The Case of Sleeping Beauty

Tyler Seacrest

April 29, 2009
Bayes’ Rule Example
Bayes’ Rule Example
Bayes’ Rule Example

50%

50%

50%
Bayes’ Rule Example
Bayes’ Rule Example

Question: What is the probability a black ball was put in?
Bayes’ Rule Example
Bayes’ Rule Example
Bayes’ Rule Example
Bayes’ Rule Example

Case 1
Bayes’ Rule Example

Case 1

(Probability a black ball was initially put in is 100%)
Bayes’ Rule Example
Bayes’ Rule Example

Case 2
Bayes’ Rule Example

Case 2
(Probability a black ball was put in goes down)
Bayes’ Rule Example

Case 2

Possible contents of the urn
Case 2

Possible contents of the urn

- One of these three balls were chosen, and each case is equally likely
Bayes’ Rule Example

Case 2

Possible contents of the urn

- One of these three balls were choosen, and each case is equally likely.
- Hence, the probability a black ball was put in initially drops to 1/3.
The Case of Sleeping Beauty

A subject known as SB takes part in an experiment at the ACME probability labs. She has full knowledge of all details of the experiment.

Sunday night, SB goes to bed. While she is asleep, a coin is flipped.

She is awoken Monday morning, and asked, "From your point of view, what is the probability the coin came up heads?". If the coin actually did come up heads, this is the end of the experiment.

If the coin came up tails, SB is injected with a special drug cocktail that will put her back to sleep and make her completely forget Monday's events.

She is then awoken on Tuesday morning (which, to her, seems like Monday morning) and asked "From your point of view, what is the probability the coin came up heads?"

If you were in SB's shoes, how would you answer?
The Case of Sleeping Beauty

- A subject known as SB takes part in an experiment at the ACME probability labs. She has full knowledge of all details of the experiment.

- Sunday night, SB goes to bed. While she is asleep, a coin is flipped.

- She is awoken Monday morning, and asked, "From your point of view, what is the probability the coin came up heads?" If the coin actually did come up heads, this is the end of the experiment.

- If the coin came up tails, SB is injected with a special drug cocktail that will put her back to sleep and make her completely forget Monday's events.

- She is then awoken on Tuesday morning (which, to her, seems like Monday morning) and asked "From your point of view, what is the probability the coin came up heads?"

- If you were in SB's shoes, how would you answer?
The Case of Sleeping Beauty

- A subject known as SB takes part in an experiment at the ACME probability labs. She has full knowledge of all details of the experiment.
- Sunday night, SB goes to bed. While she is asleep, a coin is flipped.
The Case of Sleeping Beauty

- A subject known as SB takes part in an experiment at the ACME probability labs. She has full knowledge of all details of the experiment.
- Sunday night, SB goes to bed. While she is asleep, a coin is flipped.
- She is awoken Monday morning, and asked, “From your point of view, what is the probability the coin came up heads?” If the coin actually did come up heads, this is the end of the experiment.
- If the coin came up tails, SB is injected with a special drug cocktail that will put her back to sleep and make her completely forget Monday’s events.
- She is awoken on Tuesday morning (which, to her, seems like Monday morning) and asked “From your point of view, what is the probability the coin came up heads?” If the coin actually did come up heads, this is the end of the experiment.
The Case of Sleeping Beauty

▶ A subject known as SB takes part in an experiment at the ACME probability labs. She has full knowledge of all details of the experiment.

▶ Sunday night, SB goes to bed. While she is asleep, a coin is flipped.

▶ She is awoken Monday morning, and asked, “From your point of view, what is the probability the coin came up heads?” If the coin actually did come up heads, this is the end of the experiment.

▶ If the coin came up tails, SB is injected with a special drug cocktail that will put her back to sleep and make her completely forget Monday’s events.

▶ If you were in SB’s shoes, how would you answer?
The Case of Sleeping Beauty

- A subject known as SB takes part in an experiment at the ACME probability labs. She has full knowledge of all details of the experiment.
- Sunday night, SB goes to bed. While she is asleep, a coin is flipped.
- She is awoken Monday morning, and asked, “From your point of view, what is the probability the coin came up heads?” If the coin actually did come up heads, this is the end of the experiment.
- If the coin came up tails, SB is injected with a special drug cocktail that will put her back to sleep and make her completely forget Monday’s events.
- She is then awoken on Tuesday morning (which, to her, seems like Monday morning) and asked “From your point of view, what is the probability the coin came up heads?”

If you were in SB’s shoes, how would you answer?
The Case of Sleeping Beauty

- A subject known as SB takes part in an experiment at the ACME probability labs. She has full knowledge of all details of the experiment.
- Sunday night, SB goes to bed. While she is asleep, a coin is flipped.
- She is awoken Monday morning, and asked, “From your point of view, what is the probability the coin came up heads?” If the coin actually did come up heads, this is the end of the experiment.
- If the coin came up tails, SB is injected with a special drug cocktail that will put her back to sleep and make her completely forget Monday’s events.
- She is then awoken on Tuesday morning (which, to her, seems like Monday morning) and asked “From your point of view, what is the probability the coin came up heads?”
- If you were in SB’s shoes, how would you answer?
The Pictoral Version
The Pictoral Version
The Pictoral Version

Monday:

[Two alarm clocks with faces representing morning and afternoon activities]
The Pictoral Version

Monday:

Tuesday: X
Question: From SB's point of view, what is the probability the coin came up heads?
Monday:  
Tuesday:  

Question: From SB's point of view, what is the probability the coin came up heads?

Each of these three possible awakenings seem identical to SB.
The Pictoral Version

Monday:

Tuesday:

Question: From SB's point of view, what is the probability the coin came up heads?

- Each of these three possible awakenings seem identical to SB.
- Ideas?
The Argument of David Lewis, the Halfer

Let $p$ be SB's current estimation of the probability the coin is heads. Certainly, Sunday night $p = 1/2$.

(⋆) In order for $p$ to change, SB must gain new, relevant information.

SB knows she will be awakened, hence by being woken up she gains no new information.

Hence, during the experiment $p = 1/2$.
Let $p$ be SB’s current estimation of the probability the coin is heads. Certainly, Sunday night $p = 1/2$. 
Let $p$ be SB’s current estimation of the probability the coin is heads. Certainly, Sunday night $p = 1/2$.

(*) In order for $p$ to change, SB must gain new, relevant information.
The Argument of David Lewis, the Halfer

Let \( p \) be SB’s current estimation of the probability the coin is heads. Certainly, Sunday night \( p = 1/2 \).

\( \star \) In order for \( p \) to change, SB must gain new, relevant information.

SB knows she will be awakened, hence by being woken up she gains no new information.
The Argument of David Lewis, the Halfer

Monday: ⌚️ ⌚️
Tuesday: ⌚️

Let $p$ be SB’s current estimation of the probability the coin is heads. Certainly, Sunday night $p = 1/2$.

(*) In order for $p$ to change, SB must gain new, relevant information.

SB knows she will be awakened, hence by being woken up she gains no new information.

Hence, during the experiment $p = 1/2$. 
The Standard Argument of a Thirder

Suppose the experiment is repeated 20 times. On average, we'd expect roughly 10 awakenings to be with the coin heads, and 20 awakenings to be with the coin tails, for a total of 30 awakenings. SB has no reason to think a given awakening is more likely than another, hence $p = \frac{10}{30} = \frac{1}{3}$. 

Monday: 🕒_heads, 🕒_heads

Tuesday: 🕒_heads, 🕒_tails
The Standard Argument of a Thirder

▶ Suppose the experiment is repeated 20 times.

Monday: 🕒 🕒

Tuesday: ❌ 🕒
Suppose the experiment is repeated 20 times.

On average, we’d expect roughly 10 awakenings to be with the coin heads, and 20 awakenings to be with the coin tails, for a total of 30 awakenings.
The Standard Argument of a Thirder

▶ Suppose the experiment is repeated 20 times.
▶ On average, we’d expect roughly 10 awakenings to be with the coin heads, and 20 awakenings to be with the coin tails, for a total of 30 awakenings.
▶ SB has no reason to think a given awakening is more likely than another, hence $p = 10/30 = 1/3$. 
Another Halfer argument

According to a thirder, $p = \frac{1}{2}$ on Sunday, and $p = \frac{1}{3}$ on Monday.

However, SB not only gained no new information, there was no funny business with forgetfulness drugs at that point. How could a rational person with no new information and with no cognitive lapses change her probability estimation?
Another Halfer argument

According to a thirder, \( p = 1/2 \) on Sunday, and \( p = 1/3 \) on Monday.
Another Halfer argument

According to a thirder, $p = 1/2$ on Sunday, and $p = 1/3$ on Monday.

However, SB not only gained no new information, there was no funny business with forgetfulness drugs at that point. How could a rational person with no new information and with no cognitive lapses change her probability estimation?
Suppose instead the coin flip happened Monday night, after SB had awakened once. (This should have no substantive effect on the experiment.)

If SB is told it is Monday, then $p = \frac{1}{2}$, since the coin flip is a future event. Let's use Bayes' rule to figure out the consequences of this assertion.
Suppose instead the coin flip happened Monday night, after SB had awakened once. (This should have no substantive effect on the experiment.)
The Argument of Adam Elga, a thirder

- Suppose instead the coin flip happened Monday night, after SB had awakened once. (This should have no substantive effect on the experiment.)

\[
\begin{aligned}
\text{Monday:} & \quad \text{\begin{itemize}
\item Alarm clock
\item SB awake
\item Coin flipped
\end{itemize}} \\
\text{Tuesday:} & \quad \text{\begin{itemize}
\item Alarm clock
\item SB awake
\item No coin flip
\end{itemize}}
\end{aligned}
\]

- Let's use Bayes' rule to figure out the consequences of this assertion.
Suppose instead the coin flip happened Monday night, after SB had awakened once. (This should have no substantive effect on the experiment.)

If SB is told it is Monday, then \( p = \frac{1}{2} \), since the coin flip is a future event.
Suppose instead the coin flip happened Monday night, after SB had awakened once. (This should have no substantive effect on the experiment.)

If SB is told it is Monday, then $p = 1/2$, since the coin flip is a future event.

Let’s use Bayes’ rule to figure out the consequences of this assertion.
The Argument of Adam Elga, a thirder

Monday:

Tuesday:

Bayes' Rule is

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

Let

\[ A = \text{Coin is Heads} \quad \text{and} \quad B = \text{It is Monday} \]

Bayes' Rule says

\[ P(\text{Heads}|\text{Monday}) = \frac{P(\text{Monday}|\text{Heads})P(\text{Heads})}{P(\text{Monday})} \]
The Argument of Adam Elga, a thirder

Monday:

Tuesday:

Bayes’ Rule is

\[ P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \. \]
Bayes’ Rule is

\[ P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \].

Let \( A = \text{Coin is Heads} \) and \( B = \text{It is Monday} \).
The Argument of Adam Elga, a thirder

Bayes’ Rule is

\[ P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}. \]

Let \( A = \text{Coin is Heads} \) and \( B = \text{It is Monday} \).

Bayes’ Rule says

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads})P(\text{Heads})}{P(\text{Monday})}. \]
The Argument of Adam Elga, a thirder

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads}) P(\text{Heads})}{P(\text{Monday})}. \]
The Argument of Adam Elga, a thirder

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads})P(\text{Heads})}{P(\text{Monday})}. \]

- We already argued that \( P(\text{Heads} \mid \text{Monday}) = 1/2. \)
The Argument of Adam Elga, a thirder

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads})P(\text{Heads})}{P(\text{Monday})}. \]

- We already argued that \( P(\text{Heads} \mid \text{Monday}) = \frac{1}{2}. \)
- Clearly \( P(\text{Monday} \mid \text{Heads}) = 1. \)
The Argument of Adam Elga, a thrider

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads}) P(\text{Heads})}{P(\text{Monday})}. \]

- We already argued that \( P(\text{Heads} \mid \text{Monday}) = \frac{1}{2} \).
- Clearly \( P(\text{Monday} \mid \text{Heads}) = 1 \).
- I’m not going to claim exactly what \( P(\text{Monday}) \) is, but certainly it is less than 1 (i.e. it could be Tuesday).
The Argument of Adam Elga, a thirder

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads})P(\text{Heads})}{P(\text{Monday})}. \]

- We already argued that \( P(\text{Heads} \mid \text{Monday}) = \frac{1}{2} \).
- Clearly \( P(\text{Monday} \mid \text{Heads}) = 1 \).
- I’m not going to claim exactly what \( P(\text{Monday}) \) is, but certainly it is less than 1 (i.e. it could be Tuesday).
- \( P(\text{Heads}) = p \) is what we’re looking for.
The Argument of Adam Elga, a thirder

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads})P(\text{Heads})}{P(\text{Monday})}. \]

- We already argued that \( P(\text{Heads} \mid \text{Monday}) = \frac{1}{2} \).
- Clearly \( P(\text{Monday} \mid \text{Heads}) = 1 \).
- I’m not going to claim exactly what \( P(\text{Monday}) \) is, but certainly it is less than 1 (i.e. it could be Tuesday).
- \( P(\text{Heads}) = p \) is what we’re looking for.
- Hence

\[
\frac{1}{2} = \frac{1 \cdot p}{\text{something} < 1}.
\]
The Argument of Adam Elga, a thirder

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads})P(\text{Heads})}{P(\text{Monday})}. \]

- We already argued that \( P(\text{Heads} \mid \text{Monday}) = 1/2. \)
- Clearly \( P(\text{Monday} \mid \text{Heads}) = 1. \)
- I’m not going to claim exactly what \( P(\text{Monday}) \) is, but certainly it is less than 1 (i.e. it could be Tuesday).
- \( P(\text{Heads}) = p \) is what we’re looking for.
- Hence
  \[
  \frac{1}{2} = \frac{1 \cdot p}{\text{something}} < 1.
  \]
- Solving for \( p \), we get \( p < 1/2. \).
Response by David Lewis

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads})P(\text{Heads})}{P(\text{Monday})} \]
Response by David Lewis

\[
P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads}) P(\text{Heads})}{P(\text{Monday})}.
\]

David Lewis, who insists that \( P(\text{Heads}) = 1/2 \), says the Elga argument actually leads to the counterintuitive result that \( P(\text{Heads} \mid \text{Monday}) = 2/3 \), even though the coin has not yet been flipped!
Response by David Lewis

\[ P(\text{Heads} \mid \text{Monday}) = \frac{P(\text{Monday} \mid \text{Heads})P(\text{Heads})}{P(\text{Monday})}. \]

- David Lewis, who insists that \( P(\text{Heads}) = 1/2 \), says the Elga argument actually leads to the counterintuitive result that \( P(\text{Heads} \mid \text{Monday}) = 2/3 \), even though the coin has not yet been flipped!

- Who do you side with?
Variations

The Case of Cloning SB

The Setup
Just like Classic SB, but instead of being woken up twice in the case of tails, SB is cloned, and each copy of SB is asked what is the probability the coin came up heads.

Analysis
Most people agree in this case that $p = 1/2$ is the right answer. Some halfers argue that this case is identical to Classic SB.
Variations

**The Case of Many Awakenings**

<table>
<thead>
<tr>
<th>The Setup</th>
<th>Instead of being awoken just twice in the case of tails, SB is awoken a hundred different times, each time thinking it is still Monday.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>A thirder is forced to believe $p = 1/101$ during the experiment, which highlights the counterintuitiveness of the thirder position.</td>
</tr>
</tbody>
</table>
The Case of Gambling SB # 1

The Setup

Every day SB is awakened, a gambler comes by and allows SB to wager, even money, that the coin came up tails. Should she take the bet?

Analysis

Obviously she should. By always taking the bet, she has a 50% chance of losing, and a 50% chance of winning, but if she wins she’ll effectively be paid twice. So she comes out ahead.
Variations

The Case of Gambling SB # 2

The Setup
Every day SB is awakened, a gambler comes by and allows SB to wager, even money, that the coin came up tails. However, she will be paid later, and if she is offered the bet twice, one of the two decisions at random will be taken to be her decision.

Analysis
Now the bet is no longer in her favor. These two gambling cases give credence to the proposition that neither the halfers or thirders are right, but it depends on how the question is asked.
References