

# Logical Definitions of Lying

Chiaki Sakama

Department of Computer and Communication Sciences  
Wakayama University,  
Sakaedani, Wakayama 640-8510, Japan  
sakama@sys.wakayama-u.ac.jp

**Abstract.** This paper provides logical analyses of various definitions of *lying*. We first formulate twelve definitions of lying that appear in the literature of philosophy and were comprehensively studied by Mahon [14]. We use a propositional multi-modal logic that can represent belief and intention of agents. We then compare different definitions of lying and examine which one is best supported by both logically and empirically.

**General Terms:** theory. **Keywords:** lying, modal logic, reasoning.

## 1 Introduction

Understanding *what is lying* is necessary in identifying trustful agents, and providing ways to protect users from being deceived in multiagent societies. The problem has been studied in the field of philosophy and there is a number of different definitions of lying. Recently, James E. Mahon [14] provides a comprehensive study of lying, in which he examines twelve different definitions of lying that appear in the literature and argues which one is well-defined and empirically acceptable. He judges that one definition is acceptable or not by considering whether it satisfies four necessary conditions for lying, and by excluding those definitions that permit counter-intuitive examples. As a result, he concludes that two of the twelve definitions are the best ones.

There are several studies that provide formal account of lying. Bonatti et al. [1] study a theory of databases that could lie to users to preserve security. Firozabadi and Jones [9] define lying in terms of action logic. O’Neill [16] formulates various types of speech acts including lying using an epistemic logic. Sklar et al. [20] formulate lying with argument-based dialogues. Sakama et al. [17] provide logical analyses of different categories of dishonesties. These studies employ one definition of lying and provide its logical account for their purposes.

The goal of this paper is to provide logical formulation of different definitions of lying and compare their formal properties. We first formulate twelve definitions of lying that have been proposed in the philosophical literature and were analyzed in an informal way by Mahon [14]. To this end, we use a multi-modal logic that can represent belief and intention of agents. We then provide logical analyses for different definitions of lying and investigate their formal properties. We introduce five conditions that are requested to be satisfied by the definition of lying. The results of this paper provide formal grounds for the selection of best definitions of lying, and make us better understand what is lying.

## 2 A Logic for Belief and Intention

In this paper, we use a propositional modal logic of intentional communication [7]. A propositional modal language  $L_0$  is built from a finite set of propositional constants  $\{p, q, r, \dots\}$  on the logical connectives  $\neg, \vee, \wedge, \supset, \equiv$ , and on two families of modal operators,  $(B_a)_{a \in A}$  and  $(I_a)_{a \in A}$ , where  $A$  is a finite set of agents. Well-formed formulas (or *sentences*) in  $L_0$  are defined as usual as those belonging to a multi-modal propositional logic. Sentences in  $L_0$  will be denoted by the small Greek letters, and parentheses are employed as usual to clarify the structure of sentences.  $\top$  and  $\perp$  represent valid and contradictory sentences, respectively. The intuitive readings of  $B_a\phi$  and  $I_a\phi$  are that an agent  $a$  believes that  $\phi$  and intends that  $\phi$ , respectively. A Kripkean semantics is defined for  $L_0$ . Informally speaking,  $B_a\phi$  (resp.  $I_a\phi$ ) holds iff  $\phi$  is true in all states of affairs compatible with  $a$ 's current beliefs (resp. intentions). For example, if  $\phi$  means that it rains,  $I_a\phi$  should be read as “ $a$  intends to act in such a way that he or she brings about a state of affairs in which it rains”.<sup>1</sup> A logic  $BI_0$  is defined over  $L_0$ , that is an extension of  $KD45_n$  [12] and has the following axioms and inference rules:

- (P) All propositional tautologies.  
 (K<sub>B</sub>)  $B_a\phi \wedge B_a(\phi \supset \psi) \supset B_a\psi$     and    (K<sub>I</sub>)  $I_a\phi \wedge I_a(\phi \supset \psi) \supset I_a\psi$ .  
 (D<sub>B</sub>)  $B_a\phi \supset \neg B_a\neg\phi$     and    (D<sub>I</sub>)  $I_a\phi \supset \neg I_a\neg\phi$ .  
 (4<sub>B</sub>)  $B_a\phi \supset B_aB_a\phi$     and    (4<sub>I</sub>)  $I_a\phi \supset B_aI_a\phi$ .  
 (5<sub>B</sub>)  $\neg B_a\phi \supset B_a\neg B_a\phi$     and    (5<sub>I</sub>)  $\neg I_a\phi \supset B_a\neg I_a\phi$ .

$$(MP) \frac{\phi \quad \phi \supset \psi}{\psi}, \quad (N_B) \frac{\phi}{B_a\phi}, \quad (N_I) \frac{\phi}{I_a\phi}.$$

Remark that (N<sub>I</sub>) says that all theorems hold at all state of affairs that an agent might intend to bring about [7].

To represent a speech act of an agent, we introduce the unary predicate  $utter_{xy}$  defined over sentences in  $L_0$  with  $x, y \in A$ . An expression  $utter_{ab}(\sigma)$  means that an agent  $a$  expresses a sentence  $\sigma$  to an agent  $b$ . In particular,  $utter_{aa}(\sigma)$  means that  $a$  expresses a sentence  $\sigma$  but the statement is directed to no one. A language  $L_0^U$  is defined as  $L_0$  together with the predicate  $utter_{xy}$ . Sentences in  $L_0$  are extended to  $L_0^U$  accordingly. If an agent utters something, he/she intends the speech act and is aware of his/her utterance. This is expressed by the next axiom:

$$(U_{IB}) \quad utter_{ab}(\sigma) \supset I_a(utter_{ab}(\sigma)) \wedge B_a(utter_{ab}(\sigma)).$$

We also assume that any utterance to a hearer is recognized by the hearer, and the speaker believes the recognition by the hearer. This is expressed by the axiom:

$$(U_{BB}) \quad utter_{ab}(\sigma) \supset B_b(utter_{ab}(\sigma)) \wedge B_aB_b(utter_{ab}(\sigma)).$$

The system  $BI_0^U$ , defined over  $L_0^U$ , is the weakest extension of  $BI_0$  by the two axioms (U<sub>IB</sub>) and (U<sub>BB</sub>). If a sentence  $\phi$  is a theorem of  $BI_0^U$ , we write  $\vdash \phi$ . Note that by N<sub>B</sub> and N<sub>I</sub>, each agent believes and intends that other agents follow the same logic  $BI_0^U$ . Thus,  $B_aB_b\phi \supset B_a\neg B_b\neg\phi$  and  $B_a(I_b\phi \wedge I_b(\phi \supset \psi)) \supset B_aI_b\psi$ , for instance.

<sup>1</sup> With such an interpretation, it becomes acceptable to apply an intention operator to arbitrary sentences [7].

### 3 Definitions of Lying

Mahon [14, 15] compares different definitions of lying in the philosophical literature and argues which definitions are most intuitive and acceptable. According to [15], there are at least four necessary conditions for lying. If a person lies, then **(i)** the person makes a statement (**statement condition**), and **(ii)** the person believes the statement to be false (**untruthfulness condition**), and **(iii)** the untruthful statement is made to another person (**addressee condition**), and **(iv)** the person intends that other person's believing the untruthful statement to be true (**intention to deceive addressee condition**). In what follows, we reformulate twelve definitions of lying that are argued in [14]. In this section,  $a$  and  $b$  represent two agents and  $\sigma$  and  $\lambda$  are two sentences of  $L_0^U$ .

#### 3.1 Vrij's Definition

Vrij considers lying as "a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue" [21]. Vrij's definition is stated in [14] as follows:

(L1) *To lie (to another person) is: to attempt to create a believed-false belief without forewarning (in another person).*

(L1) is formulated in  $L_0^U$  as follows.

**Definition 3.1 (L1)**  $L_{ab}(\sigma) \stackrel{def}{=} B_a \neg \sigma \wedge I_a B_b \sigma \wedge B_a B_b B_a \sigma$ .

(L1) represents that an agent  $a$  lies to another agent  $b$  on the sentence  $\sigma$  if  $a$  believes  $\sigma$  to be false and intends  $b$ 's believing  $\sigma$ . Moreover, without forewarning,  $a$  believes that it is justified for  $b$  to believe that  $\sigma$  is believed to be true by  $a$ . (L1) does not satisfy the statement condition and the addressee condition. It satisfies the untruthfulness condition ( $B_a \neg \sigma$ ) and the intention to deceive addressee condition ( $I_a B_b \sigma$ ). Vrij's definition does not require any statement. Vrij says that "lying does not necessarily require the use of words. The athlete who fakes a foot injury after a bad performance is lying without using words" [21]. On the other hand, (L1) requires lying to be an act that happens without forewarning. Thus, "magicians are therefore not lying during their performance, as people in the audience expect to be deceived" (ibid). Mahon views (L1) too broad as a definition of lying. This is because "according to (L1), feigning a yawn, wearing a hairpiece, making a phony smile, wearing an engagement ring when one is not engaged, not wearing a wedding ring when one is married, or pretending to talk to someone on a cell phone, etc., is lying" [14]. Mahon concludes that (L1) is merely a definition of attempting to deceive and then rejects (L1) as a definition of lying.

#### 3.2 Shibles's Definition

Shible's definition of lying [18] is stated in [14] as follows:

(L2) *To lie (to another person) is: to make a believed-false statement (to another person).*

(L2) is formulated in  $L_0^U$  as follows.

**Definition 3.2 (L2)**  $L_{2ab}(\sigma) \stackrel{def}{=} (utter_{aa}(\sigma) \vee utter_{ab}(\sigma)) \wedge B_a \neg \sigma$ .

(L2) represents that an agent  $a$  lies on the sentence  $\sigma$  if  $a$  utters a believed-false sentence  $\sigma$ . (L2) satisfies the statement condition  $(utter_{aa}(\sigma) \vee utter_{ab}(\sigma))$  but does not satisfy the addressee condition. (L2) also satisfies the untruthfulness condition  $(B_a \neg \sigma)$  but does not satisfy the intention to deceive addressee condition. Different from (L1), (L2) requires that a statement be made (by uttering). This implies that, according to (L2), it is impossible for a person to lie by omitting to utter an expression [14]. On the other hand, (L2) does not require that the statement be made to anyone. Shibley says in [18] that “a lie is merely a contradiction between belief (self-talk) and expression”. Thus, “according to (L2), if a person goes into a room he believes to be empty, and utters untruthful statements, then that person is lying” [14]. Mahon rejects (L2) as “it does seem wrong that simply making an untruthful statement, to no one, is lying” (ibid).

### 3.3 Bok’s Definition

Bok’s definition of lying [2] is stated in [14] as follows:

(L3) *To lie (to another person) is: to make a statement (to another person) with the intention to deceive (the other person).*

(L3) is formulated in  $L_0^U$  as follows.

**Definition 3.3 (L3)**  $L_{3ab}(\sigma, \lambda) \stackrel{def}{=} utter_{ab}(\sigma) \wedge B_a B_b(\sigma \supset \lambda) \wedge B_a \neg \lambda \wedge I_a B_b \lambda$ .

(L3) represents that an agent  $a$  lies to another agent  $b$  on the sentence  $\sigma$  if (i)  $a$  utters a sentence  $\sigma$  to another agent  $b$ , (ii)  $a$  believes that  $b$ ’s believing  $\sigma$  leads  $b$  to believing  $\lambda$ , (iii)  $a$  believes the falsity of  $\lambda$ , and (iv) believing  $\lambda$  by  $b$  is what  $a$  intends to achieve. (L3) satisfies both the statement condition and the addressee condition  $(utter_{ab}(\sigma))$ . (L3) also satisfies the intention to deceive addressee condition  $(I_a B_b \lambda)$ . On the other hand, (L3) does not satisfy the untruthfulness condition. That is, an agent  $a$  can utter a *believed-true* sentence  $\sigma$  with the intention that  $b$  uses it to reach a wrong conclusion  $\lambda$ . Mahon rejects (L3) “because it allows for the possibility of lying by making a truthful statement”. According to this definition, if a father says to his child that “You will get a Christmas gift” and the statement makes his child believe that Santa Clause will come, then he is lying. In this case, however, it seems natural to consider that father does not lie because his child will get a Christmas gift, even if his intention is to make his child believe the existence of Santa Clause.

### 3.4 OED Definition

The *Oxford English Dictionary* defines a lie as “a false statement made with the intent to deceive”. It is redefined in [14] as follows:

(L4) *To lie (to another person) is: to make a false statement (to another person) with the intention to deceive (some person or other).*

(L4) is formulated in  $L_0^U$  as follows.

**Definition 3.4 (L4)**  $L_{4ab}(\sigma, \lambda) \stackrel{def}{=} utter_{ab}(\sigma) \wedge \neg\sigma \wedge B_a B_b(\sigma \supset \lambda) \wedge B_a \neg\lambda \wedge I_a B_b \lambda$ .

(L4) satisfies both the statement condition and the addressee condition ( $utter_{ab}(\sigma)$ ). (L4) also satisfies the intention to deceive addressee condition ( $I_a B_b \lambda$ ). On the other hand, (L4) does not satisfy the untruthfulness condition. It requires the statement to be false (**falsity condition**), rather than to be believed to be false. Mahon rejects (L4) because “one can lie by being truthful with an intention to deceive, when it just so happens that one is mistaken” [14]. For instance, if a clock strikes nine and a mother says to her child that “Nine! It’s time to go to bed, otherwise, the devil will come”. She expects that her child believes that the devil will come if he/she sits up late at night. If the clock is wrong and it is five to nine, then, according to (L4), she is lying. Else if the clock is correct, she is not lying. Thus, whether she lies or not depends on the correctness of the clock, and has no relation with devil’s coming, which is unintuitive. There are other controversial examples and Mahon concludes that “it is better to say that one person is attempting to deceive another person, rather than to say that someone is lying to someone” (ibid).

### 3.5 Coleman and Kay’s Definition

Coleman and Kay [6] strengthen the OED definition and require a lie to be a statement that is both false and believed-false. The definition is stated in [14] as follows:

(L5) *To lie (to another person) is: to make a believed-false and false statement (to another person) with the intention that that statement be believed to be true (by the other person).*

(L5) is formulated in  $L_0^U$  as follows.

**Definition 3.5 (L5)**  $L_{5ab}(\sigma) \stackrel{def}{=} utter_{ab}(\sigma) \wedge B_a \neg\sigma \wedge \neg\sigma \wedge I_a B_b \sigma$ .

(L5) satisfies both the statement condition and the addressee condition ( $utter_{ab}(\sigma)$ ). (L5) satisfies the untruthfulness condition ( $B_a \neg\sigma$ ) as well as the falsity condition ( $\neg\sigma$ ). (L5) also satisfies the intention to deceive addressee condition ( $I_a B_b \sigma$ ). (L5) is considered a special case of (L4) with  $\lambda \equiv \sigma$ . Mahon rejects (L5) because “a person is not lying simply because the untruthful statement that she makes with the intention to deceive her addressee just happens to be true” [14], and “it does seem peculiar that whether or not one is lying depends upon luck” (ibid).

### 3.6 Kupfer’s Definition

Kupfer [13] addresses that “a person lies when he asserts something to another which he believes to be false with the intention of getting the other to believe it to be true”. It is redefined in [14] as follows:

(L6) *To lie (to another person) is: to make a believed-false statement (to another person) with the intention that that statement be believed to be true (by the other person).*

(L6) is formulated in  $L_0^U$  as follows.

**Definition 3.6 (L6)**  $L\theta_{ab}(\sigma) \stackrel{def}{=} utter_{ab}(\sigma) \wedge B_a \neg \sigma \wedge I_a B_b \sigma$ .

(L6) satisfies both the statement condition and the addressee condition ( $utter_{ab}(\sigma)$ ). (L6) also satisfies the untruthfulness condition ( $B_a \neg \sigma$ ), and the intention to deceive addressee condition ( $I_a B_b \sigma$ ). According to Mahon, “if any definition of lying may lay claim to being the standard definition of lying, then it is (L6)” [14]. (L6) represents an intention to deceive about the contents of the statement that is made. On the other hand, there would be an intention to deceive about one’s belief in the truth of the statement that one makes (**believed truthfulness condition**) [15]. Borrowing an example of [14], suppose an FBI agent working undercover in a criminal organization. The crime boss notices this fact, but the FBI agent has no suspicion of this. If the crime boss says the FBI agent that there are no informants in his organization, then the boss cannot intend that the FBI agent believes this statement to be true because the boss knows that the agent is an informant. In this case, the crime boss can only intend that the FBI agent believes that the boss believes this statement to be true. According to (L6), the boss is not lying to the FBI agent. To cope with such cases, Mahon modifies (L6) as follows:

(L6\*) *To lie (to another person) is: to make a believed-false statement (to another person), either with the intention that that statement be believed to be true (by the other person), or with the intention that it be believed (by the other person) that that statement is believed to be true (by the person making the statement), or with both intentions.*

The modified version (L6\*) is formulated in  $L_0^U$  as follows.

**Definition 3.7 (L6\*)**  $L\theta_{ab}^*(\sigma) \stackrel{def}{=} utter_{ab}(\sigma) \wedge B_a \neg \sigma \wedge (I_a B_b \sigma \vee I_a B_b B_a \sigma)$ .

(L6\*) states that a lie requires either an intention to deceive about the contents of the statement that is made, or an intention to deceive about the beliefs of the person making the statement, or both. The believed truthfulness condition is represented by  $I_a B_b B_a \sigma$ . Mahon asserts that (L6) and (L6\*) are two best definitions of lying [14].

### 3.7 Frankfurt’s Definition

Frankfurt [10] also considers lying involves two distinct intentions to deceive: the one is “about the state of affairs to which he (a liar) explicitly refers and of which he is purporting to give a correct account”, and the other is “about his own beliefs and what is going on in his mind” (ibid). Frankfurt’s definition is stated in [14] as follows:

(L7) *To lie (to another person) is: to make a believed-false statement (to another person) with the intention that that statement be believed to be true (by the other person) and with the intention that it be believed (by the other person) that that statement is believed to be true (by the person making the statement).*

(L7) is formulated in  $L_0^U$  as follows.

**Definition 3.8 (L7)**  $L\gamma_{ab}(\sigma) \stackrel{def}{=} utter_{ab}(\sigma) \wedge B_a \neg \sigma \wedge I_a B_b \sigma \wedge I_a B_b B_a \sigma$ .

(L7) satisfies the statement condition, the addressee condition ( $utter_{ab}(\sigma)$ ), and the untruthfulness condition ( $B_a \neg \sigma$ ). (L7) satisfies both the intention to deceive addressee condition ( $I_a B_b \sigma$ ) and the believed truthfulness condition ( $I_a B_b B_a \sigma$ ). In contrast to (L6\*), (L7) requires two distinct intentions to deceive be present and rules out cases in which only one intention to deceive is present. So, “if the lie works, then its victim is twice deceived” [10]. According to (L7), the crime boss is not lying to the FBI agent in the example given in Section 3.6. With this and another reasons, Mahon rejects (L7).

### 3.8 Chisholm and Feehan’s Definition

Chisholm and Feehan [5] provide a complex definition of lying which is restated by [14] as follows.

*(L8) To lie (to another person) is: to make a believed-not-true or believed-false statement (to another person), under conditions that are such that, (i) it is believed (by the person making the statement) that it is justified (for the other person) to believe that that statement is believed to be true (by the person making the statement), and (ii) it is believed (by the person making the statement) that it is justified (for the other person) to believe that it is intended (by the person making the statement) that it be believed (by the other person) that that statement is believed to be true (by the person making the statement).*

(L8) is formulated in  $L_0^U$  as follows.

**Definition 3.9 (L8)**  $L\delta_{ab}(\sigma) \stackrel{def}{=} utter_{ab}(\sigma) \wedge (\neg B_a \sigma \vee B_a \neg \sigma) \wedge B_a B_b B_a \sigma \wedge B_a B_b I_a B_b B_a \sigma$ .

(L8) satisfies both the statement condition and the addressee condition ( $utter_{ab}(\sigma)$ ). However, (L8) does not always require the untruthfulness condition ( $\neg B_a \sigma \vee B_a \neg \sigma$ ), and it does not require the intention to deceive addressee condition. Although the condition  $B_a B_b B_a \sigma$  appears in (L1), the remaining conditions are quite different from those appearing in (L2)–(L7).  $B_a B_b B_a \sigma \wedge B_a B_b I_a B_b B_a \sigma$  says that  $a$  believes that  $b$  believes that not only  $a$ ’s believing  $\sigma$ , but also  $a$ ’s intention to making  $b$ ’s believing  $B_a \sigma$ . According to (L8), an untruthful statement that is made merely in play or in irony is not a lie because the speaker does not believe that the hearer believes that the speaker believes what he says. On the other hand, (L8) has “the very odd and unacceptable result that a notoriously dishonest person cannot lie to people who he knows distrust him. Their definition implies that it is self-contradictory to say that I lie when I know that others know that I am lying” [4]. Mahon also rejects (L8) by similar reasons.

### 3.9 Simpson’s Definition

Simpson [19] provides yet another definition of lying. It is stated in [14] as follows:

*(L9) To lie (to another person) is: to make a believed-false statement (to another person) with the intention that that statement be believed to be true (by the other person), and with the intention that it be believed (by the other person) that that statement is believed to be true (by the person making the statement), and with the intention that it be believed (by the other person) that it is*

*intended (by the person making the statement) that it be believed (by the other person) that that statement is believed to be true (by the person making the statement).*

(L9) is formulated in  $L_0^U$  as follows.

**Definition 3.10 (L9)**  $L9_{ab}(\sigma) \stackrel{def}{=} utter_{ab}(\sigma) \wedge B_a \neg \sigma \wedge I_a B_b \sigma \wedge I_a B_b B_a \sigma \wedge I_a B_b I_a B_b B_a \sigma$ .

(L9) satisfies the statement condition, the addressee condition ( $utter_{ab}(\sigma)$ ), and the untruthfulness condition ( $B_a \neg \sigma$ ). (L9) requires three different intentions:  $a$  intends to (i) make  $b$  believe the untruthful statement  $\sigma$  ( $I_a B_b \sigma$ ), (ii) make  $b$  believe that  $a$  believes  $\sigma$  ( $I_a B_b B_a \sigma$ ), and (iii) make  $b$  believe that  $a$  intends (ii) ( $I_a B_b I_a B_b B_a \sigma$ ). Like (L7), (L9) requires the believed truthfulness condition. Hence, Mahon rejects (L9) by the same reason as (L7).

### 3.10 Carson's Definition

Carson [3] defines lying in terms of the warranty of the statement. It is defined in [14] as follows:

(L10) *To lie (to another person) is: to make a not-believed-true, and false, statement (to another person), in a context in which the truth of the statement is thereby warranted (by the person making the statement) (to the other person), (the person making the statement) not believing that the truth of the statement is not being warranted (by the person making the statement) (to the other person).*

(L10) is formulated in  $L_0^U$  as follows.

**Definition 3.11 (L10)**  $L10_{ab}(\sigma) \stackrel{def}{=} utter_{ab}(\sigma) \wedge \neg B_a \sigma \wedge \neg \sigma \wedge B_b B_a \sigma \wedge \neg B_a \neg B_b B_a \sigma$ .

(L10) satisfies the statement condition and the addressee condition ( $utter_{ab}(\sigma)$ ). On the other hand, (L10) does not require untruthfulness but requires falsity of the statement. Also, (L10) does not require an intention to deceive the addressee. The condition  $B_b B_a \sigma$  represents the situation that a speaker warrants the truth of his/her statement to the hearer. According to (L10),  $a$  can lie only if  $b$  believes that  $a$  believes  $\sigma$ . Mahon argues that "if one makes an untruthful and false statement to an audience, and if one intends to warrant the truth of that statement to one's audience, then one is lying, even if, unbeknownst to one, the context is such that one's audience does not take one to be warranting the truth of one's statement to them, and hence, one's audience does not believe one's statement to be true" [14]. Mahon then rejects (L10).

### 3.11 Fallis's Definition

We finally provide a definition by Fallis [8], which differs from most other definitions of lying given so far. Mahon provides it as follows:

(L11) *To lie (to another person) is: to make a believed-false statement (to another person) while believing that the context is one in which the norm 'Do not say what you believe to be false' is in effect.*



In the above, ‘Do not say what you believe to be false’ is called the *Gricean conversational norm of truthfulness* [11]. (L11) is formulated in  $L_0^U$  as follows.

**Definition 3.12 (L11)**  $L11_{ab}(\sigma) \stackrel{def}{=} utter_{ab}(\sigma) \wedge B_a \neg \sigma \wedge B_a B_b (utter_{ab}(\sigma) \supset \neg B_a \neg \sigma)$ .

(L11) satisfies the statement condition and the addressee condition ( $utter_{ab}(\sigma)$ ). (L11) also satisfies the untruthfulness condition ( $B_a \neg \sigma$ ), but does not satisfy the intention to deceive addressee condition. In (L11), the norm of truthfulness is represented by  $B_a B_b (utter_{ab}(\sigma) \supset \neg B_a \neg \sigma)$ . Thus, “unlike (L10), whether or not the context is one in which the truthfulness norm ‘Do not say what you believe to be false’ is in effect is determined entirely by the beliefs of the person making the statement” [14]. Thus, if one makes an untruthful statement to an audience while believing that the truthfulness norm is in effect, then one is lying even if one’s audience does not believe one’s statement to be true. With this respect, Mahon considers that (L11) is better than (L10). On the other hand, Mahon argues that (L11) has a problem when one “make(s) an untruthful statement, while believing that one is in a context in which the norm of truthfulness is in effect, and not intend that one’s untruthful statement be believed to be true” [14]. For instance, suppose that a policeman asks a drunken man “Where is your home?” and the drunk replies “On the moon”. The policeman then considers that the drunk is just kidding, rather than lying. With this reason, Mahon does not consider (L11) appropriate.

## 4 Comparison between Different Definitions

As addressed in Section 3, Mahon considers four necessary conditions for lying: statement condition, untruthfulness condition, addressee condition, and intention to deceive addressee condition, as well as other (not necessarily required) conditions such as falsity condition (L4, etc) and believed truthfulness condition (L6\*, etc). Here we introduce five additional conditions that are considered natural to be satisfied by the definition of lying.<sup>2</sup>

- Lying on valid sentences is impossible (**inability to lie on valid sentences**, or **inability- $\top$**  for short).
- Lying on contradictory sentences is impossible (**inability to lie on contradictory sentences**, or **inability- $\perp$**  for short).
- Lying on two sentences  $\sigma$  and  $\neg \sigma$  at the same time is impossible (**inability to lie on mutually complementary sentences**, or **inability- $\neg$**  for short).
- A liar is aware of his/her dishonest act (**awareness**).
- Lying to oneself leads to contradiction (**self-contradiction**).

We examine whether twelve definitions of lying satisfy the above five conditions. In what follows,  $LIE_{ab}(\sigma)$  means one of the definitions (L1), (L2), and (L5)–(L11). We prove propositions for  $LIE_{ab}(\sigma)$ , but the same proofs are applied for (L3) and (L4) that have the additional parameter  $\lambda$ .

<sup>2</sup> The four conditions, inability- $\top$ , inability- $\perp$ , awareness, and self-contradiction, are also considered in [17].

**Proposition 4.1 (inability- $\top$ )**  $\vdash LIE_{ab}(\top) \supset \perp$  holds if  $LIE_{ab}(\sigma)$  includes either  $B_a \neg \sigma$  or  $\neg B_a \sigma$  or  $\neg \sigma$ .

*Proof.* If  $LIE_{ab}(\sigma)$  includes either  $B_a \neg \sigma$  or  $\neg B_a \sigma$ ,  $LIE_{ab}(\top)$  implies  $B_a \perp$  that implies  $\neg B_a \top$  ( $\mathbf{D_B}$ ), while  $\top$  implies  $B_a \top$  ( $\mathbf{N_B}$ ). Contradiction. If  $LIE_{ab}(\sigma)$  includes  $\neg \sigma$ ,  $LIE_{ab}(\top)$  implies  $\perp$ .  $\square$

By Proposition 4.1, we can see that (L1), (L2), (L4), (L5), (L6), (L6\*), (L7), (L8), (L9), (L10), and (L11) satisfy the property of inability to lie on valid sentences. By contrast,  $L\mathfrak{I}_{ab}(\top, \lambda)$  does not imply  $\perp$ .

**Proposition 4.2 (inability- $\perp$ )**  $\vdash LIE_{ab}(\perp) \supset \perp$  holds if  $LIE_{ab}(\sigma)$  includes either  $I_a B_b \sigma$  or  $I_a B_b B_a \sigma$  or  $B_a B_b B_a \sigma$ .

*Proof.* If  $LIE_{ab}(\sigma)$  includes  $I_a B_b \sigma$ , then  $LIE_{ab}(\perp)$  implies  $I_a B_b \perp$ , while  $B_b \top$  implies  $\neg B_b \perp$  ( $\mathbf{D_B}$ ) that implies  $I_a \neg B_b \perp$  ( $\mathbf{N_I}$ ) then  $\neg I_a B_b \perp$  ( $\mathbf{D_I}$ ). Contradiction.

If  $LIE_{ab}(\sigma)$  includes  $I_a B_b B_a \sigma$ , then  $LIE_{ab}(\perp)$  implies  $I_a B_b B_a \perp$ . On the other hand,  $B_a \top$  implies  $\neg B_a \perp$  ( $\mathbf{D_B}$ ) that implies  $B_b \neg B_a \perp$  ( $\mathbf{N_B}$ ) then  $\neg B_b B_a \perp$  ( $\mathbf{D_B}$ ). Thus,  $I_a \neg B_b B_a \perp$  ( $\mathbf{N_I}$ ), so  $\neg I_a B_b B_a \perp$  ( $\mathbf{D_I}$ ) which contradicts  $I_a B_b B_a \perp$ .

If  $LIE_{ab}(\sigma)$  includes  $B_a B_b B_a \sigma$ , then  $LIE_{ab}(\perp)$  implies  $B_a B_b B_a \perp$ . On the other hand,  $B_a \top$  implies  $\neg B_b B_a \perp$  as above, which implies  $B_b \neg B_b B_a \perp$  ( $\mathbf{N_B}$ ) then  $\neg B_b B_b B_a \perp$  ( $\mathbf{D_B}$ ). Contradiction.  $\square$

By Proposition 4.2, we can see that (L1), (L5), (L6), (L6\*), (L7), (L8), (L9), and (L10) satisfy the property of inability to lie on contradictory sentences. On the other hand,  $L11_{ab}(\perp)$  implies  $B_a B_b (utter_{ab}(\perp) \supset \neg B_a \top)$ , hence  $B_a (B_b (utter_{ab}(\perp)) \supset B_b \neg B_a \top)$  and  $B_a B_b (utter_{ab}(\perp)) \supset B_a B_b \neg B_a \top$  ( $\mathbf{K_B}$ ). Since  $utter_{ab}(\perp)$  implies  $B_a B_b (utter_{ab}(\perp))$  ( $\mathbf{U_{BB}}$ ), it holds that  $B_a B_b \neg B_a \top$  by ( $\mathbf{MP}$ ). As  $B_a (B_b \neg B_a \top \supset \neg B_b B_a \top)$ , it holds that  $B_a \neg B_b B_a \top$  thereby  $\neg B_a B_b B_a \top$  ( $\mathbf{D_B}$ ). This contradicts  $B_a B_b B_a \top$  that is obtained from  $\top$  by iteratively applying ( $\mathbf{N_B}$ ). Hence,  $L11_{ab}(\perp)$  also imply  $\perp$ . By contrast,  $L2_{ab}(\perp)$ ,  $L3_{ab}(\perp, \lambda)$ , and  $L4_{ab}(\perp, \lambda)$  do not imply  $\perp$ .

**Proposition 4.3 (inability- $\neg$ )**  $\vdash LIE_{ab}(\sigma) \wedge LIE_{ab}(\neg \sigma) \supset \perp$  holds if  $LIE_{ab}(\sigma)$  includes either  $B_a \neg \sigma$  or  $\neg \sigma$ .

*Proof.* If  $LIE_{ab}(\sigma)$  includes  $B_a \neg \sigma$ ,  $LIE_{ab}(\sigma) \wedge LIE_{ab}(\neg \sigma)$  implies  $B_a \neg \sigma \wedge B_a \sigma$ .  $B_a \neg \sigma$  implies  $\neg B_a \sigma$  ( $\mathbf{D_B}$ ). Contradiction. If  $LIE_{ab}(\sigma)$  includes  $\neg \sigma$ ,  $LIE_{ab}(\sigma) \wedge LIE_{ab}(\neg \sigma)$  implies  $\neg \sigma \wedge \sigma$ . Contradiction.  $\square$

By Proposition 4.3, we can see that (L1), (L2), (L4), (L5), (L6), (L6\*), (L7), (L9), (L10) and (L11) satisfy the property of inability to lie on mutually complementary sentences. On the other hand, (L3) and (L8) do not satisfy the property.

**Proposition 4.4 (awareness)**  $\vdash LIE_{ab}(\sigma) \supset B_a (LIE_{ab}(\sigma))$  holds for any sentence  $\sigma$  for the definitions of (L1), (L2), (L3), (L6), (L6\*), (L7), (L8), (L9), and (L11).

*Proof.* The result holds by each definition and the axioms ( $\mathbf{U_{IB}}$ ), ( $\mathbf{4_B}$ ), and ( $\mathbf{4_{IB}}$ ).  $\square$

(L4), (L5) and (L10) do not satisfy the property because they contain the condition of falsity, that is,  $\neg \sigma$  does not imply  $B_a \neg \sigma$ .

**Table 1.** Comparison of twelve definitions

	L1	L2	L3	L4	L5	L6	L6*	L7	L8	L9	L10	L11
statement*		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
addressee*		(✓)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
untruthful*	✓	✓			✓	✓	✓	✓	(✓)	✓		✓
intention*	✓		✓	✓	✓	✓	✓	✓		✓		
believed truthful							(✓)	✓		✓		
falsity				✓	✓						✓	
inability- $\top$ †	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
inability- $\perp$ †	✓				✓	✓	✓	✓	✓	✓	✓	✓
inability- $\neg$ †	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
awareness†	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓
self-contradiction†	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

\*: necessary conditions by Mahon; †: conditions considered in this paper.

**Proposition 4.5 (self-contradiction)**  $\vdash LIE_{aa}(\sigma) \supset \perp$  holds for any sentence  $\sigma$  if  $LIE_{ab}(\sigma)$  includes either  $B_a \neg \sigma \wedge I_a B_b \sigma$  or  $B_a \neg \sigma \wedge I_a B_a B_a \sigma$  or  $\neg B_a \sigma \wedge B_b B_a \sigma$ .

*Proof.* If  $LIE_{ab}(\sigma)$  includes  $B_a \neg \sigma \wedge I_a B_b \sigma$ , then  $LIE_{aa}(\sigma)$  implies  $B_a \neg \sigma \wedge I_a B_a \sigma$ .  $B_a \neg \sigma$  implies  $\neg B_a \sigma$  ( $\mathbf{D_B}$ ), which implies  $I_a \neg B_a \sigma$  ( $\mathbf{N_I}$ ). On the other hand,  $I_a B_a \sigma$  implies  $\neg I_a \neg B_a \sigma$  ( $\mathbf{D_I}$ ). Contradiction.

If  $LIE_{ab}(\sigma)$  includes  $B_a \neg \sigma \wedge I_a B_a B_a \sigma$ , then  $LIE_{aa}(\sigma)$  implies  $B_a \neg \sigma \wedge I_a B_a B_a \sigma$ .  $B_a \neg \sigma$  implies  $\neg B_a \sigma$  ( $\mathbf{D_B}$ ) that implies  $B_a \neg B_a \sigma$  ( $\mathbf{5_B}$ ) then  $\neg B_a B_a \sigma$  ( $\mathbf{D_B}$ ). Thus,  $I_a \neg B_a B_a \sigma$  ( $\mathbf{N_I}$ ), so  $\neg I_a B_a B_a \sigma$  ( $\mathbf{D_I}$ ) which contradicts  $I_a B_a B_a \sigma$ .

If  $LIE_{ab}(\sigma)$  includes  $\neg B_a \sigma \wedge B_b B_a \sigma$ , then  $LIE_{aa}(\sigma)$  implies  $\neg B_a \sigma \wedge B_a B_a \sigma$ .  $\neg B_a \sigma$  implies  $B_a \neg B_a \sigma$  ( $\mathbf{5_B}$ ), which implies  $\neg B_a B_a \sigma$  ( $\mathbf{D_B}$ ). Contradiction.  $\square$

By Proposition 4.5, we can see that (L1), (L3), (L4), (L5), (L6), (L6\*), (L7), (L9), and (L10) satisfy the property of self-contradiction. It is easy to see that (L8) and (L11) also satisfy the property, but (L2) is not.

Table 1 compares twelve definitions of lying from the viewpoint of satisfaction of various conditions explained so far. In the table,  $\checkmark$  means satisfaction of each condition, so (L1), for instance, satisfies conditions of untruthfulness, intention to deceive addressee, inability- $\top$ , inability- $\perp$ , inability- $\neg$ , awareness, and self-contradiction. ( $\checkmark$ ) means that the condition is included as a disjunct.

By the table, we can observe that (L7) and (L9) satisfy most conditions. (L5), (L6), and (L6\*) follow them. (L5) does not satisfy the awareness condition, which appears unintuitive. Mahon argues that believed truthfulness is often too strong. As a result, Mahon concludes that (L6) and (L6\*) are best definitions among the twelve definitions. Since (L6) and (L6\*) also satisfy the additional five conditions, inability- $\top$ , inability- $\perp$ , inability- $\neg$ , awareness, and self-contradiction, we also support these two definitions.

Figure 1 shows relationship between different definitions of lying. In the figure,  $X \rightarrow Y$  represents that  $X$  implies  $Y$  under the logic  $L_0^U$ . In the figure, we can observe that (L4) and (L9) are relatively stronger, while (L2) is relatively weaker. On the other hand, (L1), (L8) or (L10) has no implication relations with other definitions.

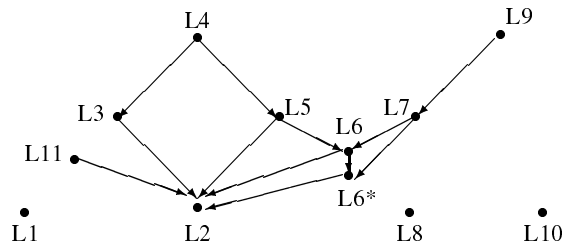


Fig. 1. Relationship between twelve definitions

## 5 Conclusion

In this paper, we gave logical definitions of lying and analyzed their formal properties. The results of this paper provide a formal ground for Mahon's informal argument that Kuper's definition (L6) and its modification (L6\*) are most intuitive. The logic of belief and intention used in this paper is simple but expressive for abstracting the act of lying. The five new conditions that logically justify the act of lying, together with four necessary conditions by Mahon that empirically support lying, serve as criteria for judging whether yet another definition of lying is appropriate or not.

## References

1. Bonatti, P. A., Kraus, S. and Subrahmanian, V. S.: Foundations of secure deductive databases. *IEEE Transactions on Knowledge and Data Engineering* 7(3), 406–422 (1995)
2. Bok, S.: *Lying: Moral Choice in Public and Private Life*. Random House (1978)
3. Carson, T. L.: The definition of lying. *Noûs* 40(2), 284–306 (2006)
4. Carson, T. L.: *Lying and deception: theory and practice*. Oxford University Press (2010)
5. Chisholm, R. M. and Feehan, T. D.: The intent to deceive. *Journal of Philosophy* 74(3), 143–159 (1977)
6. Coleman, L. and Kay, P.: Prototype semantics: the English word lie. *Language* 57(1), 26–44 (1981)
7. Colombetti, M.: A modal logic of intentional communication. *Mathematical Social Sciences* 38, 171–196 (1999)
8. Fallis, D.: What is lying? *Journal of Philosophy* 106(1), 29–56 (2009)
9. Firozabadi, B. S. and Jones, A. J. I.: On the characterisation of a trusting agent – aspects of a formal approach. In: C. Castelfranchi and Y. H. Tan (eds.), *Trust and Deception in Virtual Societies*, Kluwer Academic Publishers, pp. 157–168 (2001)
10. Frankfurt, H. G.: The faintest passion. In: *Necessity, Volition and Love*. Cambridge University Press (1992)
11. Grice, H. P.: *Studies in the Ways of Words*. Harvard University Press (1989)
12. Halpern, J. and Moses, J.: A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence* 54, 349–379 (1992)
13. Kupfer, J.: The moral presumption against lying. *Review of Metaphysics* 36, 103–126 (1982)
14. Mahon, J. E.: Two definitions of lying. *J. Applied Philosophy* 22(2), 211–230 (2008)
15. Mahon, J. E.: The definition of lying and deception. *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/lying-definition/> (2008)

16. O'Neill, B.: A formal system for understanding lies and deceit. *Jerusalem Conference on Biblical Economics* (2003)
17. Sakama, C., Caminada, M. and Herzig, A.: A logical account of lying. In: *Proc. 12th European Conference on Logics in Artificial Intelligence, Lecture Notes in Artificial Intelligence* 6341, pp. 286–299 (2010)
18. Shibles, W.: *Lying: A Critical Analysis*. The Language Press (1985)
19. Simpson, D.: Lying, liars and language. *Philosophy and Phenomenological Research* 52: 623–629 (2007)
20. Sklar, E., Parsons, S. and Davies, M.: When is it okay to lie? A simple model of contradiction in agent-based dialogues. *Argumentation in Multi-Agent Systems, Lecture Notes in Computer Sciences* 3366, pp. 251–261, Springer (2005)
21. Vrij, A.: *Detecting lies and deceit: the psychology of lying and the implications for professional practice*. John Wiley & Sons (2000)