

# The surprise examination

Jelle Gerbrandy (jelle@gerbrandy.com)\*  
*Dipartimento di Informatica, Università di Torino*

**Abstract.** We examine a version of the surprise examination paradox using dynamic epistemic logic. We claim that the difficulties in the puzzle arise from the assumption that announcements are in general successful: the hearer will come to believe that they are true. This principle fails in certain specific cases, and we show that the announcement in the surprise exam paradox is an example. In dynamic epistemic logic, announcements that are not successful can be informative anyway. This analysis explains how sentences that are true but cannot consistently be believed, can be informative anyway, resolving many of the puzzling aspects of the puzzle.

## 1. The Puzzle

The surprise examination paradox exists in so many variants that it is probably more correct to speak of a family of puzzles than of a single one. More than a hundred papers have been written about different versions of this puzzle that revolves around *reductio ad absurdum* on the basis of a prediction about the lack of knowledge on the part of the audience. Chapter 7 of Sorensen (1988) and Chow (1998) provide organised and relatively non-partisan overviews of representative selections from the literature.

This paper is one more on the list. In it, I will present a new tack on the surprise exam paradox, one that is directly inspired by a semantics of information change that in the past years has received considerable attention: dynamic epistemic semantics. I will show how this logic is a natural framework for representing the puzzle, and how it sheds new light on the reasoning involved.

It is important to avoid confusion about which member of the family we are analysing. In the following version, which is very loosely based on the scenario in Landman (1986), some ambiguities in the original formulations are avoided.

Three numbered boxes are opened in sequence by a quiz master, starting from number 1, then number 2, and finally number 3. One of the boxes is known to contain a considerable amount of money,

---

\* Research supported by Lagrange Project of the Fondazione CRT. The author would like to thank Guido Boella, Paul Harrenstein, Barteld Kooi, and Luigi Sauro for useful comments on earlier versions of this paper, and the University of Turin for its hospitality.

the other two are empty. A player, say Marilyn Vos-Savant, who is assumed to be a perfect reasoner, gets the money if she *knows*, just before the box containing the money is opened, that this box is the one with the money in it. Marilyn is not allowed to guess; she must have a convincing argument that the money is in the box that is about to be opened.

From a game-theoretical point of view, this is an unusual game. Whether Marilyn wins the game depends solely on what she knows, not on what she does. There is no point in the game where she has to make a choice between alternative actions, so the notion of a strategy, which is central in game theory, does not make much sense here.

Actually the money is in the second box, and the quiz master knows this. The only thing that Marilyn knows of the game is that exactly one box contains the money, so Marilyn has no way of knowing that the money is in box 2, also not after the first box has been opened. That means that she cannot win the game.

Suppose that the quiz master, before opening any boxes, announces to the public, but within earshot of Marilyn: “Marilyn cannot win the game.” As we have seen, this is true.

Note that if the money were in the last box, Marilyn would be able to win the game. The quiz master is only justified in making his announcement if the money is in the first or the second box.

Now Marilyn reasons as follows. “Suppose the money is in the third box. In that case, I will know that the money is in that box at the moment when all other boxes are opened, and I will win the game. So, if the quiz master tells the truth, the last box must be empty.”

Having excluded the third box, at this point it seems that Marilyn can win the game, because she will know after the first box has been opened that the second box contains the money. However, she continues reasoning as follows:

“I know now that the money is not in the third box. But then, the money cannot be in the second box either, because then I could win the game, which contradicts what the quiz master said. So, I can exclude the second and the third box. But then it must be in the first box, and, since I know this, I can win the game. Again, this contradicts the quiz master’s announcement, so I have to conclude that all boxes are empty.”

Even if this conclusion is not a clear-cut contradiction, it is close to it: the quiz master is telling the truth, Marilyn uses the quiz master’s announcement together with some other patently true facts as premises for a simple inductive proof, which has a conclusion that is false.

There is a further argument that ‘proves’ that the prediction of the quiz master is true, also after (or even because of) his announcement.

Marilyn concludes that all boxes are empty, which is in contradiction with what she knows of the game. She must conclude that the information she has been given is inconsistent. In any case, she has no way of knowing which box contains the money, and she will lose the game. So, the quiz master’s announcement was true after all.

Note first of all that the conclusion of this argument is not very surprising: the quiz master’s announcement was true to begin with. Secondly, note that this ‘proof’ that Marilyn cannot win depends crucially on the assumption that Marilyn reasons as described. This assumption is questionable at least: if Marilyn is a skilled or even merely competent logician she surely will not accept an inference that leads from true premises to a false conclusion.<sup>1</sup> We may just as well assume she finds another, equally plausible, argument that allows her to win the game, like the ‘proof’ in the our description of the puzzle that stops after excluding the third box. So, it seems that in the present version of the puzzle this line of reasoning does not introduce a new problem, but merely shows, a second time, that there must be something wrong with Marilyn’s argument.

To pre-empt confusion about which variant of the problem is under scrutiny, it is probably useful to highlight some features our way of seeing the puzzle. First of all, we take it to be essential that what the quiz master says is unambiguously true at the moment that he makes his assertion, and that its truth does not depend on whether he makes his announcement to Marilyn, to the audience, or not at all. The problematic inference stems from the fact that Marilyn hears the quiz master making his announcement, not from the content of the announcement itself, which, after all, would have been unproblematic if Marilyn had not heard him make it. We exclude from consideration an interpretation that finds fault with what the quiz master says, by saying that it is an imprecise or vague statement, by construing the quiz master’s announcement as self-referential, or by taking it as a prediction that can only be proven to be correct *post factum*. Secondly, the propositional attitude we are concerned with is that of ‘(not) knowing’: we place ourselves in the ‘epistemic school,’ exemplified by, e.g. Binkley (1969) and Wright and Sudbury (1977).

---

<sup>1</sup> Compare a variant where the quiz master says: “Your husband is unfaithful to you, and you cannot win the game.” Argue that Marilyn is so shocked by the first conjunct that she cannot think clearly anymore, so she certainly cannot win the game, thereby proving the quiz master to be right. This may be a valid way of reasoning, but if it is problematic, this can hardly be seen as a problem of *logic*, as it is based on particular assumptions about the psychology of our fictional character.

In short, we ignore complications arising from the use of the words ‘surprise’, ‘will’ and ‘you’ in the version of the puzzle where the quiz master says something along the lines of: “The contents of the box to be opened will be a surprise to you,” but instead focus on the more straightforward (and arguably different) version where he is taken to say: “The contents of the box with the money are not known by Marilyn after the previous boxes have been opened.” This is not to say that other versions are not interesting in their own right; we just think that our approach does not have much to add to existing work discussing these.

Generalising, the problem under consideration is this: the quiz master gives Marilyn information that is true, and, by an argument that seems impeccable, Marilyn concludes the contrary. What goes wrong?

Our answer, in a nutshell, will be this. The general principle that after you learn a sentence, you believe that this sentence is true, is not tenable for sentences that express that the hearer does not have certain information. The announcement of the quiz master is an example of such a sentence. In this particular case, we will argue, the quiz master’s announcement provides Marilyn with precisely the information to win the game. The quiz master, confusingly, states a fact that becomes false by virtue solely of Marilyn learning of its truth. In our scenario, Marilyn’s argument goes awry at the point where she concludes that the second box must be empty, because this conclusion is based on the, admittedly reasonable, but actually false, assumption that the quiz master’s announcement remains true after she learned of it.

## 2. A formal language

One strategy for resolving puzzles such as the surprise examination is to make the reasoning in the puzzle formally precise. The language we will use for this purpose is a version of Dynamic Epistemic Logic: an extension of modal propositional logic with sentences that express change of information.

This language contains the usual logical connectives from epistemic logic:  $\neg$  for negation and  $\wedge$  and  $\vee$  for conjunction and disjunction, and an operator  $\square$  that stands for “Marilyn knows...”.<sup>2</sup> The dynamic character of the language is obtained by adding a unary operator  $[\varphi]$  for any sentence  $\varphi$  to the language: a sentence of the form  $[\varphi]\psi$  is to

---

<sup>2</sup> In the following, we will use the phrases like ‘knowing’, ‘believing’ and ‘having the information that’ interchangeably. We do not think that the distinction between these concepts are of importance in a puzzle that describes a perfect reasoner that is provided with true information only.

express that after Marilyn learns that  $\varphi$ ,  $\psi$  is true. We will use this operator to represent the change in Marilyn's knowledge during the game that occur, respectively, when a box is opened (Marilyn learns that it is empty), and, more interestingly, when the quiz master makes his announcement (Marilyn learns that she cannot win).

In our formalisation, we will use three propositional variables  $p_1$ ,  $p_2$  and  $p_3$  standing respectively for the fact that the money is in the first, second or third box. For example, the proposition that Marilyn does not know that the second box contains the money after the first box has been opened can be represented by the sentence  $[\neg p_1]\neg\Box p_2$ .

It is straightforward to transcribe the proposition that Marilyn cannot win in dynamic epistemic logic. There are three cases: the money is in the first, the second, or the third box. If the money is in the third box, Marilyn does not know this after she learnt that it is not in the first and not in the second; if the money is in box 2, Marilyn does not know this after learning it is not in box 1; and finally, Marilyn does not know it is in box 1.

$$\begin{aligned} & p_1 \wedge \neg\Box p_1 \\ \vee & p_2 \wedge [\neg p_1]\neg\Box p_2 \\ \vee & p_3 \wedge [\neg p_1][\neg p_2]\neg\Box p_3 \end{aligned}$$

Let us abbreviate this sentence as `not_win`. This sentence expresses a prediction about what Marilyn will know about the box with the money, solely on the basis of seeing the previous boxes being opened.

Anticipating a precise definition of the operator  $[\varphi]$ , we can represent Marilyn's reasoning by a formal inference in our language. With the axioms we will present below, it follows that the sentence `not_win` is equivalent with the following sentence of classical modal logic:  $(p_1 \wedge \neg\Box p_1) \vee (p_2 \wedge \neg\Box(\neg p_1 \rightarrow p_2)) \vee (p_3 \wedge \neg\Box((\neg p_1 \wedge \neg p_2) \rightarrow p_3))$ . Given this equivalence, it follows from some of the classical laws of epistemic logic that if Marilyn knows that `not_win` is true and that at least one box contains the money, her knowledge is inconsistent. More precisely, in the logic *K4* it follows from  $\Box(p_1 \vee p_2 \vee p_3)$  and  $\Box\text{not\_win}$  that  $\Box\perp$ . We will not bore the reader with an explicit inference, as its details are straightforward.

The fact that the Marilyn's reasoning can be represented in a formal proof in classical epistemic logic explains why her argument seems correct. This has led some authors to conclude that the fault lies with some rule of epistemic logic. In the following, we will argue that it is not necessary to find fault with the formal proof, as we can blame a basic premiss of the proof, namely that after the announcement of `not_win` Marilyn knows that `not_win` is true.

### 3. Update semantics

Dynamic semantics is a general term used for semantics of change, but sometimes it is used in a more specific sense to describe an approach in philosophic logic and semantics of natural language in which it is argued that replacing a semantics based on truth values with one based in change of information is useful to explain a number of phenomena, such as reasoning with defaults (in the cited paper of Veltman (1996)), anaphora (Groenendijk and Stokhof (1991)) and presuppositions (Beaver (2001)). The work of Gillies (2001) is more germane to the present discussion: Gillies argues that part of the difficulties surrounding sentences that are consistent but cannot be consistently believed stem from the fact that truth-value based semantics are not fine-grained enough as tools to explain the meaning of such sentences.

The system of Update Semantics that is introduced in the introductory section of Veltman (1996) is probably the most basic form of dynamic epistemic semantics. The information of a person is modeled by the set of the states of affairs that are consistent with her knowledge. In the simplest case, such a state of affairs can be described exhaustively by a set of propositional variables. Let us say that a *situation*  $s$  is a function that assigns to each propositional variable a truth-value (either 0 or 1). An *information state*  $\sigma$ , then, is a set of these situations.

The meaning of a sentence in update semantics is identified with a function that says how an information state changes as the result of learning that sentence. More formally, we define for each sentence  $\varphi$  in the classical modal fragment of our language a function  $\llbracket\varphi\rrbracket$ . Using postfix notation, we write  $\sigma\llbracket\varphi\rrbracket$  for the result of applying the function  $\llbracket\varphi\rrbracket$  to an information state  $\sigma$ , so that  $\sigma\llbracket\varphi\rrbracket$  is the information state that results if  $\varphi$  is learned in  $\sigma$ . If learning a sentence does not change an information state  $\sigma$ , i.e. if  $\sigma\llbracket\varphi\rrbracket = \sigma$ , then we say that  $\varphi$  is *accepted* in  $\sigma$ .

In the following definition of the information change potential of a sentence, we deviate somewhat from the original definition of Update Semantics.

$$\begin{aligned}\sigma\llbracket p \rrbracket &= \{s \in \sigma \mid p \text{ is true in } s\} \\ \sigma\llbracket \neg\varphi \rrbracket &= \{s \in \sigma \mid s \notin \sigma\llbracket\varphi\rrbracket\} \\ \sigma\llbracket \varphi \wedge \psi \rrbracket &= \sigma\llbracket\varphi\rrbracket \cap \sigma\llbracket\psi\rrbracket \\ \sigma\llbracket \Box\varphi \rrbracket &= \sigma \text{ if } \varphi \text{ is accepted in } \sigma \\ &\quad \emptyset \text{ otherwise}\end{aligned}$$

Learning  $p$  means excluding all situations from your state in which  $p$  is false; learning that  $\varphi$  is not the case is excluding all situations that would remain if you were to learn that  $\varphi$ ; and learning that  $\varphi$  and  $\psi$  are

true is the same as excluding all states that you would exclude when learning either  $\varphi$  or  $\psi$ .

The final clause for  $\Box\varphi$  reflects a principle of introspection: we assume that our agent is aware of what she knows, and in particular of the fact that she knows  $\varphi$ . That means that a sentence  $\Box\varphi$  cannot provide her with information that is new for her. In effect, a sentence  $\Box\varphi$  works as a test: either our agent already knew  $\varphi$ , in which case  $\Box\varphi$  does not provide any new information (and her information state remains the same), or she did not, in which case  $\Box\varphi$  is inconsistent with her information, and the result of updating with this sentence is the inconsistent, empty, state.

Update Semantics is exclusively about the information change brought about by sentences, and deliberately agnostic about the truth of sentences. As our puzzle makes explicit reference to the truth of the quiz master's announcement, we need a more traditional, truth-based semantics as well. It turns out that we can easily define an update function on the basis of a truth-value based semantics that is equivalent to our version of Update Semantics, a fact that corroborates the claim that these definitions are basically correct.

A *Kripke model* is a pair  $\sigma, s$ , where  $s$  is a situation and  $\sigma$  is an information state. The idea is that  $s$  represents the state of affairs as it actually obtains. We can define the truth of a sentence in a Kripke model in the standard way:

$$\begin{aligned} \sigma, s \models p & \text{ iff } p \text{ is true in } s \\ \sigma, s \models \neg\varphi & \text{ iff } \sigma, s \not\models \varphi \\ \sigma, s \models \varphi \wedge \psi & \text{ iff } \sigma, s \models \varphi \text{ and } \sigma, s \models \psi \\ \sigma, s \models \Box\varphi & \text{ iff for all } t \in \sigma \text{ it holds that } \sigma, t \models \varphi \\ \sigma, s \models [\varphi]\psi & \text{ iff } \sigma[[\varphi]], s \models \psi \end{aligned}$$

The function  $[[\varphi]]$  that is used in the definition of  $[\varphi]\psi$  can inductively defined as in update semantics, but we can also define it on the basis of our classical definition of truth:<sup>3</sup>

$$\sigma[[\varphi]] = \{s \in \sigma \mid \sigma, s \models \varphi\}$$

The idea is that an update of  $\sigma$  with  $\varphi$  means retaining only the situations in which  $\varphi$  is true (in the context of our agent being in

---

<sup>3</sup> This definition can be seen as a simplified case of what has by now become a standard definition of information change in multi-agent systems, in its different incarnations of Landman (1986), Plaza (1989), Gerbrandy and Groeneveld (1997) and Baltag and Moss (2004). Note also that this 'reduction' of dynamic semantics to classical semantics is not as straightforward when we add specifically 'dynamic' operators, as is done in Veltman's original paper.

information state  $\sigma$ ). Roughly, we can say that  $\sigma[[\varphi]]$  is the intersection of  $\sigma$  with the proposition expressed by  $\varphi$ . This new function is defined for a wider range of sentences than in update semantics, but for the classical modal fragment, i.e. for all sentences without subformulas of the form  $[\varphi]\psi$ , the functions defined here are the same as those of update semantics. The observation that these two different definitions result in the same function can be seen as justifying the claim that the function  $[[\varphi]]$  is the 'right' one

There is a sound and complete logic for this semantics, that contains the axioms of *K45*, a functionality axiom that states that  $\neg[\varphi]\psi$  is equivalent with  $[\varphi]\neg\psi$ , a distribution axiom  $[\varphi](\varphi \rightarrow \psi) \rightarrow ([\varphi]\varphi \rightarrow [\varphi]\psi)$ , an axiom that says that an update does not change the truth value of a proposition,  $[\varphi]p \leftrightarrow p$ , and an axiom that describes the effect of an update on the information state:  $[\varphi]\Box\psi \leftrightarrow \Box(\varphi \rightarrow [\varphi]\psi)$ .

#### 4. Unsuccessful sentences

One property that has been considered uncontroversial for a natural notion of updating or learning is what Alchourrón et al. (1985) in their classical paper call *success*: after learning a sentence  $\varphi$ , you come to believe that  $\varphi$  (cf. also Stalnaker (1978) on assertion). This property seems evident to the point of triviality: what else could learning something mean if not coming to believe it is true?

Be that as it may, in Update Semantics an update is not always successful. The reason is that in our language, we can express that our agent lacks certain knowledge, and, since her knowledge can change as the result of an update, there is no guarantee that this lack of knowledge will persist. More specifically, our agent at a certain moment in time may accept  $\neg\Box p$ , then learn that  $p$  is true, and afterwards, accept  $\Box p$ .

Update Semantics predicts that for sentences that express lack of information on the part of the hearer in combination with providing other, positive information, success fails. This can seem counterintuitive at first sight.

As an example, consider the sentence  $p \wedge \neg\Box p$ , which expresses both that  $p$ , and that our agent does not know that  $p$ . This sentence is consistent, but there is no consistent information state in which it is accepted. Let us call sentences that are consistent but cannot consistently be known 'Moorean,' after the related Moore paradox. Clearly, if we want to define an update function in which Moorean sentences can be informative in a non-trivial way, we have to give up the principle of success.

Even if saying “The money is in box 2, but you don’t know it” is a somewhat perverse way of packaging the information you want to convey, the sentence *can* express information that may very well be true, and the hearer should be able to understand this and incorporate the new information into the information she already has. In the classical logic of belief and belief change, it is hard to see how this could be done. There seem to be only two options open, both unsatisfactory: the hearer either adds the new sentence to her already existing stock of beliefs (in the way prescribed by the postulates of Alchourrón et al. (1985)), in which case her beliefs become inconsistent, or she rejects the information entirely, in which case she does not make any use of the information offered to her. Update semantics (and dynamic epistemic logic in general) provides us with a middle way: an agent can learn that such a sentence is true (that is, true before she updated with the sentence) without coming to believe that the sentence is true (after the update).

To see how this works, let  $\sigma$  be a state in which it is not known whether  $p$ ; i.e. it contains situations in which  $p$  is true and situations in which  $p$  is false. Applying the definitions above, the information state after updating with  $p \wedge \neg \Box p$  is the intersection of the information state that results if the first conjunct  $p$  is learned, with that of the information state that results if the second conjunct  $\neg \Box p$  is learned. Since the second conjunct provides the agent with no new information, the result is a state that consists of exactly the situations from  $\sigma$  where  $p$  is true. In this resulting state it is, of course, known that  $p$  is true and so the sentence we started with,  $p \wedge \neg \Box p$ , is not accepted.

Success is a property that has a certain intuitive appeal, and it would be unfortunate if we were to lose it altogether. Fortunately, it holds that an update with a sentence will only fail to be successful if that sentence contains a negative occurrence of  $\Box$  (i.e. in the scope of an uneven number of negations): success only fails in ‘pathological’ cases.

It is perhaps interesting to consider the Ramsey test in this context, which states that one is warranted to believe an implication ‘if  $\varphi$  then  $\psi$ ’ just in case it holds that after you learn that  $\varphi$ , you accept that  $\psi$ . Again, this principle, although reasonable at first sight and probably valid in general, fails if  $\psi$  is about the information of the Ramsey tester. For example, the fact that after you learn that there is life on Mars, you believe that this is so, says more about our gullibility than about your extraterrestrial expertise, while if your belief that “if there is life on Mars, I know that there is” were justified, you would qualify for a research position at NASA.

Even if blatant occurrences of Moorean sentences are probably rare in daily life, assertions of lack of knowledge of the hearer do appear, more or less hidden, in a range of puzzles, and an analysis using dynamic epistemic semantics can give an independently motivated account of the reasoning involved in them. Particular examples are the puzzle of the muddy children or the wise men (Plaza (1989), Gerbrandy (1998), van Ditmarsch and Kooi (pear)), the Conway Paradox (also known as ‘Mr. Sum and Mr. Product’) (van Emde Boas et al. (1984), and again Plaza (1989)), the Fitch Paradox (van Benthem (1997)), the puzzle of the designated student (invented by Sorensen (1988)) and the problem at hand, the paradox of the surprise exam (Gerbrandy (1998), van Ditmarsch and Kooi (pear)).

## 5. A formal analysis

All ingredients are now in place to formally represent the reasoning in the puzzle of the surprise exam. Suppose  $s_1$ ,  $s_2$  and  $s_3$  are situations in which the money is in the first, second and third box, respectively. At the outset Marilyn does not know which of these three situations obtains, and her information state can be represented by the set  $\{s_1, s_2, s_3\}$ . In fact, the money is in the second box, so  $s_2$  is the actual situation. The Kripke model representing the situation at the outset of the puzzle is therefore  $\{s_1, s_2, s_3\}, s_2$ .

The sentence `not_win` is true in this model:  $\{s_1, s_2, s_3\}, s_2 \models \text{not\_win}$ . Note that if  $s_1$  were the real world, `not_win` would be true as well, but if  $s_3$  were the real world, then the sentence would be false. All this seems to be in accordance with our intuitions that Marilyn can win only if the money is in the last box.

Consider now what happens if Marilyn updates with `not_win` in this model. Her state after learning this sentence is the set  $\{s_1, s_2\}$  in which she has excluded  $s_3$  as a possibility, because if  $s_3$  were the actual situation, `not_win` would be false.

In this new situation,  $\{s_1, s_2\}, s_2$ , Marilyn *can* win the game: she knows after an update with  $p_1$  that the money has to be in the second box. Indeed, in this new model, the sentence `not_win` is false. But Marilyn does not know this; as far as she knows, the money could be in the first box as well, and in that case, she cannot win, so  $\neg\Box\text{not\_win}$  is true.

If `not_win` indeed correctly paraphrases what the quiz master meant, then it seems that we have resolved the puzzle, in the sense that we have a plausible logical theory in which it holds that the quiz master was correct in making his announcement and Marilyn can consistently

learn the proposition expressed by the announcement without drawing any conclusions that are false.

If update semantics describes what a logically competent agent should do, we can pinpoint the mistake in Marilyn's reasoning at the moment that she continues her argument after excluding the third box. She argues that the money cannot be in the second box, because that would contradict what the quiz master said. To be sure, she is correct in concluding that, now, after the new information provided by the quiz master, she can win if the money is in box 2, and she is correct in that the quiz master said that she couldn't, but she is not correct in seeing a contradiction between these two claims. What the quiz master said (i.e. the proposition expressed by `not_win`) was true before he made his announcement, but she cannot assume that the sentence will remain true after being announced. In fact, `not_win` is false in the updated model. So, in update semantics, the first step in the inductive proof is correct: Marilyn can exclude the last box. The argument halts there.

The fact that in update semantics there might be sentences that are true, but become false solely on the basis of the fact that they are learned, might be confusing, but it certainly is not paradoxical.

## 6. Conclusions

In the introduction of this paper, I summarised the main question as follows:

The quiz master gives Marilyn information that is true, and, by an argument that seems impeccable, Marilyn concludes the contrary. What goes wrong?

We argued that basically, nothing needs to be wrong with learning a true proposition and concluding that its contrary is true. If update semantics is a reasonable account of information change, there indeed exist situations and sentences where such an unusual argument is valid. One example is the Moorean  $p \wedge \neg \Box p$ .

That is not to say that Marilyn does not make a mistake in her argument. In her argument, she assumes that announcements are successful, in the sense that after she learns a sentence, she will believe that sentence to be true. In making this assumption, she is in good company (most of the postulates of Alchourrón et al. (1985) have been criticised at some point, but the postulate of success has rarely been challenged). However, when considering the learning of sentences that express lack of information on the part of the hearer, this assumption is not tenable. The quiz master's announcement is such a sentence, and its lack of success explains where Marilyn's inference fails.

## References

- Alchourrón, C. E., P. Gärdenfors, and D. Makinson: 1985, 'On the Logic of Theory Change: Partial Meet Contraction and Revision Functions'. *Journal of Symbolic Logic* **50**, 510–530.
- Baltag, A. and L. S. Moss: 2004, 'Logics for Epistemic Programs'. *Synthese* **139**(2), 165–224.
- Beaver, D.: 2001, *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications, Stanford.
- Binkley, R.: 1969, 'The Surprise Examination in Modal Logic'. *Journal of Philosophy* **65**, 127–136.
- Chow, T. Y.: 1998, 'The Surprise Examination or Unexpected Hanging Paradox'. *American Mathematical Monthly* **105**(1), 41–51.
- Gerbrandy, J.: 1998, 'Bisimulations on Planet Kripke'. Ph.D. thesis, Universiteit van Amsterdam. ILLC Dissertation Series DS-1999-01.
- Gerbrandy, J. and W. Groeneveld: 1997, 'Reasoning about Information Change'. *Journal of Logic, Language and Information* **6**(2), 147–169.
- Gillies, A. S.: 2001, 'A New Solution to Moore's Paradox'. *Philosophical Studies* **105**, 237–250.
- Groenendijk, J. and M. Stokhof: 1991, 'Dynamic Predicate Logic'. *Linguistics and Philosophy* **14**(1), 39–100.
- Landman, F.: 1986, 'Towards a Theory of Information'. Ph.D. thesis, Universiteit van Amsterdam. Also appeared as GRASS 6 with Foris Publications, Dordrecht, Holland/Cinnaminson, U.S.A.
- Plaza, J.: 1989, 'Logics of Public Communications'. In: M. Emrich, M. Pfeifer, M. Hadzikadic, and Z. Ras (eds.): *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*. New York: Academic Press, pp. 201–216.
- Sorensen, R.: 1988, *Blindspots*. Clarendon Press.
- Stalnaker, R. C.: 1978, 'Assertion'. In: P. Cole (ed.): *Pragmatics (Syntax and Semantics 9)*. New York: Academic Press, pp. 315–312.
- van Benthem, J.: 1997, 'On What One may Come to Know'. *Analysis* **64**(2), 95–105.
- van Ditmarsch, H. and B. Kooi: to appear, 'The Secret of my Success'. *Synthese*.
- van Emde Boas, P., J. Groenendijk, and M. Stokhof: 1984, 'The Conway Paradox: Its Solution in an Epistemic Framework'. In: J. Groenendijk, T. M. V. Janssen, and M. Stokhof (eds.): *Truth, Interpretation and Information: Selected Papers from the Third Amsterdam Colloquium*. Dordrecht: Foris Publications, pp. 159–182.
- Veltman, F.: 1996, 'Defaults in Update Semantics'. *Journal of Philosophical Logic* **25**, 221–261.
- Wright, C. and A. Sudbury: 1977, 'The Paradox of the Unexpected Examination'. *Australasian Journal of Philosophy* **55**, 41–58.