# The Paradox of Infallibility

*Daniel Rönnedal*

*Stockholm University*

*Abstract*

This paper discusses a new paradox, the paradox of infallibility. Let us define infallibility in the following way: (Def I) *t* is infallible if and only if (iff) everything *t* believes is true, where *t* is any term. (Def I) entails the following proposition: (I) It is necessary that for every individual *x*, *x* is infallible iff every proposition *x* believes is true. However, (I) seems to be inconsistent with the following proposition (P): It is possible that there is some individual who believes exactly one proposition, namely that she is not infallible. So, it seems to be the case that either (I) or (P) must be false. Yet, (I) is simply a consequence of (Def I) and (P) clearly seems to be true. This is the puzzle. I discuss five possible solutions to the problem and mention some arguments for and against these solutions.

*Keywords*: Paradoxes, Infallibility, Epistemic paradoxes, Dialetheism, Propositional quantifiers.

## 1. Introduction

In this paper, I introduce a new paradox, the paradox of infallibility. Intuitively, this puzzle can be formulated in the following way. Assume that someone is infallible if and only if (iff) everything she believes is true and that there is an individual who believes exactly one proposition, namely the proposition that she is not infallible. Suppose that this individual is infallible. Then everything she believes is true. Hence, she is not infallible, since she believes that she is not infallible. So, if she is infallible, she is not infallible. Suppose that she is not infallible. Then everything she believes is true, since the only proposition she believes is the proposition that she is not infallible. Accordingly, she is infallible. Consequently, if she is not infallible, she is infallible. It follows that she is infallible iff she is not infallible. But this is clearly a contradiction. Hence, it cannot be the case that someone is infallible iff everything she believes is true and that it is possible that there is an individual who believes exactly one proposition, namely the proposition that she is not infallible.

I will now describe the paradox more carefully. I will use the following definition of infallibility:

(Def I) $It =_{df} \forall A(BtA \supset A)$, where *t* is any term. For every *t*, *t* is infallible iff for every (proposition) A, if *t* believes that A, then A (is true).

*Argumenta* 8, 1 (2022): 189-197
ISSN 2465-2334
DOI 10.14275/2465-2334/20210.ron

© 2021 Daniel Rönnedal

(Def I) is a metalinguistic definition. This means that '$\forall A(B\mathit{t}A \supset A)$' can be replaced by '$\mathit{It}$' in any context whatsoever, and vice versa; '$\mathit{It}$' is simply an abbreviation of '$\forall A(B\mathit{t}A \supset A)$'. In (Def I), '$\mathit{I}$' is a predicate, '$\mathit{t}$' is a term, '$\forall$' is a propositional quantifier, 'B' is a doxastic operator, and 'A' is a propositional variable. All these symbols are used in a standard way. From (Def I) we can derive the following proposition:

(I) $\Box\forall x(\mathit{Ix} \equiv \forall A(B\mathit{x}A \supset A))$. It is necessary that for every (individual) *x*: *x* is infallible iff for every (proposition) A, if *x* believes that A then A.

Note that the first quantifier in (I) varies over individuals while the second varies over propositions or sentences. '$\Box$' is the standard (absolute) necessity operator. Hence, '$\Box A$' is true in a possible world iff 'A' is true in every possible world.[1] It should be obvious that (I) follows from (Def I), since if we replace '$\mathit{Ix}$' by '$\forall A(B\mathit{x}A \supset A)$' in (I) we obtain '$\Box\forall x(\forall A(B\mathit{x}A \supset A) \equiv \forall A(B\mathit{x}A \supset A))$', which obviously is valid. The inconsistency argument (see below) shows (or seems to show) that (I) is incompatible with the following proposition:

(P) $\Diamond\exists x(B\mathit{x}{\sim}\mathit{Ix} \;\&\; \forall A(B\mathit{x}A \supset \Box(A \equiv {\sim}\mathit{Ix})))$. It is possible that there is some individual who believes exactly one proposition, namely that she is not infallible. More precisely, the informal reading of (P) is as follows: It is possible that there is some individual *x* such that *x* believes that it is not the case that *x* is infallible and for every proposition A, if *x* believes that A, then it is necessary that A iff it is not the case that *x* is infallible.

Again, note that the first quantifier in (P) varies over individuals while the second varies over propositions or sentences. '$\exists$' is a standard propositional quantifier and '$\Diamond$' is the standard (absolute) possibility operator. Accordingly, '$\Diamond A$' is true in a possible world iff 'A' is true in some possible world. '$\forall A(B\mathit{x}A \supset \Box(A \equiv {\sim}\mathit{Ix}))$' does not say that *x* has only one belief, but it says that if *x* believes A then A is necessarily equivalent with (and so identical to) the proposition that *x* is not infallible. Hence, '$\Diamond\exists x(B\mathit{x}{\sim}\mathit{Ix} \;\&\; \forall A(B\mathit{x}A \supset \Box(A \equiv {\sim}\mathit{Ix})))$' is a reasonable symbolisation of the proposition that it is possible that there is some individual who believes exactly one proposition, namely that she is not infallible.

Furthermore, note that (P) only says that it is *possible* that there is an individual of a certain kind. It does not claim that there (*actually*) is an individual of this type. Probably, it is not the case that there is some (actual) individual who believes exactly one proposition, namely that she is not infallible. Still, this does not entail that (P) is false. In other words, (P) is compatible with the following formula: '${\sim}\exists x(B\mathit{x}{\sim}\mathit{Ix} \;\&\; \forall A(B\mathit{x}A \supset \Box(A \equiv {\sim}\mathit{Ix})))$'.

I will now show that {(I), (P)} seems to be inconsistent. To establish this, I will assume that (I) and (P) are true in some possible world $w_0$ and derive a contradiction. I will call this derivation 'the inconsistency argument'. '$\Diamond$', '$\exists$', '$\Box$', '$\forall$' and '${\sim}\forall$' in the deduction below are standard derivation rules. 'PL' means that the step follows by ordinary propositional reasoning. Intuitively, 'A, *w*' says that 'A' is true in the possible world *w*. Here is the derivation:

[1] For more on modal logic, see, for example, Blackburn, De Rijke, Venema 2001, Chellas 1980 and Hughes and Cresswell 1968.

**The Inconsistency Argument**

| | |
|---|---|
| (1) $\Box\forall x(Ix \equiv \forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \mathrm{A})), w_0$ | [Assumption] |
| (2) $\Diamond\exists x(\mathrm{B}x{\sim}Ix \;\&\; \forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv {\sim}Ix))), w_0$ | [Assumption] |
| (3) $\exists x(\mathrm{B}x{\sim}Ix \;\&\; \forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv {\sim}Ix))), w_1$ | [2, $\Diamond$] |
| (4) $\mathrm{B}c{\sim}Ic \;\&\; \forall \mathrm{A}(\mathrm{B}c\mathrm{A} \supset \Box(\mathrm{A} \equiv {\sim}Ic)), w_1$ | [3, $\exists$] |
| (5) $\mathrm{B}c{\sim}Ic, w_1$ | [4, PL] |
| (6) $\forall \mathrm{A}(\mathrm{B}c\mathrm{A} \supset \Box(\mathrm{A} \equiv {\sim}Ic)), w_1$ | [4, PL] |
| (7) $\forall x(Ix \equiv \forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \mathrm{A})), w_1$ | [1, $\Box$] |
| (8) $Ic \equiv \forall \mathrm{A}(\mathrm{B}c\mathrm{A} \supset \mathrm{A}), w_1$ | [7, $\forall$] |
| (9) $Ic, w_1$ | [Assumption] |
| (10) $\forall \mathrm{A}(\mathrm{B}c\mathrm{A} \supset \mathrm{A}), w_1$ | [8, 9, PL] |
| (11) $\mathrm{B}c{\sim}Ic \supset {\sim}Ic, w_1$ | [10, $\forall$] |
| (12) ${\sim}Ic, w_1$ | [5, 11, PL] |
| (13) $Ic \;\&\; {\sim}Ic, w_1$ | [9, 12, PL] |
| (14) ${\sim}Ic, w_1$ | [Assumption] |
| (15) ${\sim}\forall \mathrm{A}(\mathrm{B}c\mathrm{A} \supset \mathrm{A}), w_1$ | [8, 14, PL] |
| (16) ${\sim}(\mathrm{B}c\mathrm{X} \supset \mathrm{X}), w_1$ | [15, ${\sim}\forall$] |
| (17) $\mathrm{B}c\mathrm{X}, w_1$ | [16, PL] |
| (18) ${\sim}\mathrm{X}, w_1$ | [16, PL] |
| (19) $\mathrm{B}c\mathrm{X} \supset \Box(\mathrm{X} \equiv {\sim}Ic), w_1$ | [6, $\forall$] |
| (20) $\Box(\mathrm{X} \equiv {\sim}Ic), w_1$ | [17, 19, PL] |
| (21) $\mathrm{X} \equiv {\sim}Ic, w_1$ | [20, $\Box$] |
| (22) ${\sim}{\sim}Ic, w_1$ | [18, 21, PL] |
| (23) ${\sim}Ic \;\&\; {\sim}{\sim}Ic, w_1$ | [14, 22, PL] |

The individual *c* is either infallible or not in $w_1$. At step (9), we assume that *c* is infallible in $w_1$. This leads to a contradiction at step (13). At step (14), we assume that *c* is not infallible in $w_1$. This also leads to a contradiction at step (23). Accordingly, both assumptions lead to a contradiction. Hence, (1) and (2) cannot both be true in $w_0$. Since $w_0$ was arbitrary, we conclude that {(I), (P)} is inconsistent. The inconsistency argument clearly seems to be valid. So, either (1) = (I) or (2) = (P) (or both) must be false. Yet, both (I) and (P) appear to be true. (I) follows from (Def I) and (P) is intuitively plausible. Furthermore, the following argument supports (P). It is conceivable that there is some individual who believes exactly one proposition, namely that she is not infallible. Hence, it is (at least prima facie) reasonable to assume that it is possible that there is some individual of the required kind. This is the paradox of infallibility.[2]

[2] Two anonymous reviewers have challenged this claim. According to the first, it is not so obvious that it is conceivable that there is some individual who believes exactly one proposition, namely that she is not infallible. Such beliefs could not count as rational, according to the reviewer. Similarly, someone could assert 'this sentence is false', but couldn't be warranted to assert it. According to the second, the inconsistency argument shows that the existence of the 'modest believer' (i.e. a subject who believes just one thing, that is, that she is not infallible) is logically impossible. I do not assume that conceivability entails possibility. So, I would still say that the scenario is conceivable, but that this fact does not entail that it is possible. Even if the existence of the 'modest believer' should turn out to be impossible such a believer might be conceivable. If we assume that conceivability entails possibility, we should instead say that the scenario seems to be

Before I turn to the discussion of the possible solutions, I would like to briefly address one possible objection to the way the paradox of infallibility is formulated in this paper (this objection was raised by an anonymous reviewer of the paper). According to this objection, the definition of infallibility plays very little role in generating the puzzle. (P) could be formulated without reference to infallibility, as 'it is possible that there is some individual who believes just one proposition: that something she believes is not true', and still it would have paradoxical consequences. Let us call this sentence (P′). The real problem does not concern the concept of infallibility but the self-referential nature of (P′).

I am in general sympathetic to this kind of view and to the claim that the real problem does not concern the concept of infallibility. In fact, according to the solution that seems most plausible to me, solution 5 below, we can solve the paradox without changing our definition of infallibility. Furthermore, I agree that (P′) is problematic and that (P′) is similar to (P). Accordingly, it is possible that the puzzles generated by (P′) and (P) have similar 'solutions'. Therefore, (P′) (and its paradoxical nature) is interesting on its own. However, (P′) and (P) do not say exactly the same thing and (I) is an essential assumption in our inconsistency argument. Without this assumption we cannot derive a contradiction. Therefore, the paradox of infallibility, as it is formulated in this paper, is not the exact same paradox as the paradox generated by (P′). The paradox of infallibility should be interesting to anyone who philosophises about the concept of infallibility,[3] not only to anyone who philosophises about paradoxes. We want to know if and how we can solve various paradoxes, but we also want to know if the concept of infallibility is consistent or not. The paradox of infallibility is a potential threat to anyone who thinks that the concept of infallibility is consistent; (P′) is not, at least not in itself. The discussion of the paradox of infallibility, as formulated in this paper, should therefore not be replaced by a discussion of (P′) and its paradoxical nature.

Consider, for example, the debate between a classical theist and an atheist. The theist wants to claim that God is infallible. The atheist might respond that the paradox of infallibility shows that the concept of infallibility is inconsistent and that God therefore cannot be infallible. The theist might perhaps respond in the same way as the anonymous reviewer. She might claim that the concept of infallibility plays very little role in generating the puzzle and that the concept of infallibility is consistent. Or again, consider the discussion between an ideal observer theorist in metaethics and a critic. The ideal observer theorist might want to assert that an ideal observer is infallible. The critic might insist that the ideal observer theory is wrong since the paradox of infallibility shows that the concept of infallibility is inconsistent. The ideal observer theorist might perhaps respond in the same way as the anonymous reviewer and try to show that the paradox of infallibility does not establish that the concept of infallibility is inconsistent, etc.

conceivable (even though, in fact, it is not, because it is impossible). I agree that someone who believes in a contradiction cannot be perfectly rational. However, I am inclined to think that it is still possible for someone to believe in contradictions. So, this is not necessarily a problem for the conceivability argument.

[3] This might, for example, include some epistemologists, some (doxastic) logicians, some philosophers of religion and some moral philosophers.

If we reformulate the paradox and drop the concept of infallibility, we cannot understand these kinds of debates.

## 2. Possible Solutions

Is it possible to solve the paradox of infallibility? In this section, I will consider five conceivable solutions. Personally, I am inclined to believe that the last suggestion is the most promising. However, no proposal is without problems.

### 2.1 Solution 1

According to the first solution, we should accept dialetheism. According to this theory, there are sentences that are both true and false. If we accept this idea, we might also accept the proposition that it is possible to deduce a contradiction from {(I), (P)} even though both (I) and (P) are true. This might be perfectly reasonable if there are true contradictions.

Still, there are problems with this solution. Dialetheism is dubitable and even if the theory were true, it is not obvious that *every* contradiction is genuine (true). Consequently, even a dialetheist might think that the paradox of infallibility is problematic. Therefore, it seems unlikely that this solution should turn out to be the most plausible overall.[4]

### 2.2 Solution 2

According to the second solution, we should reject (P) because it is impossible that there is someone who believes that she is not infallible. '$\Diamond\exists x(\mathrm{B}x{\sim}Ix \,\&\, \forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv {\sim}Ix)))$' entails '$\Diamond\exists x\mathrm{B}x{\sim}Ix$' (this is easy to see since '$\Diamond\exists x(\mathrm{A} \,\&\, \mathrm{B})$' entails '$\Diamond\exists x\mathrm{A}$'). So, if '${\sim}\Diamond\exists x\mathrm{B}x{\sim}Ix$' is true (valid), then '${\sim}\Diamond\exists x(\mathrm{B}x{\sim}Ix \,\&\, \forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv {\sim}Ix)))$' is true (valid) (again, the proof is easy and can be left to the reader). Accordingly, if we can establish that '${\sim}\Diamond\exists x\mathrm{B}x{\sim}Ix$' is true (valid), we may conclude that (P) is false (necessarily false).

According to standard doxastic logic, '$\mathrm{B}c\mathrm{A}$' is true in a possible world $w$ iff 'A' is true in every possible world that is doxastically accessible from $w$ for $c$. Furthermore, many doxastic logicians assume that for every individual $c$ and for every possible world $w$ there is a possible world $w'$ such that $w'$ is doxastically accessible from $w$ for $c$, and that if a possible world $w'$ is doxastically accessible from a possible world $w$ for an individual $c$, then $w'$ is doxastically accessible from $w'$ for $c$.[5] Suppose that this is correct. Then we can show that '${\sim}\Diamond\exists x\mathrm{B}x{\sim}Ix$' is valid in the following way. Assume that '${\sim}\Diamond\exists x\mathrm{B}x{\sim}Ix$' is not true in some possible world $w_0$. Then '$\Diamond\exists x\mathrm{B}x{\sim}Ix$' is true in $w_0$. Hence, '$\exists x\mathrm{B}x{\sim}Ix$' is true in some possible world $w_1$. Accordingly, '$\mathrm{B}c{\sim}Ic$' is true in $w_1$ (where $c$ is some arbitrary individual). By assumption, there is a possible world $w_2$ that is doxastically accessible from $w_1$ for $c$. Consequently, '${\sim}Ic$' is true in $w_2$. By definition, '${\sim}Ic$' is equivalent with '${\sim}\forall \mathrm{A}(\mathrm{B}c\mathrm{A} \supset \mathrm{A})$'. Hence, '${\sim}\forall \mathrm{A}(\mathrm{B}c\mathrm{A} \supset \mathrm{A})$' is true in $w_2$. It follows that '${\sim}(\mathrm{B}c\mathrm{X} \supset \mathrm{X})$' is true in $w_2$ (for some arbitrary X). Therefore, '$\mathrm{B}c\mathrm{X}$' is true in $w_2$ and 'X' is false in $w_2$. By

[4] For more on dialetheism, see, for example, Priest, Berto and Weber 2018.
[5] For more on doxastic logic, see, for example, Fagin, Halpern, Moses and Vardi 1995 and Meyer and van der Hoek 1995.

assumption, $w_2$ is doxastically accessible from $w_2$ for *c*. Hence, 'X' is true in $w_2$. But this is absurd. It follows that our original hypothesis cannot be true. In other words, it is not possible that there is someone who believes that she is not infallible. It follows that (P) is false (and indeed necessarily false). This solves the paradox of infallibility.

The problem with this solution is that standard doxastic logic only seems to make sense if we assume that we are dealing with perfectly rational individuals. According to orthodox doxastic logic, it is necessary that every individual believes every logical truth. Furthermore, according to the assumptions above, it is necessary that no individual has any inconsistent beliefs and it is necessary that every individual believes that everything she believes is true. It seems very implausible to assume that this holds for *every* individual. So, if we assume that we are quantifying over *every* individual in (P) and not only over perfectly rational agents, it clearly seems to be possible that there is someone who believes that she is not infallible. In fact, there are probably many (*actual*) persons who believe this. And if there *is* someone who believes this, then certainly it is *possible* that there is some individual of this kind. Consequently, our second solution to the paradox of infallibility is quite problematic.[6]

## 2.3 Solution 3

According to the third solution, we should reject (P) because it is impossible that there is someone who believes only one proposition, namely the proposition that she is not infallible. '$\Diamond\exists x(Bx{\sim}Ix \,\&\, \forall A(BxA \supset \Box(A \equiv {\sim}Ix)))$' entails '$\Diamond\exists x\forall A(BxA \supset \Box(A \equiv {\sim}Ix))$'. Therefore, if '${\sim}\Diamond\exists x\forall A(BxA \supset \Box(A \equiv {\sim}Ix))$' is true (valid), then '${\sim}\Diamond\exists x(Bx{\sim}Ix \,\&\, \forall A(BxA \supset \Box(A \equiv {\sim}Ix)))$' is true (valid). Hence, if we can show that '${\sim}\Diamond\exists x\forall A(BxA \supset \Box(A \equiv {\sim}Ix))$' is true (valid), we may conclude that (P) is false (necessarily false). Why should we believe that it is impossible that there is someone who believes only one proposition, namely the proposition that she is not infallible? Well, according to this solution, we should believe this because it is impossible that there is someone who believes only one proposition, period. We can only have beliefs if we believe many things. To believe anything at all we need a whole web of beliefs. If this is true, we should reject (P). Hence, we can avoid the paradox of infallibility.

Is it true that it is impossible to believe only one proposition? This seems to depend on what we mean by 'impossible'. Perhaps it is historically (temporally) and naturally impossible. But the problem with this solution is that we are not (primarily) interested in these kinds of possibilities in this paper. (P) is supposed to be speaking about logical or metaphysical possibility. And it certainly seems to be logically or metaphysically possible that there is someone who believes on-

[6] An anonymous reviewer has suggested that solution 2 is clearly absurd and that the proper way to reject P is to claim that it is impossible that there is someone who believes that she is not infallible AND that this is her only belief. The falsity of this proposition can be argued on the basis of its self-referential structure, its similarity to the Liar (a belief that is true when false and vice versa). I tend to agree with the general sentiment of this view. If the inconsistency argument is sound (and we assume (Def I)), we must reject (P). The solution that seems most promising to me, solution 5, is similar to the solution suggested by the reviewer. However, solution 5 does not entail that (P) is false.

ly one proposition, even though it is perhaps not historically or naturally possible. If this is the case, we cannot use the third solution to solve the paradox of infallibility.

### 2.4 Solution 4

According to the fourth solution, we should reject the definition of infallibility (Def I) that we use to derive (I), and if (Def I) is not true (or correct), we have no reason to believe that (I) is true. Therefore, we can also reject (I). Hence, this solves the paradox.

The problem with this solution is that it is difficult to come up with some other definition of infallibility that is reasonable and that does not lead to similar problems. Let us consider one alternative attempt. Instead of (Def I) we should use the following definition of infallibility:

(Def I′) $It =_{df} \Box\forall A(BtA \supset A)$, where $t$ is any term. For every $t$, $t$ is infallible iff it is *necessary* that for every (proposition) A, if $t$ believes that A, then A (is true).

According to this definition, no one is infallible if it is *possible* that something she believes is false; it is not enough that everything she believes is true. (Def I′) does not entail (I), but it does entail something similar, namely (I′):

(I′) $\Box\forall x(Ix \equiv \Box\forall A(BxA \supset A))$. It is necessary that for every (individual) $x$: $x$ is infallible iff it is *necessary* that for every (proposition) A, if $x$ believes that A then A.

However, if we try to replace (I) by (I′) in the inconsistency argument, it breaks down. So, we cannot use this deduction to show that {(I′), (P)} is inconsistent. Consequently, if we use (Def I′) instead of (Def I) to define the concept of infallibility, we can solve the paradox of infallibility. Intuitively, (Def I′) is even more plausible than (Def I). Hence, the fourth solution seems to be one of the more plausible. Nevertheless, it is not unproblematic, for we can show that (I′) is inconsistent with the following alternative to (P):

(P″) $\Diamond\exists x(Bx{\sim}Ix \,\&\, \Box\forall A(BxA \supset \Box(A \equiv {\sim}Ix)))$. It is possible that there is some individual $x$ such that $x$ believes that it is not the case that $x$ is infallible and it is necessary that for every proposition A, if $x$ believes that A, then it is necessary that A iff it is not the case that $x$ is infallible.

That is, we can prove that {(I′), (P″)} is inconsistent (the argument for this is similar to the inconsistency argument; see above). And (P″) seems to be true. So, even though we can use (I′) to avoid our original problem, we can derive a contradiction from {(I′), (P″)}. Therefore, it is doubtful that this is the best solution to the paradox.[7]

[7] An anonymous reviewer has a strong feeling that the paradox of infallibility has nothing specifically to do with the definition of infallibility. It has to do with truth, and thus indirectly with infallibility defined in terms of true beliefs (see the introduction). I am inclined to believe that this is true, or approximately true. According to solution 5, which seems most promising to me, the paradox of infallibility can be solved without changing the definition of the concept of infallibility in this paper. This solution has to do with the way we should understand propositional quantifiers (and therefore also with self-reference). However, I do not think we should take this for granted and assume that the concept of infallibility is consistent without any discussion.

### 2.5 Solution 5

According to the fifth and last solution, we should reject the inconsistency argument. It is not necessarily anything wrong with (Def I), (I) or (P), but the deduction is not valid. If the argument for the conclusion that {(I), (P)} is inconsistent fails, then of course we have solved the paradox of infallibility.

But what is wrong with the inconsistency argument? It clearly seems to be valid. The problematic step, according to this solution, is step (11). The universal quantifier cannot be instantiated with any sentence whatsoever. The quantifier in step (10) is a propositional quantifier and in step (11) we have instantiated A with '∼*Ic*'. However, '∼*Ic*' is simply an abbreviation of '∼∀A(B*c*A ⊃ A)' and this sentence includes a propositional quantifier. It is a well-known fact that it is problematic to allow universally quantified sentences to be instantiated with universally quantified sentences when we use ∀-elimination for propositional quantifiers. If we allow such instances, several problematic consequences follow. Consider, for example, the following difficulty. Intuitively, '∀XA' says 'For all propositions X: A'. '∀XA' is true if and only if 'A[B/X]' for every proposition B, where 'A[B/X]' is the result of replacing all free occurrences of the propositional variable 'X' in A by B. Now, let A = ∀XX and assume that our substitution-instances can include any formula whatsoever. Then 'A[A/X] = A', for '∀XX[∀XX/X] = ∀X∀XX = ∀XX'. Hence, the truth-conditions for '∀XX' include '∀XX' itself. That is, to know the truth-value of '∀XX' we must first know the truth-value of '∀XX'. This clearly seems to be viciously circular. In a recursive definition of truth, the truth-conditions for a complex sentence should be defined in terms of simpler sentences. So, there are independently good reasons to suppose that we cannot replace 'X' by any formula whatsoever when we drop the quantifier in a sentence of the following form '∀XA'. We should not replace 'X' by a formula that includes a propositional quantifier.[8] If this is correct, step (11) does not follow from step (10). Hence, the inconsistency argument fails. {(I), (P)} is not inconsistent (or at least we have not seen any reason to believe that it is). Consequently, we can avoid the paradox of infallibility.

This solution seems to be the most promising to me. However, it is not entirely unproblematic. The solution entails that we treat '∀' as a 'substitutional' quantifier that varies over sentences and not as an 'objectual' quantifier that varies over propositions. The paradox of infallibility might still be a problem for everyone who wants to use 'objectual' propositional quantifiers that vary directly over propositions and for everyone who wants the elimination rule for ∀ to be unlimited.

I conclude that we should take the paradox of infallibility seriously.[9]

[8] Some systems of this kind are developed in Rönnedal 2019. For more on propositional quantifiers, see, for example, Lewis and Langford 1932: 178-98, Kripke 1959, Bull 1969 and Fine 1970.

[9] I would like to thank two anonymous reviewers for some interesting comments on an earlier version of this paper.

## References

Blackburn, P., De Rijke, M. and Venema, Y. 2001, *Modal Logic*, Cambridge: Cambridge University Press.

Bull, R. 1969, "On Modal Logic with Propositional Quantifiers", *The Journal of Symbolic Logic*, 34, 2, 257-63.

Chellas, B.F. 1980, *Modal Logic: An Introduction*, Cambridge: Cambridge University Press.

Fagin, R., Halpern, J.Y., Moses, Y. and Vardi, M.Y. 1995, *Reasoning About Knowledge*, Cambridge, MA: The MIT Press.

Fine, K. 1970, "Propositional Quantifiers in Modal Logic", *Theoria*, 36, 336-46.

Hughes, G.E. and Cresswell, M.J. 1968, *An Introduction to Modal Logic*, London: Routledge (Reprinted 1990).

Kripke, S. 1959, "A Completeness Theorem in Modal Logic", *Journal of Symbolic Logic*, 24, 1, 1-14.

Lewis, C.I. and Langford, C.H. 1932, *Symbolic Logic*, New York: Dover Publications, Second edition 1959.

Meyer, J.-J. Ch. and van der Hoek, W. 1995, *Epistemic Logic for AI and Computer Science*, Cambridge: Cambridge University Press.

Priest, G., Berto, F. and Weber, Z. 2019, "Dialetheism", in Zalta, N.E. (ed.), *The Stanford Encyclopedia of Philosophy* (2019 Edition), https://plato.stanford.edu/entries/dialetheism/.

Rönnedal, D. 2019, "Semantic Tableau Versions of Some Normal Modal Systems with Propositional Quantifiers", *Organon F*, 26, 3, 505-36.