The Asylum of Doctor Tarr and Professor Fether

Adapted from: *The Lady or the Tiger?* by Raymond Smullyan
(With apologies to Edgar Allan Poe)

Inspector Craig of Scotland Yard was called to investigate an insane asylum where it was suspected that something was wrong. The asylum was run by two doctors named Doctor Tarr and Professor Fether. There were other doctors on the staff as well. The only inhabitants were patients and doctors – the doctors constituted the entire staff.

Each inhabitant of the asylum, patient or doctor, was either sane or insane. Moreover, the sane ones were *totally* sane and a hundred percent accurate in all their beliefs; all true propositions they knew to be true and all false propositions they knew to be false. The insane ones were totally inaccurate in their beliefs; all true propositions they believed to be false and all false propositions they believed to be true. It is to be assumed also that all the inhabitants were always honest – whatever they said, they really believed.

While Craig was waiting to interview Doctor Tarr and Professor Fether, he spoke separately to two inhabitants whose last names were Jones and Smith.

"Tell me," Craig asked Jones, "what do you know about Mr. Smith?"

"You should call him *Doctor* Smith," replied Jones. "He is a doctor on our staff."

Some time later, Craig met Smith and asked, "What do you know about Jones? Is he sane or insane?"

"He is insane," replied Smith.

The inspector mulled over the situation for a while, and then had private interviews with Doctor Tarr and Professor Fether. Here is the interview with Doctor Tarr:

Craig: Tell me, Doctor Tarr, are all the doctors in this asylum sane?

Tarr: Of course they are!

Craig: What about the patients? Are they all insane?

Tarr: At least one of them is.

The second answer struck Craig as a surprisingly modest claim! Why was Doctor Tarr being so cautious?

Craig then had his interview with Professor Fether, which went as follows:

Craig: Surely it is true that all the patients in this asylum are insane, isn't it?

Fether: Of course it is true! What kind of asylum do you think we are running?

Craig: What about the doctors? Are they all sane?

Fether: At least one of them is.

Craig: What about Doctor Tarr? Is he sane?

Fether: Of course he is! How dare you ask me such a question?

At this point, Craig realized the full horror of the situation! What was it?

After you have solved the problem and proved your answer, consider the following two variations.

1. Was the last question really necessary? That is, could Craig have deduced the same conclusion without knowing that Fether thinks that Tarr is sane?

2. When Craig asked Fether, "What about the doctors? Are they all sane?", suppose Fether had just replied, "At least Doctor Tarr is sane." Would the conclusion have been the same?

On the one hand, if Doctor Tarr is sane, it certainly implies that at least one doctor is sane. On the other hand, the negation of the statement "Doctor Tarr is sane" is much weaker than the negation of the sentence "At least one doctor is sane." Resolve this dilemma.

Solution

Begin with the statements of Jones and Smith. Jones thinks Smith is a doctor, and Smith thinks Jones is insane. If Jones is sane, then Smith is a doctor; but Smith incorrectly thinks Jones is insane, hence Smith is an insane doctor. If Jones is insane, then his belief that Smith is a doctor is incorrect, i.e. Smith is a patient; but Smith correctly believes that Jones is insane, so Smith is a sane patient. Thus Smith is either an insane doctor or a sane patient, and there is indeed something wrong at the asylum.

The same conclusion can be drawn by making assumptions about Smith's sanity: if Smith is sane, then Jones is insane, and thus Smith is a patient; if Smith is insane, then Jones is sane and Smith is a doctor. One could also make assumptions about Smith's status: if Smith is a doctor, then Jones is sane and Smith is insane; if Smith is a patient, then Jones is insane and Smith is sane. Interestingly, no conclusions at all can be drawn about Jones at this point.

Now consider the interviews with Tarr and Fether. Tarr thinks all the doctors are sane, and Fether thinks that all the patients are insane. We already know that these statements cannot both be true, so Tarr and Fether cannot both be sane. Also, Fether thinks that Tarr is sane, so if Tarr were sane Fether would also be sane. Thus Tarr is insane, and since Fether thinks that Tarr is sane, Fether must also be insane.

Tarr incorrectly thinks that at least one patient is insane, and therefore all the patients are sane. Similarly, Fether thinks that at least one doctor is sane, so all the doctors are insane.

Variation 1

The answer (to the second question) in the first variation is "Almost, but not quite." One can argue as follows. Suppose Jones is sane, i.e. Smith is an insane doctor. Then Tarr must be insane, since he thinks all the doctors are sane. Thus his belief that at least one patient is insane is wrong, so all the patients are sane. Thus Fether is insane, since he thinks that all the patients are insane. So his belief that at least one doctor is sane is wrong.

On the other hand, if Jones is insane, i.e. Smith is a sane patient, we conclude that Fether is insane since he believes all patients are insane. Thus his belief that at least one doctor is sane is wrong, so all the doctors are insane, and in particular Tarr is insane. (We do not even need to use Tarr's belief that all the doctors are sane!) So Tarr's belief that at least one patient is insane is wrong.

There is, however, a subtle fallacy in the first half of the argument: suppose there are no patients! (The existence of at least one patient was strongly implied but never explicitly stated in the problem; existence of a patient can be deduced from the problem as stated, including the last question, but not without the last question. Indeed, in the problem as stated, if Smith is an insane doctor, then Jones is sane, hence a patient, so either Smith or Jones is a patient.) If there are no patients, then the statements "All the patients are sane" and "All the patients are insane" are both technically correct. Thus, if there are no patients,

Fether is sane, and it could be that all the doctors except Tarr and Smith are sane. In this case, Jones would be a sane doctor. (Note that the same problem does not occur in the second half of the argument, since we know from the outset that there is at least one doctor, in fact more than two.)

There is one other technical problem with the way the first variation is stated: it asks not "can it be deduced . . ." but "could Craig have deduced . . ." We have no way of knowing whether Craig himself could have made the deduction.

Variation 2

In the second variation, one can still conclude that Tarr and Fether are both insane, and that all the patients are sane. However, it cannot be concluded that all the doctors are insane. The dilemma can be resolved as follows. Consider

Statement 1: If Doctor Tarr is sane, then at least one doctor is sane.

Statement 1 is certainly a true implication. However, Statement 1 is not really the relevant one for this problem; instead, we really need to consider

Statement 2: If Fether believes that Doctor Tarr is sane, then he believes that at least one doctor is sane.

This is *not* necessarily a true implication. If Fether were sane, then Statement 2 would be true. However, if Fether is insane (as we know to be the case through other information), he believes the true Statement 1 to be false; thus we cannot conclude that he thinks at least one doctor is sane (he may or may not.)

However, this analysis leads to another apparent contradiction. If Fether is insane and believes the true Statement 1 to be false, the way truth of an implication is defined he must believe Statement 1P is true and Statement 1Q is false:

Statement 1P: Doctor Tarr is sane.
Statement 1Q: At least one doctor is sane.

Thus Statement 1Q must actually be true! This argument can be made also in the problem as stated to obtain a contradiction.

This apparent contradiction can be explained away by saying that if Fether is insane, he does not believe that if Statement 1 is false Statement 1Q must also be false. However, further thought reveals many other logical contradictions in the problem as stated which make the entire problem ridiculous. For example, if Jones is truly insane, he not only believes that Smith is a doctor, but also believes his name is not Smith; instead he believes that most if not all the other inhabitants are named Smith. In particular, he believes both Tarr and Fether are named Smith, so he could not have made the response he did to Craig.

There are other logical difficulties with the whole problem. One interesting question, raised by a student in a discussion, is whether it even makes sense at all to ask whether a fictitious statement is true or false.