AN UNDECIDABLE ASPECT OF THE UNEXPECTED HANGING PROBLEM

JACK M. HOLTZMAN

Introduction

An apparent paradox which has received some attention has appeared in several guises (unexpected hanging, unexpected examination, etc.). There have been a number of explanations and critiques of explanations of this apparent paradox; a survey with 40 references is given in Margalit and Bar-Hillel [1].

Although a number of papers have claimed to have resolved the paradox, subsequent papers typically start off with a lack of complete satisfaction with previous explanations. The crux of a number of papers is an explanation of how an announced unexpected hanging can indeed be unexpected on the last day. The alternative, not to have an unexpected hanging on the last day, appears to lead, by a notorious backward-in-days reasoning, to the paradoxical impossibility of the unexpected hanging on any previous day.

In the formulation given here, an unexpected hanging is precluded on the last day. The backwards chain of reasoning is broken by an undecidable aspect which arises *because* of the conditions which precluded a last day unexpected hanging. Adapting terminology used in other contexts, we use *decidable* to mean that there is an effective procedure that, for each positive integer n, provides the answer as to whether n satisfies a condition C (to be defined) which characterizes the problem. Decidability is discussed in several chapters of Barwise [2].

A paper on this problem related to decidability is by Meltzer [3]. However, Meltzer was only concerned with an unexpected event on the last day and, furthermore, that paper was criticized by Meltzer and Good [4] (although Meltzer maintained his view that a decidability-related notion should play a role).

The Problem and Assumptions

I will summarize (once again) one version of the problem. A man was sentenced on Saturday to hanging at noon on one of the following seven days. The judge also said that the prisoner would not know which day until he was informed the morning of the hanging. The prisoner concluded that the hanging could not take place. He could not be hanged on Saturday because there would be no surprise on the last day. Nor could he be hanged on Friday because Saturday was already ruled out. Continuing this reasoning backwards among the days, he eliminated all the days. To his surprise, he was hanged on Wednesday.

Consider the judge's claim:

A_1. The prisoner will be hanged on one of the following n days (n=7 in the above but we will let n be a variable in the following) and

A_2. The hanging will be unexpected,

to which we add

A_3. The prisoner has no doubt about the judge's intention and ability to have him hanged.

We focus here on the satisfaction of C, the conjunction of A_1, A_2, and A_3. That is, it is required by C that the hanging takes place, that it comes unexpectedly, and the unexpectedness is not due to the prisoner having doubt as to the judge's intent and ability to have him hanged. A_3 is included here to sharpen the definition of the problem, making it more difficult to explain away. In fact, it is the inclusion of A_3 that distinguishes the orientation of this paper from a number of others. In these other approaches, emphasis is on explaining how an unexpected hanging can occur on the last day (as mentioned previously) and assumptions like A_3 would have to be disallowed. Rather than preclude A_3 because it disallows a last day unexpected hanging, we concede the last day and seek other days for the unexpected hanging.

We shall say that a value of n satisfies C if C can be satisfied with that value of n in A_1. Since C cannot be satisfied on the last day, n=1 does not satisfy C. As discussed on pp. 264,265 of Margalit and Bar-Hillel, an unexpected hanging can take place, which means that some n satisfies C. It would thus seem that there should be a minimum n satisfying C.

However, a minimum n satisfying C cannot be determined. That is, if we assume that n_{min} satisfies C and $n<n_{min}$ does not, then C cannot be satisfied with $(n_{min}-1)$ days. Hence, on D, the first of n_{min} days, the hanging must take place, since passing the first day leaves only $(n_{min}-1)$ days left which is too small for satisfaction of C. The prisoner, using knowledge of n_{min}, can deduce that the

hanging must take place on D to satisfy C. Thus, the hanging will not be unexpected,[1] contradicting the definition of n_{min}. (Note that while a general answer as to whether each n satisfies C is not available, that does not preclude answering the question for some n.)

More formally and more generally, define the partial function f: $\Omega \rightarrow [0,1]$ where Ω is a subset of the set of positive integers $N+$. On Ω there is an effective procedure yielding:

$$f(n) = 1 \quad \textit{if } n \textit{ satisfies } C$$
$$\quad = 0 \quad \textit{if } n \textit{ does not satisfy } C$$

The totality of f on $N+$ is decidability. For any n, $f(n+1)=1$ implies $f(n) \neq 0$ so that if f is total $f(n)=0$ implies $f(n+1)=0$ for all n. And then, since $f(1)=0$, $f(n)=0$ for all n. Thus, with decidability, the apparent paradox reappears. With the undecidable aspect (the undecidability of the jump in $f(n)$ to 1), the prisoner's backward chain of reasoning is broken.

Discussion

To clarify the above, we elaborate on the questions:

1. Why does the decidability issue challenge the prisoner's reasoning?
2. Why is there an undecidable aspect?

In answer to question (1), the prisoner's reasoning implicity assumes decidability when the only alternatives considered for each day are whether or not an unexpected hanging can take place. In a case where decidability is an issue, the prisoner's argument is incomplete without a demonstration of decidability.

Undecidability is suggested by the following consideration. What can be deduced from the assumptions can be deduced by both the judge and the prisoner. But the judge and the prisoner cannot both deduce something which must represent a surprise to the prisoner (e.g., an n_{min}). Hence, the suggestion of undecidability. Moreover, the prisoner can be unaware of a choice made by the judge which is not deducible from the assumptions.

Acknowledgement

I am grateful to Richard Ascher for bringing my attention to this problem and for stimulating discussions.

48 SHADOWLAWN DRIVE
LIVINGSTON, NEW JERSEY 07039
USA

NOTE

[1] A possible objection to this is that the hanging on D could be said to be unexpected because the prisoner may not know whether the hanging will be unexpected, allowing the prisoner to think that the judge may possibly pass D. This objection can be obviously fixed by adding assumption A_4: The prisoner believes A_2. It can be shown that A_4 is not problematical in our formulation as it is in formulations that try to retain unexpectedness on the last day, However, there is an argument against needing A_4 based on D being different than a last day where the prisoner's disbelief is forced to arise for unexpectedness to occur. Without a priori assuming either A_4 or its negation, it is possible that the prisoner makes the deduction.

REFERENCES

[1] A. Margalit and M. Bar-Hillel, "Expecting the Unexpected", *Philosophia*, vol. 13, October 1983, pp. 263–289.

[2] J. Barwise, Editor, *Handbook of Mathematical Logic*, North-Holland Publishing Co., Amsterdam, 1977.

[3] B. Meltzer, "The Third Possibility", *Mind*, vol. 73, July 1964, pp. 430–433.

[4] B. Meltzer and I.J. Good, "Two Forms of the Prediction Paradox", *British Journal for the Philosophy of Science*, vol. 16, May 1965, pp. 50–51.