One Hundred Prisoners and a Light Bulb

Hans van Ditmarsch & Barteld Kooi

Illustrations by Elancheziyan

Springer

One Hundred Prisoners and a Light Bulb

Hans van Ditmarsch • Barteld Kooi

One Hundred Prisoners and a Light Bulb

Illustrations by Elancheziyan

Springer

Copernicus Books is a brand of Springer

Hans van Ditmarsch
LORIA, CNRS
Université de Lorraine
Vandoeuvre-lès-Nancy
France

Barteld Kooi
Faculty of Philosophy
University of Groningen
Groningen
The Netherlands

Illustrations (c)2015 by Hans van Ditmarsch and Barteld Kooi

ISBN 978-3-319-16693-3 ISBN 978-3-319-16694-0 (eBook)
DOI 10.1007/978-3-319-16694-0

Library of Congress Control Number: 2015933648

Mathematics Subject Classification (2010): 00A08

Honderd gevangenen en een gloeilamp
Original Dutch edition published by (c) Epsilon Uitgaven, Amsterdam, 2013

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Copernicus is a brand of Springer

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

This puzzlebook presents 11 different puzzles about knowledge and ignorance. Each puzzle is treated in depth in a separate chapter, and each chapter also contains additional puzzles for which the answers can be found at the back of the book. A constant theme in these puzzles is that the persons involved make announcements about what they know and do not know, and then later appear to contradict themselves. Such knowledge puzzles have played an important role in the development of an area known as dynamic epistemic logic. A separate stand-alone chapter gives an introduction to dynamic epistemic logic.

The illustrations for this book were made by Elancheziyan. Elancheziyan is a Tamil speaking Indian illustrator living in Chennai. Hans has an associate position at the Institute of Mathematical Sciences (IMSc) in Chennai, India. By the intermediation of his IMSc host Ramanujam, and the kind assistance of Shubashree Desikan, who acted as a Tamil-English interpreter, he got in contact with Elancheziyan. How the illustrations to each chapter came about is story in itself, and we are very grateful for Elancheziyan's essential part in this joint enterprise.

We wish to thank Paul Levrie and Vaishnavi Sundararajan for their substantial and very much appreciated efforts to proofread the final version of the manuscript. Peter van Emde Boas has indefatigably provided details on the history of the Consecutive Numbers riddle, and has much encouraged us in writing this book. We wish to thank Allen Mann, Springer, for his encouragement and for getting us started on this project. Nicolas Meyer from the ENS des Mines in Nancy found an embarrassing error in a light bulb protocol when Hans gave a course there, only a few weeks before we handed over the manuscript. He is one of many. If one were to go back all the 25 years of teaching logic and puzzles at colleges, universities, and summer schools, a much longer list of thanks to students and colleagues would be appropriate: by making an example of one, we wish to thank them all. No doubt, there will still be many remaining errors. They are all the responsibility of the authors.

Nancy, France, and Groningen,
the Netherlands
25 December 2014

Hans van Ditmarsch
and Barteld Kooi

Contents

1
Consecutive Numbers

Anne and Bill get to hear the following: "Given are two natural numbers. They are consecutive numbers. I am going to whisper one of these numbers to Anne and the other number to Bill." This happens. Anne and Bill now have the following conversation.

- *Anne: "I don't know your number."*
- *Bill: "I don't know your number."*
- *Anne: "I know your number."*
- *Bill: "I know your number."*

First they don't know the numbers, and then they do. How is that possible? What surely is one of the two numbers?

The natural numbers are the numbers 0, 1, 2, 3, etc. Numbers are consecutive if they are one apart. It is important for the formulation of the riddle that Anne and Bill are simultaneously aware of this scenario, and also know that they both are aware of this scenario, etc. Therefore, they are being spoken to, instead of, for example, both receiving written instructions. It is therefore too that the numbers are whispered into their ears—the whispering creates common knowledge that they have received that information. We can imagine the setting of this riddle as Anne, Bill, and the speaker sitting round a table, such that the speaker has to lean forward to Anne in order to whisper to her, and subsequently has to lean forward to Bill and whisper to him.

1.1 Which Numbers Are Possible?

We solve the riddle by analyzing the developing scenario piecemeal. The first bit of information is as follows:

- Given are two natural numbers.

H. van Ditmarsch, B. Kooi, *One Hundred Prisoners and a Light Bulb*,
DOI 10.1007/978-3-319-16694-0_1

We do not know yet what these numbers are, but apparently there are two relevant variables: the number x that Anne is going to hear and the number y that Bill is going to hear. The question is then to determine the pair (x, y). We also know that x and y are *natural numbers*: 0, 1, 2, etc. So, the possible pairs are $(0, 0)$, $(0, 1)$, $(1000, 243)$, etc. Of course there are infinitely many such pairs. The state space consisting of all such pairs looks as follows—to simplify the representation we write xy instead of (x, y), and for convenience we order the number pairs in a grid.

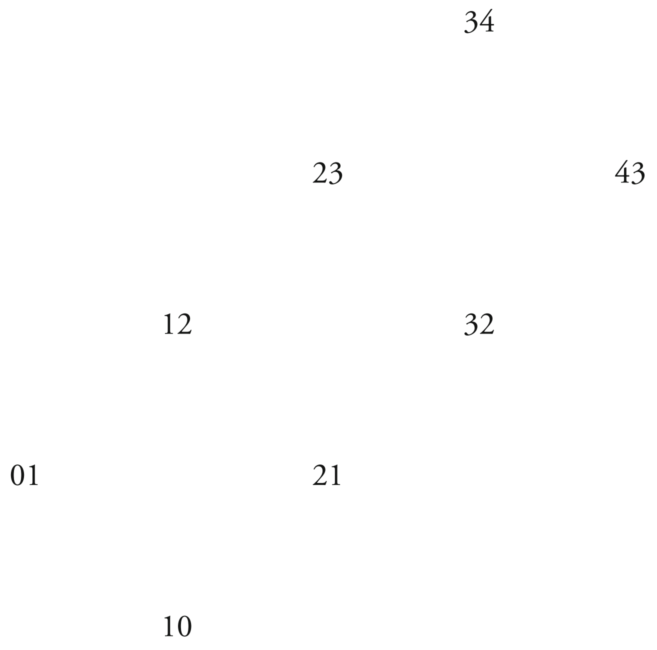
⋮	⋮	24	34	44
03	13	23	33	43
02	12	22	32	42
01	11	21	31	⋯
00	10	20	30	⋯

The number pair $(1, 2)$ is different from the number pair $(2, 1)$: The first of each pair is the number that Anne is going to hear, whereas the second of each pair is the number that Bill is going to hear. In $(1, 2)$, Anne is going to hear 1, and in $(2, 1)$ she is going to hear 2.

The next bit of information is that

- They are consecutive numbers.

This means that the only possible number pairs (x, y) are those where $x = y + 1$ or $y = x + 1$. Hence, only these pairs remain:

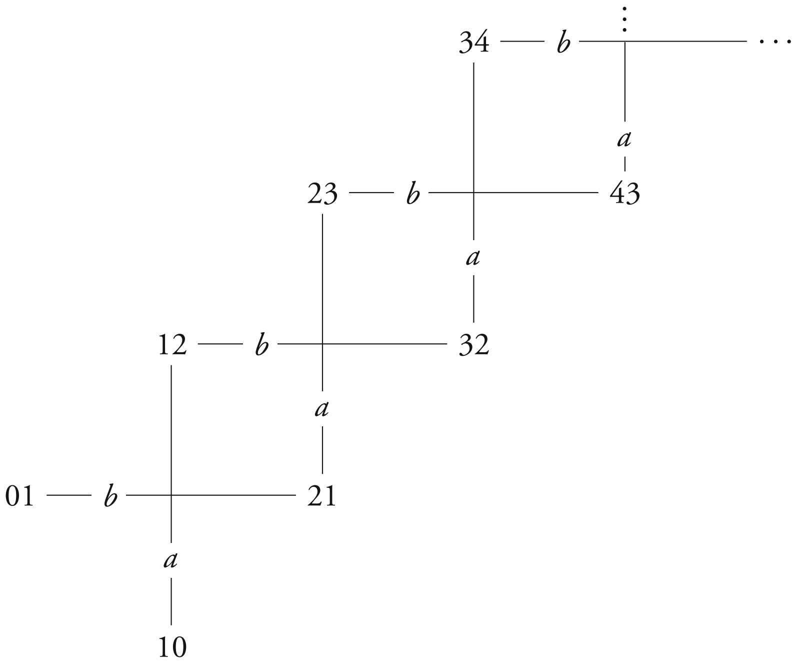


1.2 What Anne and Bill Know

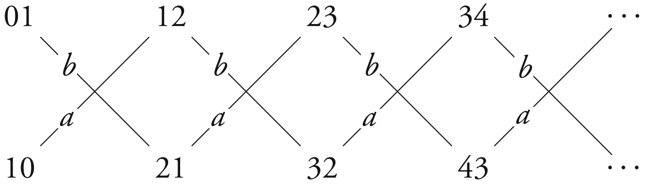
So far, your perspective, as reader, is the same as Anne's and Bill's: The numbers are natural numbers, and they are consecutive. These are all the possibilities that we have to take into account. We cannot distinguish among these pairs. The next bit of information makes Anne's and Bill's perspective different from your perspective as reader:

- "I am going to whisper one of these numbers to Anne and the other number to Bill." This happens.

Suppose that the whispered numbers were 5 to Anne and 4 to Bill. After Anne hears 5, she knows that Bill's number is 4 or 6. She can rule out all number pairs except $(5, 4)$ and $(5, 6)$. Bill's view of the situation is different from Anne's. He hears 4. After that, the remaining number pairs from his perspective are $(5, 4)$ and $(3, 4)$. You, the reader, cannot rule out any number pair! But you still have learnt something, namely what Anne and Bill learnt about any number pair and about each other. We can make the information change visible in the given set of consecutive number pairs: We can indicate which pairs are indistinguishable for Anne or for Bill after the whispering has taken place. A visual means is to link such pairs by an edge labeled with *a* for Anne, or *b* for Bill. We get:



We might as well have the figure topple over a bit to save space on the page:



In fact, we simply have two infinitely long chains of number pairs, with alternating labels. So, alternatively, just one of those is as follows:

10 —*a*— 12 —*b*— 32 —*a*— 34 —*b*— · · ·

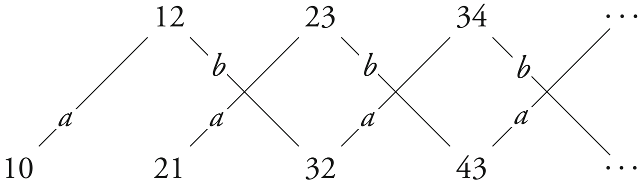
Anne's and Bill's perspectives are now different from each other and also from your perspective as a reader. Before the whispering action, all number pairs were equally possible for Anne, for Bill, and for you. After the whispering, all number pairs remain possible for you—they can equally well be 3 and 4, or 5 and 4, or 89 and 88—but for Anne and Bill this is no longer the case: If Anne were to have 3, she would know that the other number cannot be 88, but only 2 or 4. What you have learnt as a reader is that Anne and Bill now have this knowledge.

1.3 Informative Announcements

A figure such as the above we call a *model* of the description of the initial state of the riddle. We changed the model piecemeal with every new bit of information in the problem description. There were two sorts of changes: eliminating number pairs (for example, those number pairs that were not consecutive numbers), and indicating which number pairs could be distinguished by Anne and by Bill (for example, that Anne can distinguish (2, 3) from (5, 6), but not (2, 3) from (2, 1)). Next on our list of problem-solving activities is to convert each announcement by Anne and Bill into some such model transforming operation. In this riddle, all further changes are of the first kind: elimination of number pairs. The crucial aspect here is that we do not treat Anne's announcement differently from the "announcements" of the anonymous speaker who informs Anne and Bill in the beginning. Anne and Bill both hear their own announcements, and know from one another that they both hear what they say, and so on. And also, you as a reader can be said to be "hearing" the announcements: You have to imagine yourself as silent bystander present at the interaction between the initial speaker and Anne and Bill, and at their subsequent announcements. Let us take the first announcement:

- Anne: "I don't know your number."

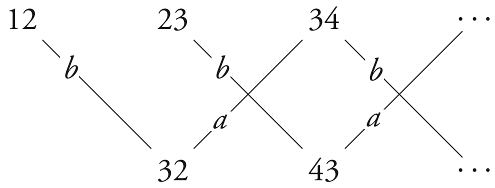
When would Anne have known what Bill's number is? Suppose Anne had heard 0. She knows that Bill's number is one more or one less than her own. It cannot be -1, as this is not a natural number. Therefore, the only remaining possibility is that Bill's number is 1. So, Anne then *knows* that Bill has 1. However, as she says, "I don't know your number," we can rule out the number pair (0, 1). And not just we, but also Bill. The change is public (for Anne and for Bill), because Anne said it aloud. If she had, for example, written it on a piece of paper, this might have created uncertainty in her whether the message had reached Bill, or uncertainty in Bill whether Anne knew that the message had reached him, and so on. The message would not have been public. Given that the change is public, the result is as follows:



It is now crucial to observe that this is a different model, and that it may therefore satisfy different propositions. Propositions that were false before may now be true, and propositions that were true before may now be false. This will explain why saying, “I don’t know your number” now and “I know your number” later only appears to be in contradiction, but is not really a contradiction. These observations are about different information states of the system. The announcements help us to resolve our uncertainty about what the number pair is. Similarly, it will help Anne and Bill to resolve their uncertainty. We continue our analysis by processing the next announcement:

- Bill: “I don’t know your number.”

When would Bill have known what was Anne’s number? There are two possibilities. In the first place, Bill would have known Anne’s number if the number pair had been $(2, 1)$. If Bill has 1, then he can imagine Anne to have 0 and 2. Given that 0 is no longer possible after Anne’s (first) announcement, only 2 remains. So, Bill then knows that Anne’s number is 2. But there is yet another pair where Bill would have known Anne’s number, namely $(1, 0)$. Now, just like Anne in the case of $(0, 1)$, Bill would have known that Anne has 1 because -1 is not allowed. Because Bill said, “I don’t know your number,” neither of these two pairs can be the actual pair. The resulting situation is as follows:



This brings us to the third announcement:

- Anne: “I know your number.”

We can see in the model that this is true for the number pairs $(2, 3)$ and $(1, 2)$, as there is then no alternative left for Anne. We can alternatively see this as the conclusion of a valid argument. For example, for the pair $(2, 3)$:

> If Anne has 2, then she now knows that Bill has 3, because, if Bill were to have 1, he would have said in the second announcement that he knew Anne’s number. But he did not.

All other number pairs have become impossible because of her announcement. The resulting model is therefore,

This depicts that if the numbers are 1 and 2, then Anne and Bill know this, know from one another that they know this, etc. It is common knowledge between them. If the numbers are 3 and 2, then they also have common knowledge of the numbers. Although both $(1,2)$ and $(2,3)$ are in the model, this does not mean that if the numbers are 1 and 2, then Anne and Bill also consider it possible that they are 2 and 3: There is no link for *a* or for *b* in the model. But you, as a reader, cannot determine which of the two pairs must be actually the case. We now get to the last announcement:

- Bill: "I know your number."

This proposition is already true for both remaining number pairs. Therefore, nothing changes. We could also have said: This last announcement was not informative. Anne already knew that Bill knew her number, and they both knew this.

This solves the riddle. All four announcements were truthful. The contradiction between "I don't know your number" and "I know your number" is not a contradiction in the riddle, because these announcements are made at different moments. What was true before can be false later. After the four announcements, the remaining number pairs are $(1,2)$ and $(2,3)$. You cannot choose between these two pairs. But the number 2 occurs in both pairs, and is therefore certainly one of the two numbers.

1.4 Versions

Puzzle 1 *Suppose that the actual numbers are neither* 1 *and* 2*, nor* 2 *and* 3*, but* 4 *and* 5*. The four announcements can no longer all be made truthfully. What is going wrong? How often does "I don't know your number" have to be repeated for Anne and Bill to get to know the other number, and by whom?*

Puzzle 2 *An alternative presentation of the riddle is as follows:*

> *Anne and Bill* **each have a natural number on their forehead***. They are consecutive numbers. Anne and Bill now have the following conversation.*
>
> - *Anne: "I don't know* **my** *number."*
> - *Bill: "I don't know* **my** *number."*
> - *Anne: "I know* **my** *number."*
> - *Bill: "I know* **my** *number."*
>
> *What difference does this formulation make for the solution?*

Puzzle 3 *Suppose that the numbers are not consecutive, but* **two** *apart. So, the riddle will be as follows:*

> *Anne and Bill get to hear the following: "Given are two natural numbers. The numbers are two apart. I am going to whisper one of these numbers to Anne and the other number to Bill." This happens. Anne and Bill now have the following conversation.*
>
> - *Anne: "I don't know your number."*
> - *Bill: "I don't know your number."*
> - *Anne: "I know your number."*
> - *Bill: "I know your number."*

What does the model look like in this case, and how it is transformed due to the announcements? And what if the numbers are m apart, where m is a natural number?

Puzzle 4 *Suppose there is a third person playing the game, Catherine. Now, the riddle is:*

> *Anne, Bill,* **and Catherine** *each have a natural number on their forehead. They are consecutive numbers. Suppose, for example, that the numbers are* 3, 4, *and* 5 *(respectively). What sort of conversation is possible between Anne, Bill, and Catherine, on knowledge and ignorance of each other's number, in order to find out their own number?*

Puzzle 5 *Anne and Bill have a natural number on their forehead. It is known that the sum of these two numbers is equal to* 3 *or* 5. *Anne and Bill may now consecutively announce whether they know their own number. Show that they can have the following conversation:*

- *Anne: "I don't know my number."*
- *Bill: "I don't know my number."*
- *Anne: "I know my number."*
- *Bill: "I know my number."*

(After Conway et al. (1977); see the history section below.)

1.5 History

An original source for the riddle is found straight at the beginning of *A Mathematician's Miscellany* by Littlewood (1953, p. 4):

> There is an indefinite supply of cards marked 1 and 2 on opposite sides, and of cards marked 2 and 3, 3 and 4, and so on. A card is drawn at random by

> a referee and held between the players A, B so that each sees one side only. Either player may veto the round, but if it is played the player seeing the higher number wins. The point now is that every round is vetoed. If A sees a 1 the other side is 2 and he must veto. If he sees a 2 the other side is 1 or 3; if 1 then B must veto; if he does not then A must. And so on by induction.

In the Littlewood version, there is no "solution" (every round is vetoed), and the synchronization is left open to interpretation (who vetoes first?). But a player seeing number x on one side of the playing card is uncertain if the number on the other side is $x + 1$ or $x - 1$. Only when a player is seeing the number 1 can he be certain about the other number, namely that it is 2 (the number 0 is ruled out). This version is also treated, slightly differently, by Gardner (1977):

> You are one of two contestants in the following game: An umpire chooses two consecutive positive integers entirely at random and writes the two numbers on slips of paper, which he then hands out randomly to the two players. Each looks at their number and either agrees or disagrees to play. If both players agree, the person with the higher number must pay that many dollars to their opponent. You only agree to play when the expected payout favors you. Obviously, you would agree if your number was 1. For what other values should you agree to play?
>
> Assume infinite resources for payouts. I.e. it does not matter how high the numbers are, the payment can be made.

A far more general version of the riddle is found in *A Headache-Causing Problem* by Conway et al. (1977). This is a contribution to an honorary volume "presented to Hendrik W. Lenstra on the occasion of his doctoral examination." The treatment is light, for example, the initials of the third author are "U.S.S.R." This is because Paterson and Conway discussed the riddle while waiting in transit on Moscow airport (as van Emde Boas recently found out).

> There are n persons, all having a natural number on their forehead. It is known that the sum of these n numbers is equal to one of at most n possible given numbers. The n players may now consecutively announce if they know their own number, until one of them says that he or she knows it. Prove that this will happen eventually.

The last publication in this series of original sources is then *The Conway Paradox: Its Solution in an Epistemic Framework* by van Emde Boas, Groenendijk, and Stokhof, orginally presented at the Amsterdam Colloquium in 1980, afterwards published in *Mathematical Centre Tract No. 135* in 1981, and finally published in book format in (van Emde Boas et al. 1984). This publication is an important precursor of dynamic epistemic logic. It also provides a very accurate historical section, on which this overview is based. After their publication, the consecutive numbers riddle became known as the *Conway paradox*. Yet another nice story comes with that: It is curious to observe that the consecutive numbers riddle, even though it is now known as the Conway paradox, is *not* a special case of the problem described in Conway et al. (1977), so that "Conway paradox" is actually a misnomer for the consecutive numbers riddle, as van Emde Boas confirms.

For example, if Anne has 3 on her forehead and Bill 2, that indeed involves uncertainty by two players about two numbers, and therefore also about two sums of numbers, but, unlike the Conway version, this is uncertainty about more than two sums: Anne is uncertain if the sum is 5 or 3, whereas Bill is uncertain if the sum is 5 or 7. And, of course, Bill is uncertain whether Anne is uncertain between sums 5 and 3, or between sums 7 and 9, and so on. An infinity of sums plays a role.

On a more abstract level (no doubt in the mind of van Emde Boas et al. at the time), there is of course a correspondence. See Puzzle 5.

2
Hangman

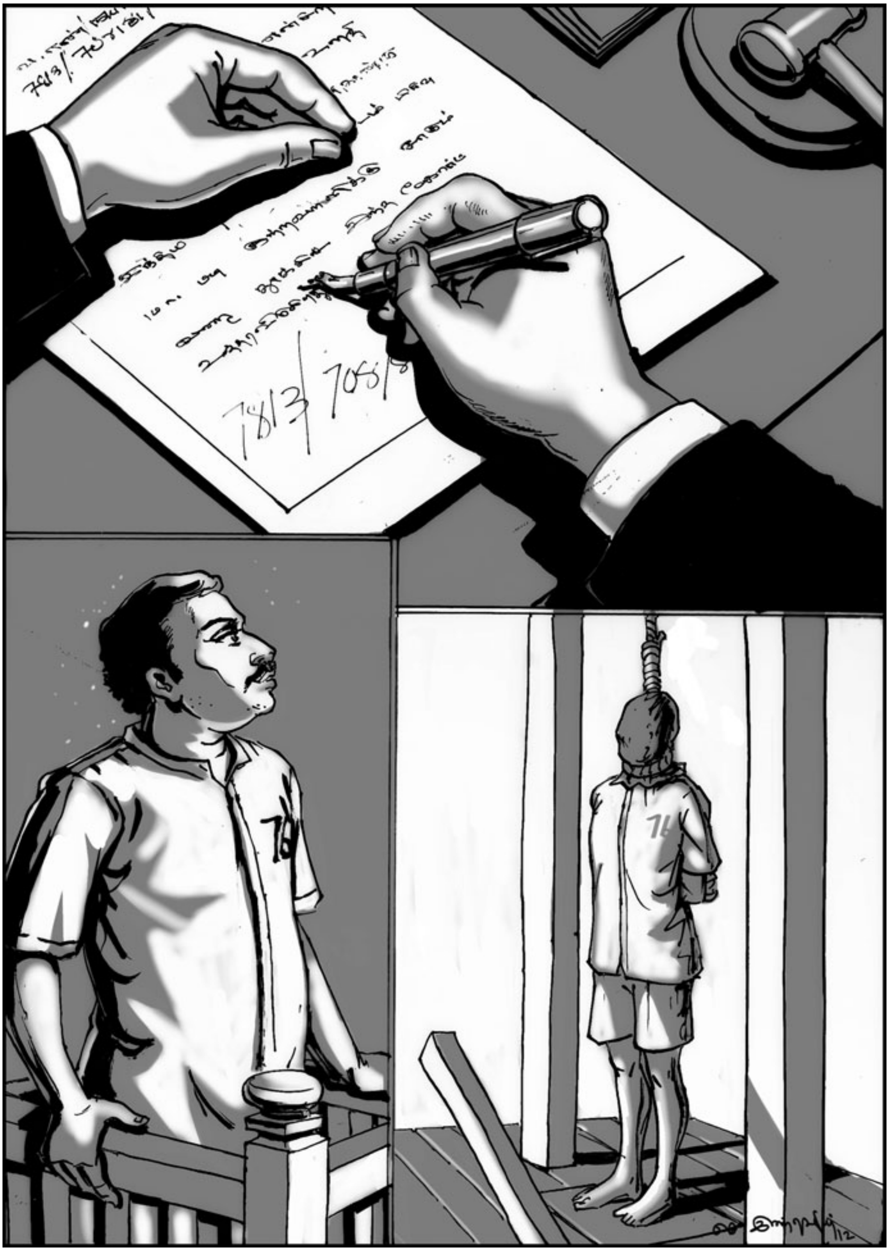
> *At a trial a prisoner is sentenced to death by the judge. The verdict reads "You will be executed next week, but the day on which you will be executed will be a surprise to you." The prisoner reasons as follows. "I cannot be executed on Friday, because in that case I would not be surprised. But given that Friday is eliminated, then I cannot be executed on Thursday either, because that would then no longer be a surprise. And so on. Therefore the execution will not take place." And so, his execution, that happened to be on Wednesday, came as a surprise.*
>
> *So, after all, the judge was right. What error does the prisoner make in his reasoning?*

The prisoner's argument is very convincing. At first sight it seems as if it cannot be refuted at all. Still, the conclusion cannot be right. The prisoner rules out that the hanging will be on Thursday, and that it will be on Wednesday, and so on, but in fact the hanging is on Wednesday. Is it not easy to make clear where the error is. And therefore it is indeed called a paradox. To find the error in the prisoner's reasoning, we first have to define what a "secret" is. Because initially the day of the hanging is a *secret*.

2.1 How to Guard a Secret?

The best way to guard a secret (like who you are in love with) is not ever to tell it to anyone. That is easier said than done. If your head is filled with the secret, it can happen to fall out of your mouth before you know it. And then it is no longer a secret. Someone might ask you why you are staring out of the window all the time, focusing on the horizon. You can then of course say that this is because you are guarding a secret. But that makes it less secret. If you really want to guard a secret, you had better not ever talk about it, because if you do, then you risk that the secret will be discovered.

H. van Ditmarsch, B. Kooi, *One Hundred Prisoners and a Light Bulb,*
DOI 10.1007/978-3-319-16694-0_2


7813/ 7081
76
76

It is also a bad idea to talk to yourself about your own secret. A classic of that kind (renewedly popular from the TV series *Once Upon a Time*) is the character Rumpelstiltskin in the Grimm Brothers fairy tale by the same name (1814). The queen promised her first-born child to Rumpelstiltskin. There is an escape clause: If the queen guesses correctly Rumpelstiltskin's name, then she can keep her child. She can guess three times. The first two guesses are incorrect. The tension rises. The queen's messenger now tells her that he saw in the forest, from a hidden place behind some bushes, a funny guy who was dancing while singing, loudly:

> Heute back ich, morgen brau ich,
> übermorgen hol ich der Königin ihr Kind;
> ach, wie gut dass niemand weiß,
> dass ich Rumpelstilzchen heiß!

Fortunately, the messenger understood German and could translate this into

> Today I'll bake; tomorrow I'll brew,
> Then I'll fetch the queen's new child,
> It is good that no one knows,
> Rumpelstiltskin is my name.

It was indeed Rumpelstiltskin who was singing this song, and so the queen finds out his name, and the third time her guess is correct: "Your name is Rumpelstiltskin." If only he had kept his mouth shut, it would have remained a secret.

The funny thing is, that the last two sentences, in a more convenient phrasing "Nobody knows that my name is Rumpelstiltskin," become false because Rumpelstiltskin is singing it. After this, it is no longer the case that nobody knows that his name is Rumpelstiltskin. The messenger now knows. This phenomenon is quite special. Apparently, it is possible to say something ("Nobody knows that I am in love with Stephanie") but because I am saying it, it becomes false. (In no time everyone, including Stephanie, knows that I am in love with her.) Usually when we say something, it remains true after we say it. But in exceptional cases, this is apparently false.

What is the relationship between hangings and fairy tales? The day of the hanging is a secret guarded by the judge, and the prisoner can only guess what the exact day is. The judge does not tell which day it is. What does it mean that the judge says that the day of the hanging will be a surprise? A surprise is something unexpected, it is something happening that you did not see coming. In the reasoning of the prisoner, "surprise" is entirely interpreted in terms of *knowledge*. The hanging is a surprise, because the prisoner does not *know* the day of the hanging in advance. A secret is no secret anymore if you are telling it to someone, just as for the secret of Rumpelstiltskin. Similarly, a surprise is not a surprise anymore if you announce it. If you want to surprise someone with a big bunch of roses, then you should not let it appear from your behavior. If you say, “I am going to surprise Stephanie tomorrow with a big bunch of roses,” then the surprise is lost when she hears about it. If Rumpelstiltskin says, “Nobody knows that my name is Rumpelstiltskin,” then someone may get to know it.

2.2 A Bridge Too Far

When the judge says that the day of the hanging will be a surprise, he risks spoiling the surprise. If he had not said anything about the hanging, not even that it was going to be next week, it would not have mattered; surely the prisoner would then have been surprised by the hanging.

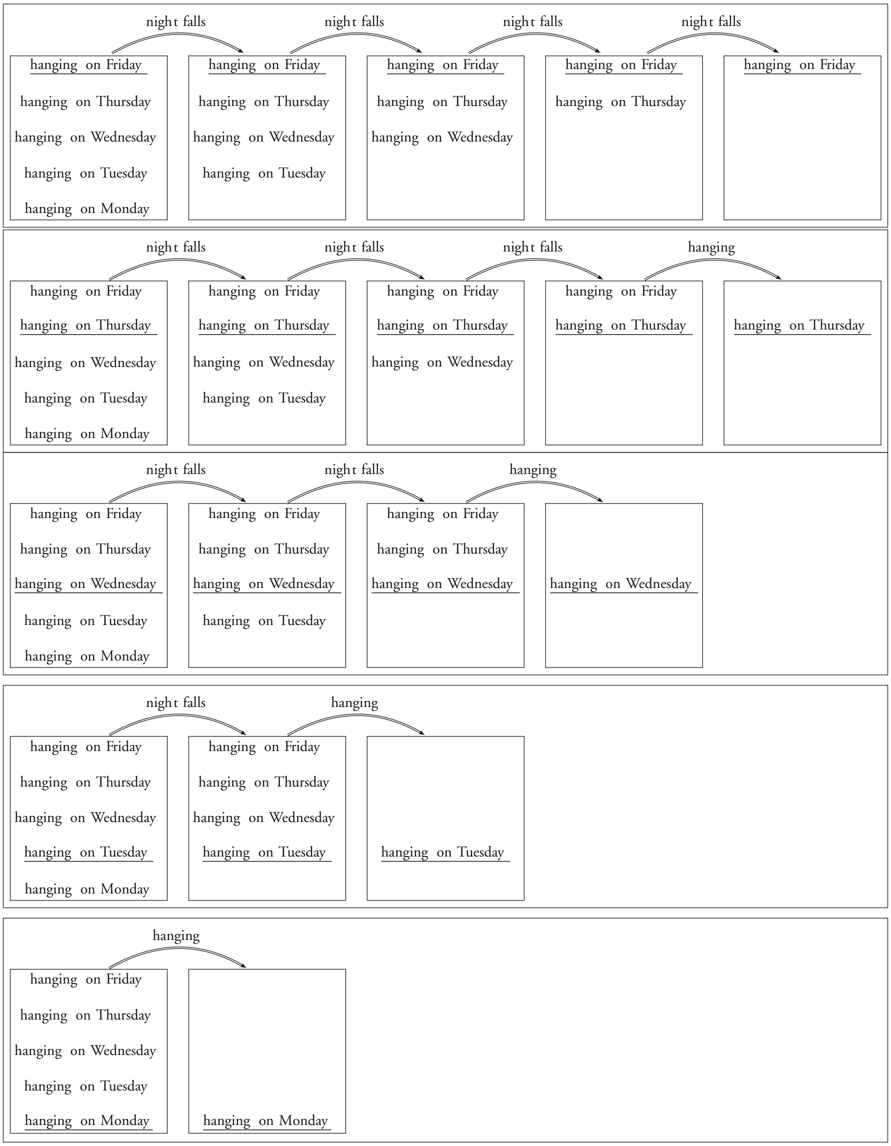
The error that the prisoner seems to make in his reasoning is that he does not realize that the judge may have spoilt the surprise by announcing it. Before the judge is saying that the day will be a surprise, the prisoner considers it possible that the hanging will take place on one of Monday, Tuesday, Wednesday, Thursday, or Friday. Now suppose nothing else has been announced about the day of the hanging. The prisoner would then know on Thursday night that the hanging will be on Friday. The hanging would then not be a surprise. On all other days, it would be a surprise. This, the judge also knows. But by saying that to the prisoner, he spoils the surprise. His announcement rules out that the hanging will be on Friday. Therefore, if the prisoner had not yet been hanged by Wednesday night, he could by that time have concluded that the hanging must be on Thursday. So now Thursday is special, instead of Friday.

However, the prisoner takes the argument further—and too far: He assumes that even after the judge’s announcement, the day of the hanging remains a surprise. And, therefore, he thinks he can rule out not only Friday but also Thursday, and Wednesday, and Tuesday, and Monday. But that is carrying it too far. Only Friday can be ruled out.

In fact, the hanging is on Wednesday. So, if the prisoner would not get more information, that would still be a surprise.

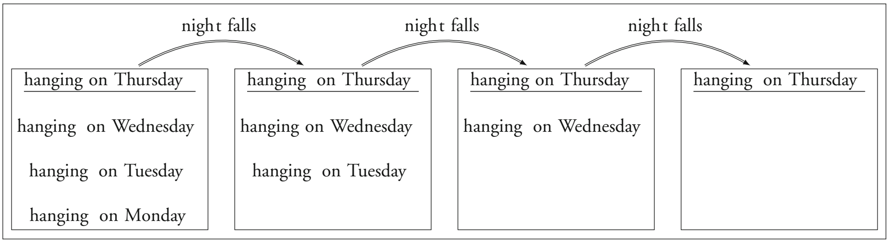
Let us illustrate this by constructing models. We assume that initially the prisoner only knows that there will be a hanging some day next week (a working day: Monday to Friday). So, this is before the judge announces that the day of the hanging will be a surprise. In that case, how will the prisoner’s information change with the passing of that coming week? Below we can see this depicted for the different days that the hanging can take place. Two events may reduce the uncertainty for the prisoner: Nightfall will rule out that the hanging is on the current day and thus reduces the uncertainty, but the hanging itself will confirm that it is on the current day and thus also reduces the uncertainty.

Hang on, what does it mean for a prisoner who has been hanged and who is dead, to know on Friday that he has been hanged on Thursday? Dead prisoners do not know anything. True enough, but this is an artifact of our setting of the riddle! In another version, the riddle concerns a surprise *exam* given by a schoolmaster to his pupils. Then, on Friday you will still know that the exam has been on Thursday. We can also imagine ourselves, as problem solvers, to be the agents observing the scenario and whose knowledge is being modeled. The problem solver will still know on Friday that the prisoner has been hanged on Thursday.



What is remarkable in these different scenarios is that there is only one occasion where the actual hanging does remove the uncertainty about the day of the hanging, for the prisoner. Namely, when the hanging is on Friday. Because (only) on that occasion the prisoner can determine the night before the hanging that the hanging will take place on Friday. So, for the prisoner there is only one day where the hanging will not be a surprise: Friday. If the judge announces that the hanging will be a surprise, this then rules out that the hanging is on Friday.

After the judge's announcement it is not necessarily so that the hanging will be a surprise. But there is now another scenario in the picture above where the hanging will not be a surprise, namely where Friday has been eliminated and where the hanging will be on Thursday. The prisoner does not know this in advance but knows that the hanging will be on Thursday when Wednesday night falls.



2.3 Versions

Puzzle 6 *Suppose that the judge has answered the question "On which day?" by "That will* not *be a surprise." On which day will the hanging then take place?*

A different wording of the riddle is not about a judge surprising a prisoner with the day of a hanging, but about a schoolmaster or teacher surprising a class with the day of an examination: "You will get an exam next week, but the day of the exam will be a surprise for you." Then, of course, the class only learns that the examination will not be on Friday. It is therefore also known as the surprise exam paradox. We now discuss a further version of that.

Puzzle 7 *Suppose the teacher, Alice, had only said that the exam would take place next week, but without saying that the exam would come as a surprise.*

During lunch break, her pupil Rineke walks past the staffroom and overhears the teacher saying to a colleague, "I am going to give my class an exam next week, and the day of the exam will be a surprise to them." The teacher did not realize that Rineke was overhearing her. What can Rineke conclude on the basis of this information about the day of the examination?

But the plot thickens. Because after lunch break Rineke says to the teacher, "I heard you say that the day of our exam next week will come as a surprise." The teacher confirms this. However, later that day, while getting her parked bike from the bikeshed, the teacher meets the staffroom colleague again, who is about to go home as well, and tells him that Rineke had overheard them earlier that day, and says, "But the day of the exam will still come as a surprise!" Unfortunately, Rineke overhears this again. What can Rineke now conclude about the day of the exam?

2.4 History

During the Second World War, the Swedish mathematician Lennart Ekbom overheard a radio message announcing a military training exercise next week. The training exercise would, of course, come as a surprise. It occurred to him that this message seemed paradoxical (Kvanvig 1998; Sorensen 1988, p. 253). The paradox was then published by O'Connor (1948). One of the responses to this publication then mentions that the exercise could still take place (Scriven 1951), what makes it even more paradoxical.

There are many versions of the paradox, the best known is the "surprise exam" version where a schoolmaster announces to his class that an examination will be given next week, but that the day will be a surprise (this first appeared in Weiss 1952). The "hangman" version of our presentation first appeared in Quine (1953).

The treatment of the puzzle differs depending on how "surprise" is interpreted. This can be done in many different ways. It can be in terms of derivability (the precise day *does not follow from* what the judge says). This approach was followed by Shaw (1958). But of course "surprise" can also be interpreted as "ignorance": lack of knowledge. This is what we have done here. "The prisoner will be surprised" then means that the prisoner does not *know* in advance when the hanging will take place.

Since 1948, more than 100 publications have appeared on the hangman paradox. They contain even more interpretations. A detailed overview of treatments of the paradox and its history is given by Sorensen (1988).

It is remarkable that all this "scientific work" has not resulted in a universally accepted solution of the paradox. Chow (1998) even calls it a meta-paradox:

> The meta-paradox consists of two seemingly incompatible facts. The first is that the surprise exam paradox seems easy to resolve. [. . .] The second (astonishing) fact is that to date nearly a hundred papers on the paradox have been published, and still no consensus on its correct resolution has been reached.

The solution given in this chapter is based on the work of Gerbrandy (1999, 2007). It is also treated by van Ditmarsch and Kooi (2005, 2006).

3
Muddy Children

> *A group of children has been playing outside and they are called back into the house by their father. The children gather round him. As one may imagine, some of them have become dirty from the play. In particular: they may have mud on their face. Children can only see whether other children are muddy, and not if there is any mud on their own face. All this is commonly known, and the children are, obviously, perfect logicians. Father now says: "At least one of you is muddy." And then: "Will those who know whether they are muddy step forward." If nobody steps forward, father keeps repeating the request. At some stage all muddy children will step forward. When will this happen if m out of k children in total are muddy, and why?*

This is a puzzling scenario. If there is more than one muddy child, all children see at least one muddy child, so they know that there is at least one muddy child. Father then says something that everyone already knows. If that is so, why say it? And why, after making the request to step forward, would he repeat this request? If nobody responds by stepping forward, what difference would it make to repeat the request? To understand that this makes a difference, we look at a simpler puzzle first.

3.1 Muddy or Not Muddy, That is the Question

Puzzle 8 *Alice and Bob are coming home from playing outside. Their father notices that they have been playing with mud, because Bob has mud on his face. They can only see mud on each other's face, but not on their own face. Of course, you can find out by looking in a mirror. Father now says "One of you has mud on his face." Bob now leaves and washes his face. However, he did not look in a mirror. How did he find out that he is muddy?*

To solve such puzzles we have to assume that all children are geniuses (what every parent will happily confirm): They are perfect logicians. Also, we assume that the father and his children are always speaking the truth, and that they have complete confidence in each other speaking the truth. If a child has mud on the face, we might as well say that *the child is muddy*.

H. van Ditmarsch, B. Kooi, *One Hundred Prisoners and a Light Bulb,*
DOI 10.1007/978-3-319-16694-0_3