle now the inherently related issue of the Incubator's equivalent in the Heads-world*: what is the equivalent of the objects of the Heads-world *plus* the objects of the Tails-world, once transferred into the Heads-world*? Now in order to preserve the relative proportions (1:2) of the balls, i.e. one ball (Heads-world) for one 2-composite RG object (Tails-world), a slight adaptation is required. Recall now that the Heads-world initially contains one red ball. Now the red ball of the Heads-world needs to be transformed into two red balls in the Heads-world*, in order to preserve the relative proportions. Now merging the objects of the Heads-world and the Tails-world into a Heads-world* while preserving their relative proportions, we get: two red balls plus one red ball and one green ball, i.e. three red balls and one green ball. We observe then that the original red ball has been transformed into two red balls, and the incoming composite RG object has been split into one red ball and one green ball.

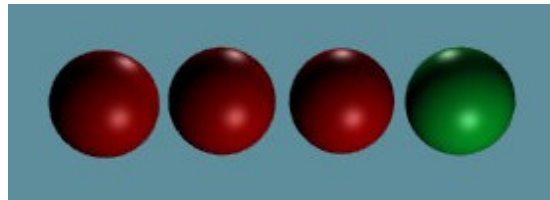


Figure 4. *The Incubator in the Heads* world*

To fix ideas, let us envisage now a slight variant of the Incubator. Suppose now that if the coin falls Heads, the Incubator creates three rooms: one with a black-bearded man, one with a white-bearded man and one with a gray-bearded man. Call Incubator* this variation. Recall that in the original experiment, the Incubator only creates one room with one black-bearded man on Heads. The situation can now be modeled with a slight variation of experiment 2, namely where the urn contains one red

ball, one green ball and one black ball on Heads (instead of only one red ball). Now the Heads-world and the Tails-world of the Incubator* can be described as follows:

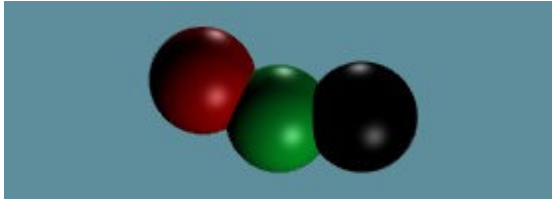


Figure 5. *The Incubator* in the Heads world*

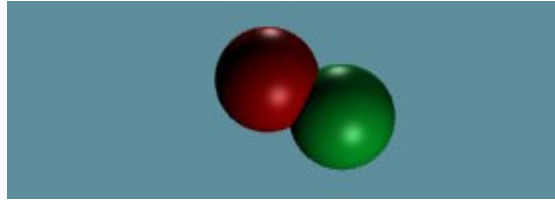


Figure 6. *The Incubator* in the Tails world*

What is then the equivalent of the Incubator* in the Tails-world* (an extension of the Tails-world which can contain multiple 2-composite objects of the RG sort)? Now reasoning as above, the RGB object of the Heads-world is transformed into one RR, one GG and one BB object once plugged into the Tails-world*. The initial proportion of the balls being 3:2, the original RG object of the Tails-world is transformed into two RG objects once plugged into the Tails-world*, in order to preserve the initial proportions. What we get finally is one RR object, one GG object, one BB object and two RG objects:

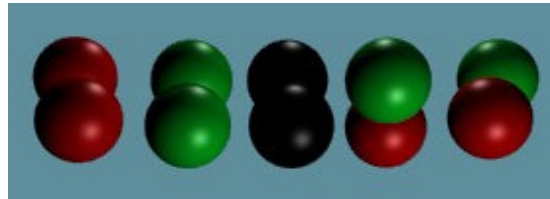


Figure 7. *The Incubator in the Tails* world*

What is then the equivalent of the Incubator* in the Heads-world* (an extension of the Heads-world which can contain multiple 3-composite objects of the RGB sort)? Reasoning as above, the RG object of the Tails-world is transformed into one RRR and one GGG object once plugged into the Heads-world*. The initial proportion of the balls being 3:2, the original RGB object of the Heads-world is transformed into three RGB objects once plugged into the Heads-world*, in order to preserve the initial proportions. What we get finally is one RRR object, one GGG object and three RGB objects:

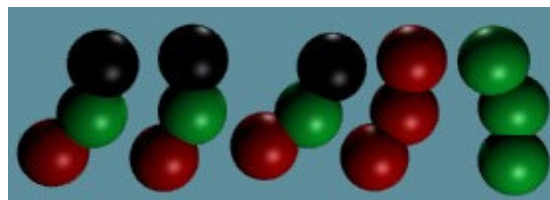


Figure 8. *The Incubator in the Heads* world*

Finally, the lesson of the Sleeping Beauty Problem is the following: our current and familiar objects or concepts such as balls, wakings, etc. should not be considered as the sole relevant classes of objects for probability purposes. We should bear in mind that according to an unformalized axiom of probability theory, a given situation is standardly modeled with the help of urns, dices, balls, etc. But the rules that allow for these simplifications lack an explicit formulation. However in certain situations, in order to reason properly, it is also necessary to take into account somewhat unfamiliar objects whose constituents are pairs of indissociable balls or of mutually unseparable wakings, etc. This lesson was anticipated by Nelson Goodman, who pointed out in *Ways of Worldmaking* that some objects which are prima facie completely different from our familiar objects also deserve consideration: ‘we do not welcome molecules or concreta as elements of our everyday world, or combine tomatoes and triangles and typewriters and tyrants and tornadoes into a single kind’.⁹ As we

⁹ Goodman (1978, p. 21).

have seen, we cannot add unrestrictedly objects of the Heads-world with objects of the Tails-world. For some objects of the Heads-world are mere parts of objects in the Tails-world. And reciprocally, some mere constituents of objects of the Tails-world are genuine and independent objects in the Heads-world. Now the status of our paradigm probabilistic object, namely a ball, appears world-relative, since it can be a whole in the Heads-world and a part in the Tails-world. Once this goodmanian step accomplished, we should be less vulnerable to certain subtle cognitive traps in probabilistic reasoning.¹⁰

References

- Arntzenius, F. (2002) Reflections on Sleeping Beauty, *Analysis*, 62-1, 53-62
- Bostrom, N. (2002) *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, New York: Routledge
- Bradley, D. (2003) Sleeping Beauty: a note on Dorr's argument for 1/3, *Analysis*, 63, 266-268
- Delahaye, J.-P. (2003) La Belle au bois dormant, la fin du monde et les extraterrestres, *Pour la Science*, 309, 98-103
- Dorr, C. (2002) Sleeping Beauty: in Defence of Elga, *Analysis*, 62, 292-296
- Elga, A. (2000) Self-locating Belief and the Sleeping Beauty Problem, *Analysis*, 60, 143-147
- Franceschi, P. (2001) Une solution pour le paradoxe de Goodman, *Dialogue*, 40: 99-123, English translation under the title 'A Solution to Goodman's Paradox', Cogprints: 2176
- Franceschi, P. (2005) Situations probabilistes pour n-univers goodmaniens, *Journal of Philosophical Research*, to appear
- Goodman, N. (1978) *Ways of Worldmaking*, Indianapolis: Hackett Publishing Company
- Hájek, A. (2002) Interpretations of Probability, *The Stanford Encyclopedia of Philosophy* (Winter 2002 Edition), E. N. Zalta (ed.), <http://plato.stanford.edu/archives/win2002/entries/probability-interpret>
- Leslie, J. (1989) *Universes*, London: Routledge
- Leslie, J. (1996) *The End of the World: the science and ethics of human extinction*, London: Routledge
- Lewis, D. (1983) Attitudes de dicto and de se, in *Philosophical Papers*, volume I, 133-159, New York: Oxford University Press
- Lewis, D. (2001) Sleeping Beauty: Reply to Elga, *Analysis*, 61, 171-176
- Monton, B. (2002) Sleeping Beauty and the Forgetful Bayesian, *Analysis*, 62, 47-53
- Quine, W.V.O. (1969) Ontological Relativity, in *Ontological Relativity and Other Essays*, Columbia University Press: New York.
- Quine, W.V.O. (1969) Propositional objects, in *Ontological Relativity and Other Essays*, Columbia University Press: New York.

¹⁰ I thank Jean-Paul Delahaye and Claude Panaccio for useful discussion.