Das 'Paradox' der unerwarteten Hinrichtung

florian@math.uni-bielefeld.de

Ein rational handelnder Richter macht die folgende Aussage über einen Gefangenen in dessen Anwesenheit: „Der Gefangene wird in der nächsten Woche gehängt und der Tag seiner Hinrichtung wird eine Überraschung für ihn sein."

Der Gefangene, der von der rationalen Handlungsweise des Richters weiß, macht folgende Beobachtung: Wenn ich am Sonntag noch lebte, müsste ich am Sonntag hingerichtet werden. Dies wäre aber keine Überraschung für mich. Also werde ich nicht am Sonntag hingerichtet. Wenn ich am Samstag noch lebte, müsste ich am Samstag hingerichtet werden, da ich am Sonntag nicht hingerichtet werde. Dies wäre aber keine Überraschung für mich. Also werde ich auch am Samstag nicht hingerichtet, usw. Also werde ich überhaupt nicht hingerichtet.

Am Mittwoch allerdings kommt der Henker und richtet den Gefangenen zu dessen Überraschung hin. Die Aussage des Richters hat sich aber trotzdem erfüllt. Was ist das Problem?

Das Problem liegt in der unklaren Eigenschaft eine 'Überraschung' zu sein. Für die Aussage des Richters gibt es mindestens zwei formale Interpretationsmöglichkeiten, die mit den beiden in Frage kommenden Handlungsweisen des Henkers korrespondieren. Zur Klärung wird im folgenden versucht, die Situation ein wenig zu formalisieren.

Sei $W = \{6, 5, 4, 3, 2, 1, 0\}$ die Menge der potentiellen Hinrichtungstage und definiere die Aussage

$$A_{-1} = \text{„}\textit{Es gibt einen Tag } t \in W\textit{, an dem der Gefangene gehängt wird}\text{“}$$

sowie die in der Realität offensichtlich richtige Aussage

$$\begin{aligned} B \quad = \quad & \text{„}\textit{Der Gefangene ist am Tag } t < 6 \textit{ noch am Leben}\text{“} \Longrightarrow \\ & \text{„}\textit{Der Tag der Hinrichtung ist nicht einer der Tage } 6, ..., t+1\text{“} \end{aligned}$$

Definiere nun induktiv für alle $t \in W$ die Aussage

$$A_t = \begin{array}{l} \text{„}\textit{Der Tag der Hinrichtung ist nicht aus den} \\ \textit{Aussagen } A_{t-1}, ..., A_{-1} \textit{ und } B \textit{ ableitbar}\text{“} \end{array}$$

1. Möglichkeit
Es gilt B. Die Aussage des Richters ist: A_0 und A_{-1}, d.h. *„Der Tag der Hinrichtung ist nicht aus den Aussagen A_{-1} und B ableitbar"* und *„Es gibt einen Tag $t \in W$, an dem der Gefangene gehängt wird"*. Dies scheint die naheliegenste Interpretation der Aussage des Richters zu sein. Der Gefangene kann wie folgt

den Sonntag $t = 0$ als Hinrichtungstag ausschließen: Angenommen der Gefangene wäre am Tag 0, d.h. sonntags noch am Leben. Es gilt A_{-1}, d.h. die Aussage *„Es gibt einen Tag $t \in W$, an dem der Gefangene gehängt wird“*. Da B gilt und $0 < 6$, fielen die Tage $6, ..., 1$ als Hinrichtungstage aus, also würde der Gefangene am Tag 0 hingerichtet. Dies steht allerdings im Widerspruch zu der Aussage A_0, d.h. *„Der Tag der Hinrichtung ist nicht aus den Aussagen A_0, A_{-1} und B ableitbar“*.
Das Problem ist nun allerdings, daß auf diese Weise keine anderen Tage von dem Gefangenen als Hinrichtungstage ausgeschlossen werden können, d.h. es ist nicht widersprüchlich, wenn der Henker den Gefangenen am Mittwoch zu dessen Überraschung hinrichtet.

2. Möglichkeit
Es gilt B. Die Aussage des Richters ist: Für alle $t \in W$ gilt A_t und A_{-1}. In diesem Fall kann der Gefangene tatsächlich alle Tage der Woche als Hinrichtungstage wie folgt ausschließen: Durch Induktion über $t \in W$ wird die Aussage

> „Aus den Aussagen $A_t, ..., A_{-1}$ und B ist ableitbar, daß der Gefangene nicht an den Tagen $t, ..., 0$ hingerichtet wird.“

gezeigt.

$t = 0$ Wie bei der 1. Möglichkeit gezeigt wurde, ist aus den Aussagen A_0, A_{-1} und B ableitbar, daß der Tag der Hinrichtung nicht $s = t = 0$ ist.

$t - 1 \to t$, $1 \leq t \leq 6$ Angenommen es gelten die Aussagen $A_t, ..., A_{-1}$ und B. Nach Induktionsvoraussetzung gilt, daß der Gefangene nicht an den Tagen $t - 1, ..., 0$ hingerichtet wird. Angenommen der Gefangene wäre am Tag t noch am Leben. Es gilt A_{-1}, d.h. die Aussage *„Es gibt einen Tag $t \in W$, an dem der Gefangene gehängt wird“*. Falls $t = 6$ dann würde der Gefangene also am Tag t hingerichtet. Falls $t < 6$, so fielen nach B die Tage $6, ..., t + 1$ als Hinrichtungstage aus, also würde der Gefangene am Tag t hingerichtet. Also würde in beiden Fällen der Gefangene am Tag t hingerichtet. Dies steht allerdings im Widerspruch zu der Aussage A_t, d.h. *„Der Tag der Hinrichtung ist nicht aus den Aussagen $A_{t-1}, ..., A_{-1}$ und B ableitbar“*. Also gilt die Behauptung.