An examination for the Unexpected Hanging Paradox

Duc M. Tran

Introduction

The Unexpected Hanging Paradox, also known as many of its variants The Surprise Examination or the surprise drill, is a popular logical paradox. The paradox is described as follow:

> "A judge tells a condemned prisoner that he will be hanged at noon on one weekday in the following week but that the execution will be a surprise to the prisoner. He will not know the day of the hanging until the executioner knocks on his cell door at noon that day.
>
> Having reflected on his sentence, the prisoner draws the conclusion that he will escape from the hanging. His reasoning is in several parts. He begins by concluding that the "surprise hanging" can't be on Friday, as if he hasn't been hanged by Thursday, there is only one day left - and so it won't be a surprise if he's hanged on Friday. Since the judge's sentence stipulated that the hanging would be a surprise to him, he concludes it cannot occur on Friday.
>
> He then reasons that the surprise hanging cannot be on Thursday either, because Friday has already been eliminated and if he hasn't been hanged by Wednesday night, the hanging must occur on Thursday, making a Thursday hanging not a surprise either. By similar reasoning he concludes that the hanging can also not occur on Wednesday, Tuesday or Monday. Joyfully he retires to his cell confident that the hanging will not occur at all.
>
> The next week, the executioner knocks on the prisoner's door at noon on Wednesday — which, despite all the above, was an utter surprise to him. Everything the judge said came true."
>
> *Source: Unexpected hanging paradox - https://en.wikipedia.org*

1. Identifying the dilemma

1.1. A simplification – the One Day Unexpected Hanging Problem

First, let us named the prisoner Fred for easier understanding. To examine further, we simplify the paradox to a new One Day Unexpected Hanging Problem:

> The judge, for old age, got heavily sick for 6 days in a row and could not order the executioner to hang Fred. However, she went back to work in Sunday; and with morbid intellectuality, she told Fred that he will be hanged tomorrow. But he will still not know the day of the hanging.
>
> Fred thought: “This is a joke! She just told me that I will be hanged tomorrow! Therefore she cannot kill me since I would not be surprise! **That means I will not be hanged tomorrow!**”

And, to his surprise, Fred met his inevitable death.

Through trimming off all days but one, we verify that the paradox exists within any day, not on the length of 7 days. Also, we find that his rush conclusion **“That means I will not be hanged tomorrow!”** is at fault for his death. But why did he, and we, cling to this deadly assumption?

1.2. Additional rules and diagram:

In order to solve the problem, we need to clarify some rules that was not written in texts:

1) *Both sides must know all rules.*
2) *A “stiff” law: If the judge logic is broken (she contradicts herself), Fred will be free.*
3) *What “surprise” means:*

- *To a confident judge: “surprise” means Fred will think he will live,* since the judge is confident that Fred will be killed. This is the original problem.
- *To a lenient judge: “surprise” means Fred will think the opposite of what will actually happen.* The judge believes that there is a chance for her to lose. This is not the original situation, but it gives some interesting results.

Also, we will examine the new, easier problem and Fred’s logical thinking in this diagram:

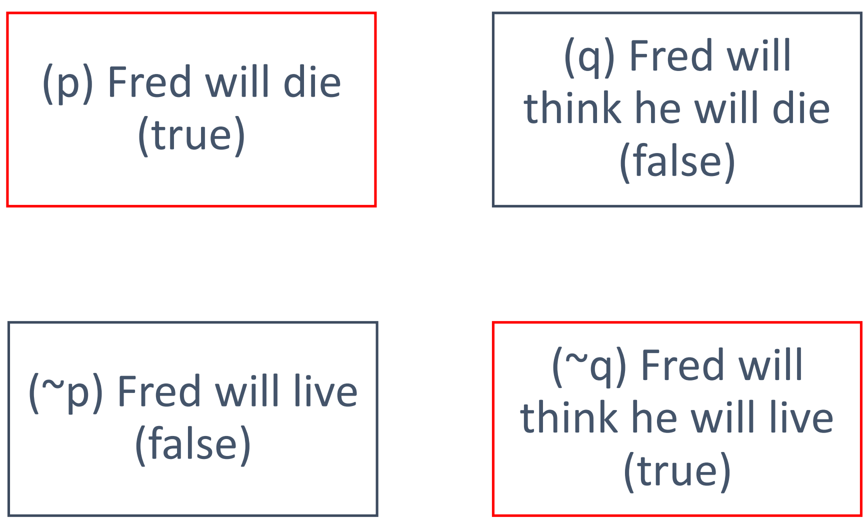


Figure 1: Events as the judge believes

In the diagram, the judge’s statements are p and ~q. We know that these two statements, to Fred and the judge, must not be violate. On the left column, p and ~p (Fred’s life) are the events that the judge control, and Fred determines events on the right column q and ~q - (Fred’s thoughts).

2. Analysis:

A quick analysis gives us a general idea of what happened: the judge gave Fred a contradiction (giving him the date of his death and announcing that he will not know the date of his death) and Fred naïvely believed that he would live simply because of that contradiction. When confronting an inconsistency, people tend to believe in thing that make them feel secure, despite whether it actually logically/physically helps them or not. That is what happened: instead of looking for the right choice and avoid the trap, Fred signed his own death warrant by believing in the thing that make him feel “safer”.

So what choice did he have?

2.1. Moore’s paradox

Before we go further, we have to take a look at Moore’s paradox:

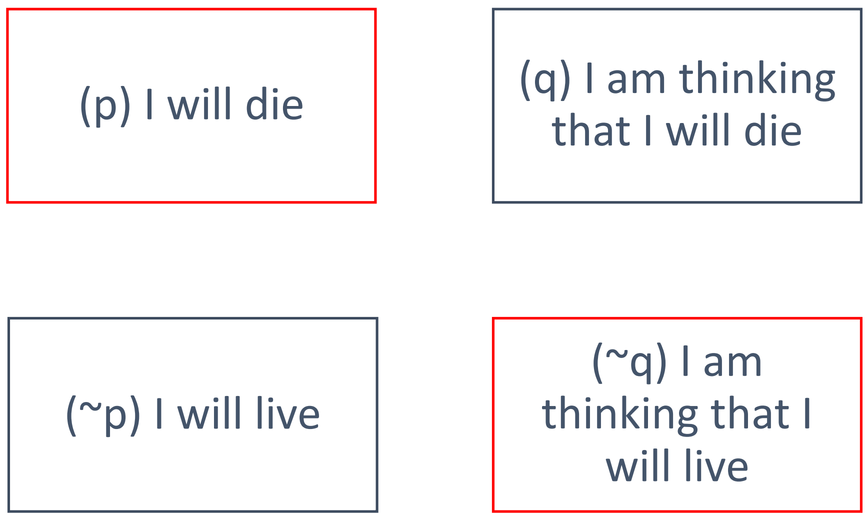


Figure 2 Events as appear to Fred

In order to avoid a cognitive dissonance, at one time, Fred can only believe in things that are in the same row; that is, (p) along with (q) or (~p) with (~q). For example, he cannot believe the judge’s statements that “he will die (p) while believing that he will live (~q)” since the judge has given him the date of his death; assuming that he is the same person with identical memory and also completely sane. This absurdity is called the “Moore’s paradox”.

We take a look at this example: “I believe A, but (I believe that) Fred believes ~A.”

- The judge: “I believe A, but (I believe that) Fred believes ~A.”

These are 2 legit statements, as the judge is not the person mentioned in the statements and did the act of believing. The judge can say whatever she wants about Fred’s belief, but Fred cannot because by making a statement he also states his belief. And if the two statements create 2 contradict belief, Fred will contradict himself:

- Fred: I believe A, but (I believe that) Fred believes ~A (?)

Fred stated that he believes A from the first statement, and immediately ignore what he just said about himself. We see the contradiction clearly when we rephrase it to: “I believe A, but (I believe that) I believes ~A. (?)” Moore’s paradox exists because we confusedly assess a first-person statement from a third-person perspective.

2.2. A quick view on the paradox

In order to avoid a cognitive dissonance, Fred can only choose statements on the same row. But by concluding that he will not be hanged, Fred enables the judge to hang him that day; in other words, thinking that he will not be hanged will kill him. But, if he knows that also (that he will be killed because of his thought), he cannot be killed!

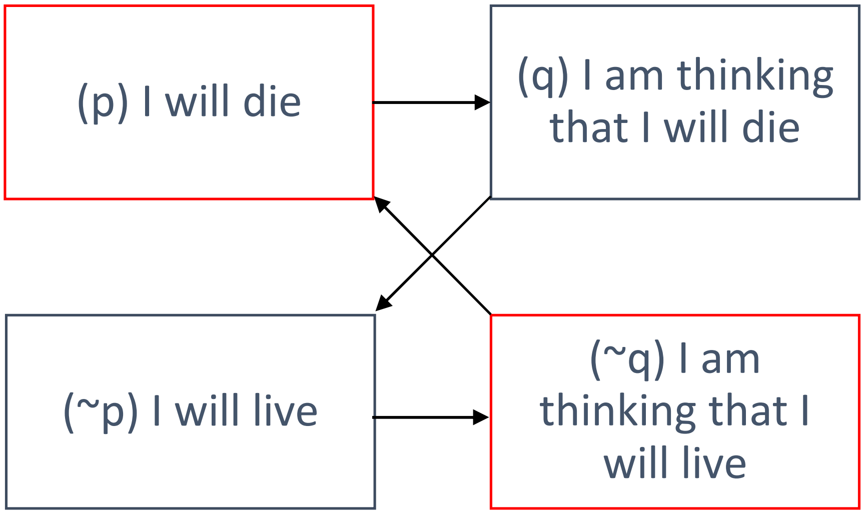


Figure 3: Events as appear to Fred: p ↔ ~p; q ↔ ~q

We have an unsolvable paradoxical loop similar to that of the liar paradox: p → q → ~p → ~q → p. However, assuming that he started reasoning from p, q → ~p is not necessary true (thinking that he will die does not mean that he will live), as I will explain in 4.1.2.

3. The solution

3.1. Some notifications

Before we go further, we have to agree on several issues:

1) What is our goal? To outwit the judge by making her statements false, or to survive?

It is nearly impossible to keep Fred alive after the judge has stated: "That Fred will be killed". In order to keep two of her statements both true, Fred will nearly always get killed no matter what, for at least one of them - the first statement - to stay true. However, Fred might outsmart the judge logically and die in honor.

2) What is the judge's goal?

The judge goal is to realize all of her words, not killing Fred. If we somehow force her in a situation when the only way for her to keep her words is to free Fred, she will have to do so.

3.2. The solution:

Knowing the judge statements, the solution is very simple: between the 2 rows, the first row is the better choice. As he has no power over what will happen in the left column (in other words, he knows p will be true

– he will die), he will have to affect the judge by changing his belief in the right column - he must think opposite of what the judge said (he must believe in q). In short, though he probably will die, the judge's words are ruined and cannot come true. The answer is Fred must think he will die.

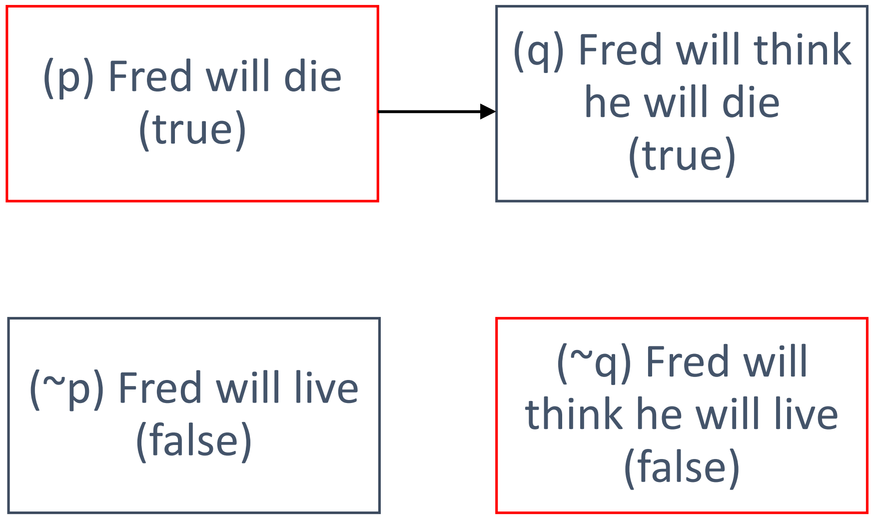


Figure 4. Fred believes in p and q

If he choose to believe in the events of the second row, as he always did, he will fell right into the judge's trap and give her what she wants: an event which she cannot decide herself (Fred's believing that he will live) is satisfied; and now she just have to finalize her words by executing Fred.

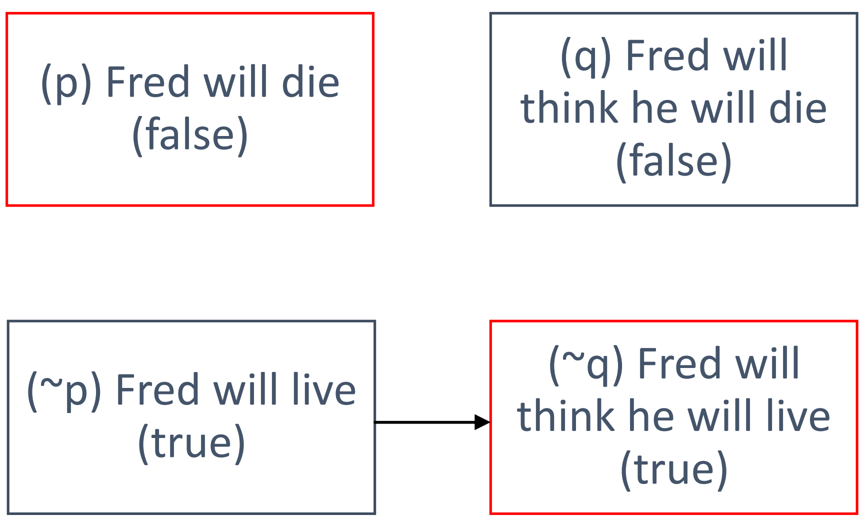


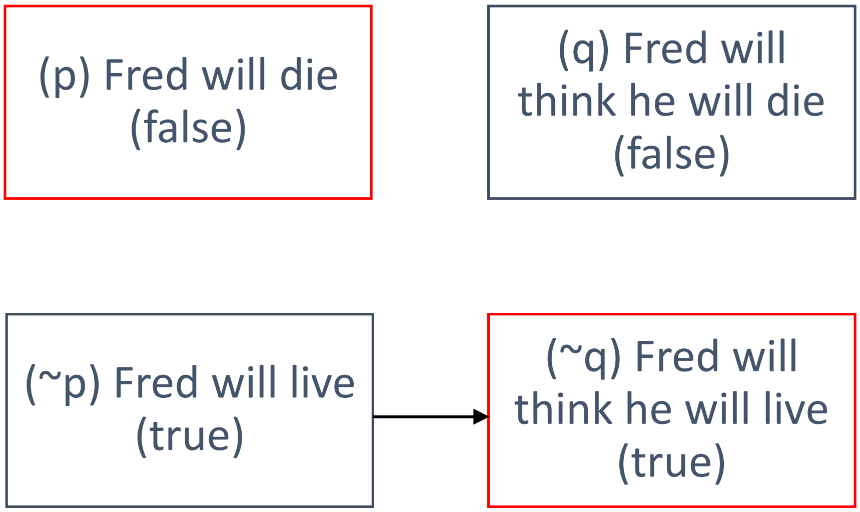
Figure 5. Fred believes in ~p and ~q

A psychological aspect of the game is how the Fred so quickly believes that he will stay alive when he discovered contradiction is discovered. By rashly concluding **"That means I will not be hanged tomorrow!"** the judge is not just free, but forced to kill him.

4. Further analysis:

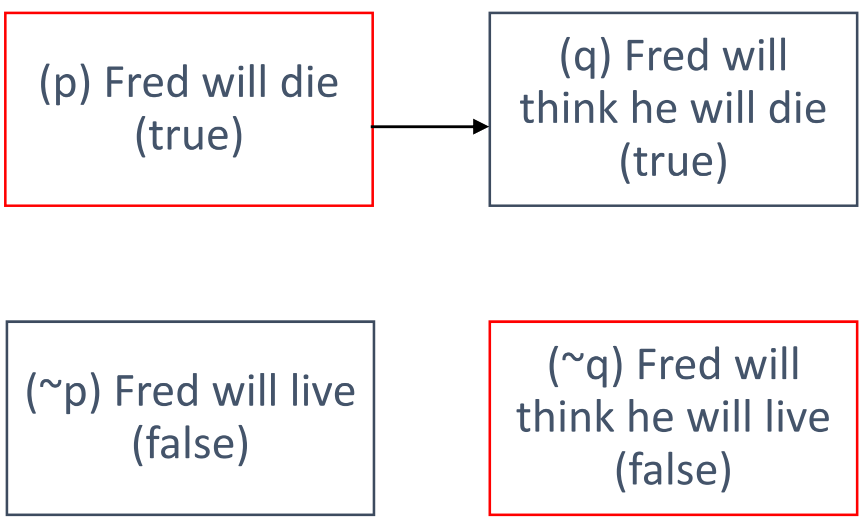
4.1. Confident judge without stiff law:

4.1.1. Fred's wrong choice:



By believing that he will survive, he gives the judge what she wants: an events which she cannot decide is satisfied. From this position, it is very easy for the judge to do her work. She does not have to satisfy Fred, and she gets to keep her words by executing him to keep both of her words, as she always did. Or she keeps Fred alive and violate both of her statements, which is unlikely.

4.1.2. Fred's better choice:



The judge is defeated: either she kills Fred and still violate her second statement (Fred is not going to be surprise), or releases him and violates both of her statements. Anyway, she will violate at least one of her statements. Of course, as 2 is worse than 1, he might still be death (as he truly believed so), but by beating the judge, at least he won the game. The judge must be wrong, one way or the other.

It is very tempting to hastily conclude that we felt in to a loop that is similar to which of the liar paradox, as I have mentioned in 2.2: If “he thinks he will die” to live, he will actually think that he will live, and die. But that is not necessarily true when we disable the “stiff law” rule in 1.1.2, which is the original problem. Thinking he will die, thus breaking the judge’s statements, does not immediately mean that he will live if the judge is not obliged to free him when he outsmarted her. For freeing Fred will cause even more of her statements wrong, there is no reason for Fred to believe that he will live by thinking that he will die.

If the stiff law rules is applied, it is true that the liar paradox’s loop might happen. One way to avoid it is for Fred to somehow convince himself that he will be dead without craving the benefit of faith, in order to keep him away from thinking that he will live. Still, this rule was not mentioned and does not seems to be implied.

4.2. Confident judge with stiff law:

4.2.1. Fred’s wrong choice:

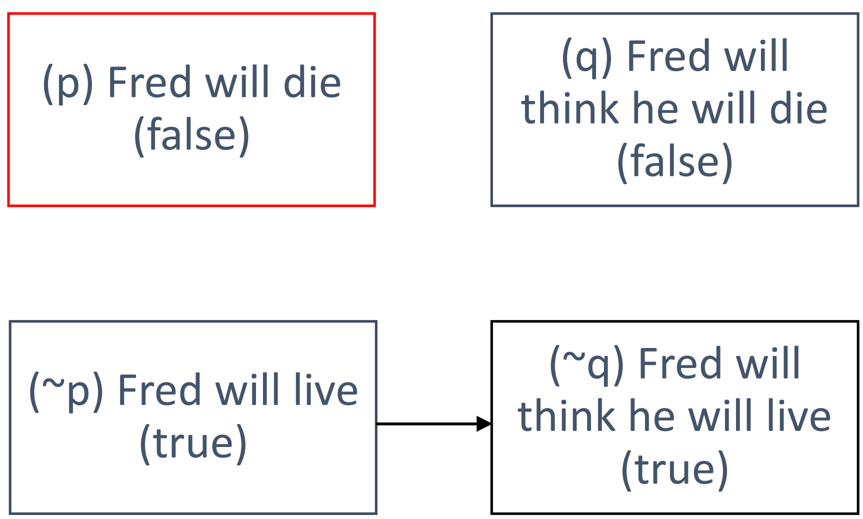
Fred’s wrong choice in this situation is the same: he will die and both of the judge’s statement are true.

4.2.2. Fred’s better choice:

The logic for Fred is the same; however, if the judge can read mind, he must not create his belief for the sake of survival, or he will actually believe in ~q and creates an infinite paradoxical loop.

4.3. Lenient judge without stiff law:

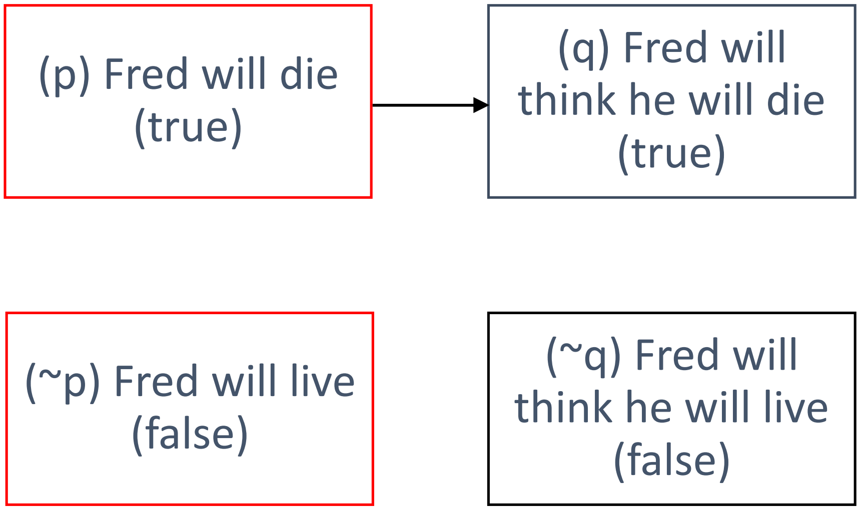
4.3.1. Fred’s wrong choice:



In the lenient judge situation, we can only know the judge’s choice for her second statements after we make our choices. Let us assume that If Fred chooses to believe in q, she will do ~p to surprise him and if Fred chooses ~q, she will do p.

The wrong choice for Fred stays the same: by choosing to believe the second row and thus ~q, Fred enabled her to realize both of her statements, as both of them need her to do p.

4.3.2. Fred’s better choice:



In this situation we encounter a more interesting situation: his chance of survival increase. Either she kills Fred and violate her second statements (he will not be surprise), or she keeps him alive and violate her first statements. Whatever she does, she will violate one of her statements. It became a 50-50 situation.

4.4. Lenient judge with stiff law:

4.4.1. Fred’s wrong choice:

Fred’s wrong choice in this situation is the same: he will die and both of the judge’s statement are true.

4.4.2. Fred’s better choice:

The logic for Fred is the same; however, he must not create his belief for the sake of survival, or he will actually believe in ~q and creates an infinite paradoxical loop.

5. How the judge increase her chance:

In this section we come back to our original 7-day Unexpected Hanging Problem. As we know that the only way to increase Fred’s chance of winning is to believe that he will die, the judge will add several more days for him to predict. For Fred can only believe in one of the days, the more days there are, the slimmer the chance of Fred to win and/or survive will become. Depends on the number of days, we have a winning probability of

$$\frac{1}{number\ of\ days}$$

To prevent him from cheating, the judge probably will have him state the day that he think he will die. But Fred cannot believe the judge also, for if she knows the day she will avoid it. So they will have to ask a referee or make some deals that are irrelevant to the paradox.

References

1. Timothy Y.Chow, The surprise Examination or Unexpected Hanging Paradox,Amer.Math.Monthly 105(1998)
2. Linville, Kent & Ring, Merrill. "Moore's Paradox Revisited". *Synthese* **87** (2): 295–309. doi:10.1007/BF00485405
3. Bovens, Luc (1995). "'P and I Will Believe that not-P': Diachronic Constraints on Rational Belief". *Mind* **104** (416): 737–760. doi:10.1093/mind/104.416.737
4.