Performatives in a Rationally Based Speech Act Theory[*]

Philip R. Cohen
Artificial Intelligence Center
and
Center for the Study of Language and Information
SRI International
333 Ravenswood Ave.
Menlo Park, CA 94025
and
Hector J. Levesque[†]
Department of Computer Science
University of Toronto

Abstract

A crucially important adequacy test of any theory of speech acts is its ability to handle performatives. This paper provides a theory of performatives as a test case for our rationally based theory of illocutionary acts. We show why "I request you..." is a request, and "I lie to you that p" is self-defeating. The analysis supports and extends earlier work of theorists such as Bach and Harnish [1] and takes issue with recent claims by Searle [10] that such performative-as-declarative analyses are doomed to failure.

1 Introduction

There is something special about performative sentences, sentences such as "I promise to return": uttering them makes them true. How and when is this possible? Not all verbs can be uttered in the first-person present tense and thereby make the sentence true. In general, the successful verbs seem to correspond to those naming illocutionary acts, but not

to perlocutionary ones such as "frighten." But, even some illocutionary verbs cannot be used performatively: e.g., "I lie to you that I didn't steal your watch" is self-defeating [12]. So, which verbs can be use performatively, and in Searle's words [10], "how do performatives work?"

Any theory of illocutionary acts needs to provide a solution to questions such as these. But, such questions are not merely of theoretical interest. Natural language database question-answering systems have been known to receive performative utterances [14], dialogue systems that recognize illocutionary acts (e.g., [6]) will need to infer the correct illocutionary force to function properly, dialogue translation systems [5] will have to cope with markers of illocutionary force that function performatively (e.g., sentence final particles in Japanese), and proposals for "agent-oriented programming languages" [7, 13], as well as Winograd and Flores' [15] COORDINATOR system, are based on performative communication. For all these systems, it is important to understand the semantics and pragmatics of such communicative acts, especially their intended effects. To do so, one needs a full theory of illocutionary acts, and a formal theory that predicts how utterances can be made true by uttering them.

The currently accepted theory of performatives is that they are in fact assertions, hence true or false, and additionally constitute the performance of the named illocutionary act, in the same way as an indirect reading of an illocutionary act is obtained from the direct illocutionary act. That is, the named illocutionary act is derived from the assertion as an indirect speech act. The most compelling defense of this performative-as-assertion analysis that we are aware is that of Bach and Harnish [1], who address many of the linguistic phenomena discussed by Sadock [9], but who, we believe, have misanalyzed indirect speech acts. However, in a recent paper, Searle [10] forcefully criticizes the performative-as-assertion approach on the following grounds:

- Assertions commit the speaker to the truth of what is asserted

- Performative statements are self-referential

- "An essential feature of any illocutionary act is the intention to perform that act"

Searle claims that accounts based on self-referential assertions are "doomed to failure" because one cannot show that being committed to having the intention to be performing the named illocutionary act entails that one in fact has that intention. Moreover, he questions that one should derive the named illocutionary act from an assertion, rather than vice-versa. However, Searle has imparted into Bach and Harnish's theory his notion of assertions as commitments to the truth without providing a precise analysis of commitment. What may be doomed to failure is any attempt to base an analysis of performatives on such a theory of assertions.

This paper provides a formal analysis of performatives that treats them as declarative utterances, not initially as assertions, does not succumb to Searle's criticisms, and does not require an entirely new class of illocutionary acts (the "declarations") as Searle and Vanderveken [12] have proposed. The analysis is offered as another adequacy criterion for our theory of illocutionary acts. That theory, more fully explicated in [3], is based on an analysis of the individual rational balance agents maintain among their beliefs, goals, intentions, commitments, and actions [2].

As desiderata for the theory of performatives, we demonstrate that the analysis meets two properties:

- A sincere utterance of "I request you to open the door" is both a request and an assertion, yet neither illocutionary act characterization is derived from the other.

- "I lie that the door is open" is self-defeating.

Briefly, the ability to capture performatives is met almost entirely because such utterances are treated as indicative mood utterances, and because illocutionary acts are defined as attempts. Since attempts depend on the speaker's beliefs, goals, and intentions, and these mental states are introspectable in our theory, if a speaker sincerely says, for example, "I request you to open the door," he must believe he did the act with the requisite beliefs, goals, and intentions. Hence, the utterance is a request.

To meet the desiderata we need first to present, albeit briefly, the theory of rational interaction, the treatment of declarative mood utterances, and then the illocutionary act definitions for requesting and asserting. Finally, we combine the various analyses to analyze performatives.

2 Abbreviated theory of rational action

Below, we give an abbreviated description of the theory of rational action upon which we erect a theory of intention. The theory is cast in a modal logic of belief, goal, action, and time. Further details of this logic can be found in [2].

2.1 Syntax

The language we use has the usual connectives of a first-order language with equality, as well as operators for the propositional attitudes and for talking about sequences of events: (BEL x p) and (GOAL x p) say that p follows from x's beliefs or goals (a.k.a choices) respectively; (AGT x e) says that x is the only agent for the sequence of events e; $e_1 \leq e_2$ says that e_1 is an initial subsequence of e_2; and finally, (HAPPENS a) and (DONE a) say that a sequence of events describable by an action expression a will happen next or has just happened, respectively. Versions of HAPPENS and DONE specifying the agent (x) are also defined.

An action expression here is built from variables ranging over sequences of events using the constructs of dynamic logic [4]: a;b is action composition; a|b is nondeterministic choice; a||b is concurrent occurrence of a and b; p? is a test action; and finally, a^* is repetition. The usual programming constructs such as IF/THEN actions and WHILE loops, can easily be formed from these. Because test actions occur frequently in our analysis, yet create considerable confusion, read p?;a as "action a occurring when p holds," and for a;p?, read "action a occurs after which p holds." We use e as a variable ranging over sequences of events, and a and b for action expressions.

We adopt the following abbreviations and domain predicates.

(BEFORE a p) $\stackrel{\text{def}}{=}$ (DONE p?;a)[1]

(AFTER a p) $\stackrel{\text{def}}{=}$ (HAPPENS a;p?)

$\diamond$p $\stackrel{\text{def}}{=}$ $\exists$e (HAPPENS e;p?).

(LATER p) $\stackrel{\text{def}}{=}$ $\neg$p $\wedge$ $\diamond$p.

$\square$p $\stackrel{\text{def}}{=}$ $\neg\diamond\neg$p.

(PRIOR p q) $\stackrel{\text{def}}{=}$ $\forall$c (HAPPENS c;q?) $\supset$ $\exists$a (a $\leq$ c) $\wedge$ (HAPPENS a;p?).
The proposition p will become true no later than q.

(KNOW x p) $\stackrel{\text{def}}{=}$ p $\wedge$ (BEL x p).

(IMPERATIVE s) means that sentence s is an imperative.

(DECLARATIVE s) means that sentence s, a string of words, is a declarative.

(MAIN-VERB s v), (TENSE s tense), (COMPLEMENT s s'), (D-OBJECT s np), (SUBJECT
s np), are all syntactic predicates intended to have the obvious meanings.[2]

(TRUE s e) means that sentence s is true with respect to some event sequence e
(which we will say has just been done.)

(REFERS np x e) means that noun phrase np refers to thing x with respect to event
e.

(FULFILL-CONDS s e e') means that event e fulfills the satisfaction conditions, rel-
ative to event e', that are imposed by sentence s.[3] For example, if s is "wash the
floor," e would be a floor-washing event.

2.2   Assumptions

The model we are developing embodies various assumptions constraining beliefs and choices
(goals). First, BEL has a "weak S5" semantics, and GOAL has a "system K" semantics.[4]
Among the remaining assumptions, the following will be used in this paper.[5]

Beliefs imply choice: $\models$ (BEL x p) $\supset$ (GOAL x p).
    This means that agents choose amongst worlds that are compatible with their beliefs.

Goals are known: $\models$(GOAL x p) $\equiv$ (BEL x (GOAL x p)).

Memory: $\models$ (DONE x (BEL x p)?;e) $\equiv$ (BEL x (DONE x (BEL x p)?;e)).
    That is, agents remember what their beliefs were.

---

[1]This differs from the BEFORE relation described in [3], which is here labelled PRIOR.

[2]Feel free to substitute your favorite syntactic predicates.

[3]TRUE REFERS, and FULFILL-CONDS are just placeholders for semantic theories of truth, reference, and
the meanings of imperatives, respectively. Their last event arguments would be used only in the interpretation
of indexicals.

[4]See other work of ours [2] for a full model theory.

[5]In other words, we only deal with semantic structures where these propositions come out true.

3 Individual Commitments and Intentions

To capture one grade of commitment that an agent might have toward his goals, we define a persistent goal, P-GOAL, to be one that the agent will not give up until he thinks certain conditions are satisfied. Specifically, we have

Definition 1 (P-GOAL x p q) $\stackrel{\text{def}}{=}$
(1) (BEL x ¬p) ∧
(2) (GOAL x (LATER p)) ∧
(3) [KNOW x   (PRIOR [(BEL x p)∨(BEL x □¬p)∨(BEL x ¬q)]
                      ¬[GOAL x (LATER p)])]].

That is, the agent x believes p is currently false, chooses that it be true later, and knows that before abandoning that choice, he must either believe it is true, believe it never will be true, or believe q, an escape clause (used to model subgoals, reasons, etc.) is false.

Intention is a species of persistent goal. We analyze two kinds of intentions, those to do actions and those to achieve propositions. Accordingly, we define INTEND$_1$ and INTEND$_2$ to take action expressions and propositions as arguments, respectively.

Definition 2 *Intention:*
(INTEND$_1$ x a q) $\stackrel{\text{def}}{=}$ (P-GOAL x [DONE x (BEL x (HAPPENS a))?;a] q).

(INTEND$_2$ x p q) $\stackrel{\text{def}}{=}$ (P-GOAL x ∃e[HAPPENS x
                                    (BEL x ∃e' (HAPPENS x e';p?))?;e;p?] q)

Intending to do an action a or achieve a proposition p is a special kind of commitment (i.e., persistent goal) to having done the action a or having achieved p.[6] However, it is not a simple commitment to having done a or e;p? for that would allow the agent to be committed to doing something accidentally or unknowingly. Instead, we require that the agent be committed to arriving at a state in which he believes he is about to do the intended action next.

This completes a brief discussion of the foundational theory of intention and commitment. Next, we proceed to define the more specific concepts needed for analyzing communicative action.

4 Utterance Events

We begin the analysis of utterance events by adopting a Gricean correlation of an utterance's features (e.g., syntactic mood or sentence-final particles in Japanese) with the speaker's mental state, termed a "core attitude" in [3, 8]. Very roughly, a declarative utterance s will be correlated with the speaker's believing the uttered sentence is true, and an imperative utterance will be correlated with the speaker's wanting the addressee to do some action that fulfills the conditions imposed by the sentence. Let us notate these correlations as:

---

[6]For simplicity, we omit here one condition from the definition of INTEND$_2$ in [2].

DECLARATIVE $\Rightarrow$ (BEL x (TRUE s e))

IMPERATIVE $\Rightarrow$ (GOAL x $\diamond \exists e'$ (DONE y e') $\wedge$ (FULFILL-CONDS s e' e)

We formalize this notation below.

Someone who thinks he is observing an utterance event will come to believe the speaker is in the correlated mental state, unless he has other beliefs to the contrary. For example, if the observer thinks the speaker is lying, he believes that the speaker does not believe the uttered sentence is true. But, because he may think the speaker takes himself to be especially convincing, the observer may still believe that the speaker thinks the observer is deceived. Hence, he would believe the speaker thinks that he thinks the speaker believes p.

This type of reasoning can continue to further levels. In general, if an utterance is produced when there are no countervailing observer beliefs at a certain level of nesting, then the result will be, at the given level of nesting, that the speaker is taken to be in the correlated mental state [8]. To be able to state such conditions, we need to be able to refer easily to what a person x believes about what y believes about what x believes etc., to arbitrary depths. To do so, we use the notion of ABEL.

Definition 3 (ABEL n x y p) $\stackrel{\text{def}}{=}$ $\underbrace{(\text{BEL x (BEL y (BEL x } \ldots (\text{BEL x p}}_{n} \underbrace{) \ldots)}_{n}$

That is, ABEL characterizes the nth alternating belief between x and y that p, built up "from outside in," i.e, starting with x's belief that p. On this basis, one can define unilateral mutual belief — what one agent believes is mutually believed — as follows.

Definition 4 (BMB x y p) $\stackrel{\text{def}}{=}$ $\forall$n(ABEL n x y p)

In other words, (BMB x y p) is the infinite conjunction (BEL x p) $\wedge$ (BEL x (BEL y p)) $\wedge \ldots$ Finally, we define mutual belief and mutual knowledge as follows.

Definition 5 (MB x y p) $\stackrel{\text{def}}{=}$ (BMB x y p) $\wedge$ (BMB y x p).

(MK x y p) $\stackrel{\text{def}}{=}$ p $\wedge$ (MB x y p).

Utterance events can produce effects at any (or no) level of alternating belief. For example, the speaker may not be trying to communicate anything to an intended observer. Illocutionary acts will be defined to require that the speaker intend to produce BMBs, that is, beliefs at all levels. In what follows, it is important to keep in mind the distinction between utterance events and full-blooded communicative acts.

4.1 Notation for Describing Utterance Events

We now provide a formal notation for this correlation of utterance form and the speaker's mental state as a kind of default axiom (cf. [8]). First, we specify who is speaking (spkr), who is observing (obs, which includes the speaker and addressee, but also others), who is

6

being addressed (addr), and what kind of sentence (s) has been spoken (indicated by Φ). We shall assume that everyone knows that a given utterance is of a given syntactic type (e.g., declarative), that speakers and addressees are observers, and that observers are known by all to be observing.[7]

Definition 6 $\Phi \Rightarrow \alpha \stackrel{def}{=}$

$\forall$ spkr, obs, addr, e, s, n
  (KNOW obs (DONE spkr e) $\wedge$ (UTTER spkr addr s e) $\wedge$ (Φ s)) $\wedge$
$\sim$(ABEL n obs spkr
    (BEFORE e
      $\sim$(GOAL spkr
        [AFTER e
          (KNOW addr (BEFORE e α))]))) $\supset$
  (ABEL n obs spkr
   (BEFORE e
     α $\wedge$ (GOAL spkr
        [AFTER e
          (KNOW addr (BEFORE e α))])))

That is, $\Phi \Rightarrow \alpha$ is an abbreviation for a quantified implication roughly to the effect that if an observer obs knows that e was just done, where e was an uttering to addressee addr of a sentence s in syntactic mood Φ, and obs does not believe that e was done when the speaker did not want the addressee to come to know that the core speaker-attitude α associated with utterances of that type held, then obs believes that the speaker in fact wanted the addressee to know that α, and so he, the observer, believes that α held just prior to the utterance. The notation states that at each level of alternating belief for which the antecedent holds, so does the consequent. The symbol '$\Rightarrow$' can now be understood as a textual-replacement "macro" operator.

Since these correlations are of the form $\forall$nP(n) $\supset$ Q(n), they imply $\forall$nP(n) $\supset$ $\forall$nQ(n). As we quantify over the positive integers indicating levels of alternative belief, we can derive the conclusion that under certain circumstances, addr thinks it is mutually believed (in our notation, BMB'ed) that the speaker spkr wants addr to know α was true.

Notice that right after the utterance, we are concerned with what mental state the observer thinks the speaker chose to bring about in the observer with that utterance. That is, the condition on utterance events involves the speaker's wanting to get the observer to know something, Without this temporal dimension, our performative analysis would fail. The analysis of performatives will say that after having uttered such a sentence, or while uttering it, the speaker believes he has just done or is doing the named illocutionary act. Typically, prior to uttering a performative, the speaker has not just performed that speech act, and so he would believe his having just done so is false. So, if the condition on utterance events in Domain Axiom 1A involved only what the speaker believed or wanted to be true prior to the

---

[7]The case of unseen observers is straightforward, but omitted here.

utterance, rather than after, all performatives would fail to achieve the observer's coming to believe anything.

We can now state the correlation between utterance form and a speaker's mental state as a domain axiom.

Domain Axiom 1 *Declaratives and Imperatives:*

$A.$ $\models$ DECLARATIVE $\Rightarrow$ (BEL spkr (TRUE s e))

$B.$ $\models$ IMPERATIVE $\Rightarrow$ (GOAL x $\Diamond \exists e'$ (DONE y e') $\wedge$ (FULFILL-CONDS s e' e)

Below, we present our definitions of illocutionary acts. Further justification can be found in [3].

5 Illocutionary Acts as Attempts

Searle [11] points out that an essential condition for requesting is that the speaker be attempting to get the addressee to perform the requested action. We take this observation one step further and define all illocutionary acts as attempts, hence defined in terms of the speaker's mental states. Attempts involve both types of goal states, GOAL (merely chosen) and INTEND (chosen with commitment), as noted below.

Definition 7 {ATTEMPT x e p q t_1} $\stackrel{def}{=}$

t_1 ?;[(BEL x $\sim$p $\wedge$ $\sim$q) $\wedge$
    (INTEND$_1$ x t_1?;e;p? (GOAL x $\Diamond$q)) $\wedge$
    (GOAL x $\Diamond$q)]?; e

That is, an attempt to achieve q via p is a complex action expression in which x is the agent of event e at time t_1, and prior to e the agent believes p and q are both false, chooses that q should eventually be true, and intends, relative to that choice, that e should produce p. So, q represents some ultimate goal that may or may not be achieved by the attempt, while p represents what it takes to make an honest effort.

5.1 Definitions of Request and Assert

To characterize a request or, for that matter, any illocutionary action, we must decide on the appropriate formulas to substitute for p and q in the definition of an attempt. We constrain illocutionary acts to be those in which the speaker is committed to understanding, that is, to achieving a state of BMB that he is in a certain mental state. Below is a definition of a speaker's requesting an addressee to achieve p.

Definition 8 {REQUEST spkr addr e p t_1} $\stackrel{def}{=}$

{ATTEMPT spkr e

8

[BMB addr spkr (BEFORE e

               (GOAL spkr $\Diamond$p $\wedge$

                   [AFTER e (INTEND$_2$ addr p [(GOAL spkr $\Diamond$p) $\wedge$

                                        (HELPFUL addr spkr)])])]]

$\exists$e$'$ (DONE addr e$'$;p?)

t$_1$}

That is, event e is a request at time t$_1$ if it is an attempt at that time to get the addressee to achieve some condition p while being committed to making public that the speaker wanted: first, that p eventually be achieved; and second, that the addressed party should intend to achieve it relative to the speaker's wanting it achieved and relative to the addressee's being helpfully disposed towards the speaker.

The illocutionary act of asserting will be defined as an attempt to make the speaker's believing the propositional content mutually believed.

Definition 9 {ASSERT spkr addr e p t$_1$} $\stackrel{\text{def}}{=}$

  {ATTEMPT spkr addr e

    [BMB addr spkr

      (BEFORE e [GOAL spkr

               (AFTER e

                     [KNOW addr (BEFORE e (BEL spkr p))])])]]

    [BMB addr spkr (BEFORE e (BEL spkr p))]

  t$_1$}

More precisely, assertions at time t$_1$ are defined as attempts in which to make an "honest effort," the speaker is committed to getting the addressee to believe that it is mutually believed that the speaker wanted prior to the utterance that the addressee would come to know that the speaker believed p held then. That is, just like a request, an assertion makes public that the speaker wants the addressee to know what mental state he was in. Although he is committed to that, what the speaker has chosen to achieve is not merely to make public his *goal* that the addressee know what mental state he was in, but to make public that he was in fact in that state of believing p. For an INFORM, the speaker would choose to achieve (KNOW addr p).

6 Performatives

To illustrate how performatives work, we show when both assertions and requests can be derived from the utterance of the performative "I request you to <do act>." The important point to notice here is that we have not had to add to our machinery; performative utterances will be treated exactly as declarative utterances, with the exception that the content of the utterance will make reference to an utterance event.

6.1 Request Reports

Let us characterize the truth conditions of the family of declarative sentences "x requests y to ⟨imperative sentence s'⟩." Let s be such a sentence. Let α be $\exists e_1 (\text{DONE y } e_1) \wedge (\text{FULFILL-CONDS s' } e_1 \text{ e})$. We ignore most syntactic considerations and indexicality for reasons of space.

Domain Axiom 2 *Present tense requests*
$\models \forall$x, y, e, t_1, (DONE t_1?;e) $\wedge$ (SUBJECT s x) $\wedge$ (D-OBJECT s y) $\wedge$
(REFERS x x e) $\wedge$ (REFERS y y e) $\supset$
(TRUE s e) $\equiv$ (DONE x {REQUEST x y e α t_1})

That is, if event e was just done and the sentence s was a present tense declarative sentence whose main verb is "request," whose subject x refers to person x, and whose direct object y refers to person y, then the sentence is true iff x just requested the addressee y to fulfill the conditions of imperative sentence s'. A bare present (or present progressive) tense sentence is true when the event being described is contemporaneous with the event of uttering it.[8] This definition applies equally well to "John requests Mary to ..." as it does when I utter "I request you to ..." For the former, such sentences are likely to be narrations of ongoing events.[9] For the latter, the event that is happening that makes the utterance true is the speaker's uttering of the sentence.

By our definition of request, for x to request y to achieve p, x has to attempt to get y to do some action intentionally to fulfill the sentence s', by making that goal mutually believed between them. Thus, to say x requested y to do something is only to say that x had the right beliefs, goals, and intentions.

6.2 Performatives Used as Requests

Next, we treat performative sentences as declaratives. This means that the effects of uttering them are described by Domain Axiom 1A. We sketch below a proof of a general theorem regarding performative requests, with s being the declarative sentence "I request you to ⟨imperative sentence s_1⟩, and α being $\exists e_1 (\text{DONE addr } e_1) \wedge (\text{FULFILL-CONDS } s_1 \text{ } e_1 \text{ e})$. We take the uttering of a sentence to be a unitary utterance event.

Theorem 1 *A Performative Request*
$\models \forall$ spkr, addr, e, n, t_1,
(MK spkr addr (DONE spkr t_1?;e) $\wedge$
            (UTTER spkr addr e s)) $\wedge$
(BEFORE t_1?;e
   (GOAL spkr [AFTER t_1?;e

---

[8]Searle [10] correctly points out that performatives can be uttered in the passive, and in the first-person plural.

[9]We are ignoring the habitual reading of bare present tense sentences because we do not have a semantics for them.

(KNOW addr

    [BEFORE $t_1?;e$ (BEL spkr (TRUE s e))]])]])) $\supset$

(DONE {REQUEST spkr addr e α t_1})

That is, we need to show that if the sentence "I request you to <imperative sentence>" has just been uttered at time t_1 sincerely, i.e., when the speaker wanted the addressee to know that he believed the sentence was true, then a direct request has taken place at t_1.

Proof sketch: Essentially, one invokes the domain axiom for declaratives at the first level of ABEL, entailing that the speaker believes that he believes that he has just done a REQUEST. Then, one expands the definition of REQUEST into an ATTEMPT, and then into its parts. The definition of ATTEMPT is based on BEL, GOAL and INTEND, the first two of which are obviously introspectable. That is, if one believes one has them one does, and vice-versa. Hence, by the memory assumption, the speaker actually had them prior to the utterance. More critically, intending to act at time t_1 is also introspectable at time t_1 because agents know what they are doing at the next instant and because there is no time to drop their commitment [2]. Thus, one can repackage these mental states up into an ATTEMPT and then a REQUEST.

6.3   Performatives Used as Assertions

We have shown that the speaker of a sincere performative utterance containing an illocutionary verb has performed the illocutionary act named by that verb. Under somewhat stronger conditions, we can also prove that the speaker has made an assertion. As before, let s be "I request you to <imperative sentence>."

Theorem 2 *Performatives Used as Assertions*

$\models \forall$ spkr, addr, e, n, t_1,

(MK spkr addr (DONE spkr $t_1?;e$) $\wedge$ (UTTER spkr addr e s)) $\wedge$

[BEFORE e (BEL spkr

        [AFTER e

         $\forall$n~(ABEL n addr spkr

            (BEFORE e

               ~(GOAL spkr

                [AFTER e

                 (KNOW addr [BEFORE e (BEL spkr (TRUE s e))]] $\supset$

(DONE spkr {ASSERT spkr addr (TRUE s e) t_1})

This default condition says that before the utterance, the speaker believed there would be no addressee belief after the utterance event (at any level n) to the effect that prior to that event the speaker did not want the addressee to come to know that the speaker believed (TRUE s e). Given Domain Axiom 1A, and the fact that BEL entails GOAL, this suffices to entail the definition of assertion. Notice that whereas requesting was derived in virtue of the content of the utterance, an assertion was derived by default assumptions regarding lack of belief in the speaker's insincerity.

11

7 'Lie' is not a performative

Some illocutionary verbs such as "lie," "hint," "insinuate," cannot be achieved performatively. The following analysis shows a general model for why such verbs naming covert acts cannot be performatively achieved.

A reasonable definition of lying is the following complex action:

Definition 10 {LIE spkr addr e p} $\stackrel{\text{def}}{=}$ (BEL spkr $\sim$p)?;{ASSERT spkr addr e p t_1}

That is, a lie is an assertion performed when the speaker believes the propositional content is false.

For "I lie to you that the door is open" to be a successful performative utterance, it would have to be true that the utterance is a lie. We would have to show that the uttering of that declarative sentence results in a lie's having been done. More generally, we provide a putative statement of the truth conditions of "x lies to y that <declarative sentence s'> ." Call the main sentence s.

Domain Axiom 3 *Supposed Truth Conditions for Performative Lying*
$\models \forall$e, x, y, t_1, (DONE t_1?;e) $\wedge$ (REFERS x x e) $\wedge$ (REFERS y y e) $\supset$

(TRUE s e) $\equiv$ (DONE {LIE x y e (TRUE s' e) t_1})

That is, if s and s' are declarative sentences of the appropriate syntactic form, x refers to x and y refers to y, then s is true iff in performing it at time t_1, x was lying that sentence s' is true.

So we can prove the following. Let the sentence s be "I lie to you that <declarative sentence s'>."

Theorem 3 *Lies are not performative*
$\models\forall$ spkr, addr, e, n
(MK spkr addr [(DONE spkr t_1?;e) $\wedge$ (UTTER spkr addr e s)]) $\supset$
     $\sim$(DONE {LIE spkr addr e (TRUE s e) t_1})

In other words, you cannot perform a lie by saying "I lie that ..."

Proof Sketch: Assume that it is mutually believed that the speaker has uttered declarative sentence s. Now, apply Domain Axiom 1A. By assumption, the first conjunct of the antecedent holds. There are then two cases to consider. First, assume (**) the second conjunct holds (say, at level n = 1), i.e., the addressee does not believe the speaker did not want him to know that he believed s' was true. In virtue of the supposed truth conditions on lying, spkr would have to have been lying. By expanding its definition, and using the memory and introspectability properties of BEL, GOAL, and INTEND the addressee can conclude that, before the utterance, the speaker wanted him not to know that the speaker believes that in uttering S, he was lying. But, this contradicts the assumption (**). Since the speaker in fact uttered the sentence, that assumption is false, and the addressee believes the speaker did not in fact want him to know

that he believed the sentence was true. This renders impossible the intentions to be achieved in asserting, which are constitutive of lying as well.

Now, assume (**) is false, so the addressee in fact believes the speaker did not want him to know that s' was true. Again, this immediately makes the speaker's intentions in asserting, and hence lying, impossible to achieve. So, in neither case is the utterance a lie. If the addressee believes the speaker is a competent speaker of the language, the speaker must have intended some other interpretation.

8 Conclusion

Requesting works well as a performative verb because requesting requires only that the agent has made an attempt, and need not have succeeded in getting the hearer to do the requested action, or even to form the right beliefs. Some verbs cannot be used performatively, such as "frighten," because they require something beyond a mere attempt. Hence, such verbs would name action expressions that required a particular proposition p be true after the utterance event. When the utterance event does not guarantee such a p, the use of the performative verb will not be possible.

On the other hand, certain utterances (performative or not), when performed by the right people in the right circumstances, make certain institutional facts hold. So, when a clergyman, judge, or ship captain says "I now pronounce you husband and wife," the man and woman in question are married. In our framework, there would be a domain axiom whose antecedent characterizes the circumstances, participants, and nature of the utterance event, and whose consequent asserts that an institutional fact is true. The axiom is justified not by the nature of rational action, but by the existence of an institution. Such utterances could be made with a performative prefix provided such attempts are made into successes by the institution.

This paper has shown that treating performative utterances as declarative sentences is a viable analysis, in spite of Searle's criticisms. The performative use of an illocutionary verb is self-guaranteeing when the named illocutionary act consists in the speaker's making an attempt to make public his mental state. In such cases, if the speaker thinks he has done so, then he has. However, we do not derive the named illocutionary act from the assertion, nor vice-versa. Instead, both derivations may be made from the utterance event, but the assertive one is in fact harder to obtain as it has extra conditions that need to be satisfied.

References

[1] K. Bach and R. Harnish. *Linguistic Communication and Speech Acts.* M. I. T. Press, Cambridge, Massachusetts, 1979.

[2] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(3), 1990.

[3] P. R. Cohen and H. J. Levesque. Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. M.I.T. Press, Cambridge, Massachusetts, in press.

[4] D. Harel. *First-Order Dynamic Logic*. Springer-Verlag, New York City, New York, 1979.

[5] K. Kogure, H. Iida, K. Yoshimoto, H. Maeda, M. Kume, and S. Kato. A method of analyzing Japanese speech act types. In *Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, 1986.

[6] D. J. Litman and J. F. Allen. A plan recognition model for subdialogues in conversation. Technical report, Department of Computer Science, Univ. of Rochester, Rochester, New York, November 1984.

[7] J. McCarthy. ELEPHANT: a programming language based on speech acts. Unpublished ms., Dept. of Computer Science, Stanford University, 1989.

[8] C. R. Perrault. An application of default logic to speech act theory. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. M.I.T. Press, Cambridge, Massachusetts, in press.

[9] J. Sadock. *Toward a Linguistic Theory of Speech Acts*. Academic Press, New York, 1984.

[10] J. Searle. How performatives work. *Linguistics and Philosophy*, 12:535–558, 1989.

[11] J. R. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, Cambridge, 1969.

[12] J. R. Searle and D. Vanderveken. *Foundations of Illocutionary Logic*. Cambridge Univ. Press, New York City, New York, 1985.

[13] Y. Shoham. Agent oriented programming. Unpublished ms., Dept. of Computer Science, Stanford University, October 1989.

[14] H. Tennant. Evaluation of natural language processors. Technical Report T-103, Coordinated Science Laboratory, University of Illinois, Urbana, Illinois, November 1980. Ph. D. Thesis.

[15] T. Winograd and F. Flores. *Understanding Computers and Cognition: A New Foundation for Design*. Ablex Publishing Co., Norwood, New Jersey, 1986.

A   Model Theory

A model M is a structure $\langle \Theta, P, E, Agt, T, B, G, \Phi \rangle$, where Θ is a set, P is a set of people, E is a set of primitive event types, $Agt \in [E \longrightarrow P]$ specifies the agent of an event, $T \subseteq [Z \longrightarrow E]$ is a set of possible courses of events (or worlds) specified as a function from the integers to

elements of E, $B \subseteq T \times P \times Z \times T$ is the belief accessibility relation, $G \subseteq T \times P \times Z \times T$ is the goal accessibility relation, and Φ interprets predicates. Formulas will be evaluated with respect to some possible course of events, hereafter some *possible world*, and an "index" into that possible world, that is, at a particular point in the course of events.

A.1 Definitions

- $D = \Theta \bigcup P \bigcup E^*$, specifying the domain of quantification. That is, one can quantify over things, people, and sequences of (types of) primitive events. Given this, $\Phi \subseteq [Pred^k \times T \times Z \times D^k]$.

- $AGT \subseteq E^* \times P$, where $x \in AGT[e_1, \ldots, e_n]$ iff there is an i such that $x = Agt(e_i)$. That is, AGT specifies the partial agents of a sequence of events.

A.2 Satisfaction

Assume M is a model, σ a sequence of events, n an integer, v a set of bindings of variables to objects in D, and if $v \in [Vars \longrightarrow D]$, then v_d^{x} is that function which yields d for x and is the same as v everywhere else. We now specify what it means for M, σ, v, n to *satisfy* a wff α, which we write as $M, \sigma, v, n \models \alpha$. Because of formulas involving actions, this definition depends on what it means for an expression a to *occur* between index points n and m. This, we write as $M, \sigma, v, n[\![\mathsf{a}]\!]m$, and is itself defined in terms of satisfaction. The definitions are as follows:[10]

1. $M, \sigma, v, n \models (\mathsf{BEL\ x\ p})$ iff for all σ^* such that $B(\sigma, v(\mathsf{x}), n, \sigma^*)$, $M, \sigma^*, v, n \models \mathsf{p}$.

2. $M, \sigma, v, n \models (\mathsf{BMB\ x\ y\ p})$ iff for all $\sigma_1, \ldots \sigma_m$ such that $\sigma_1 = \sigma$ and $B(\sigma_{2i-1}, v(\mathsf{x}), n, \sigma_{2i})$ and $B(\sigma_{2i}, v(\mathsf{y}), n, \sigma_{2i+1})$ for every $1 \leq i \leq m/2$, we have that $M, \sigma_m, v, n \models \mathsf{p}$. This is one-sided mutual belief (that is, x thinks it is mutually believed).

3. $M, \sigma, v, n \models (\mathsf{GOAL\ x\ p})$ iff for all σ^* such that $G(\sigma, v(\mathsf{x}), n, \sigma^*)$, $M, \sigma^*, v, n \models \mathsf{p}$.

4. $M, \sigma, v, n \models (\mathsf{AGT\ x_1\ \ldots\ x_k\ e})$ iff $v(\mathsf{e}) = S_1 S_2 \ldots S_m$ and $\{x \mid x = Agt(S_j) \& \mathsf{e} \in S_j\} = \{v(\mathsf{x}_j)\}$. Thus the x_j are the *only* agents of e. One can easily define inductively $(\mathsf{AGT\ x_1 \ldots x_k\ a})$, where a is an arbitrary action expression.

5. $M, \sigma, v, n \models (\mathsf{e_1 \leq e_2})$ iff $v(\mathsf{e_1})$ starts $v(\mathsf{e_2})$.

6. $M, \sigma, v, n \models (\mathsf{HAPPENS\ a})$ iff $\exists m, m \geq n$, such that $M, \sigma, v, n[\![\mathsf{a}]\!]m$. That is, a describes a sequence of events that happens "next" (after n).

7. $M, \sigma, v, n \models (\mathsf{DONE\ a})$ iff $\exists m, m \leq n$, such that $M, \sigma, v, m[\![\mathsf{a}]\!]n$. That is, a describes a sequence of events that *just* happened (before n).

---

[10]For conciseness, we omit that part of the definition that deals with the constructs of first-order logic.

8. $M, \sigma, v, n \models (\mathsf{DOING}\ \mathsf{a})$ iff $\exists m \leq n$ and an $n' \geq n$ such that $m[\![\mathsf{a}]\!]n'$. That is, the doing of a overlaps the current time. Notice that $(\mathsf{HAPPENS}\ \mathsf{a})$ and $(\mathsf{DONE}\ \mathsf{a})$ both imply $(\mathsf{DOING}\ \mathsf{a})$.

9. $M, \sigma, v, n \models \langle\text{Time-Proposition}\rangle$ iff $\langle\text{Time-Proposition}\rangle = n$.

Turning now to the occurrence of actions, we have:

1. $M, \sigma, v, n[\![\mathsf{e}]\!]n + m$ (where e is an event variable) iff $v(\mathsf{e}) = S_1 S_2 \ldots S_m$ and $S_i \subseteq \sigma(n + i), 1 \leq i \leq m$. Intuitively, e denotes some sequence of events of length m which appears next after n in the world σ.

2. $M, \sigma, v, n[\![\mathsf{a;b}]\!]m$ iff $\exists k, n \leq k \leq m$, such that $M, \sigma, v, n[\![\mathsf{a}]\!]k$ and $M, \sigma, v, k[\![\mathsf{b}]\!]m$. The action described by a and then that described by b occurs.

3. $M, \sigma, v, n[\![\mathsf{a} \mid \mathsf{b}]\!]m$ iff $M, \sigma, v, n[\![\mathsf{a}]\!]m$ or $M, \sigma, v, n[\![\mathsf{b}]\!]m$. Either the action described by a or that described by b occurs within the interval.

4. $M, \sigma, v, n[\![\mathsf{a} \parallel \mathsf{b}]\!]m$ iff $M, \sigma, v, n[\![\mathsf{a}]\!]m$ and $M, \sigma, v, n[\![\mathsf{b}]\!]m$. That is, both the action described by a and that described by b occur during the interval.

5. $M, \sigma, v, n[\![\mathsf{p?}]\!]n$ iff $M, \sigma, v, n \models \mathsf{p}$. The test action, $\mathsf{p?}$, involves no events at all, but occurs if p holds, or "blocks" (fails), when p is false.

6. $M, \sigma, v, n[\![\mathsf{a^*}]\!]m$ iff $\exists n_1, \ldots, n_k$ where $n_1 = n$ and $n_k = m$ and for every i such that $1 \leq i \leq k - 1$, $M, \sigma, v, n_i[\![\mathsf{a}]\!]n_{i+1}$. The iterative action $\mathsf{a^*}$ occurs between n and m provided only a sequence of what is described by a occurs within the interval.

A wff p is *satisfiable* if there is at least one M, world σ, index n, and assignment v such that $M, \sigma, v, n \models \mathsf{p}$. A wff p is *valid*, iff for every M, world σ, event index n, and assignment of variables v, $M, \sigma, v, n \models \mathsf{p}$.

B   Proofs of Theorems

B.1   Performative Requests

Let s be "I request you to $\mathsf{s_1}$."

Theorem 1 *A Performative Request*

$\models \forall$ spkr, addr, e, n, $\mathsf{t_1}$,

(MK spkr addr (DONE spkr $\mathsf{t_1}$?;e) $\wedge$

       (UTTER spkr addr e s)) $\wedge$

(BEFORE $\mathsf{t_1}$?;e

(GOAL spkr [AFTER t_1?;e

        (KNOW addr

           [BEFORE t_1?;e (BEL spkr (TRUE s e))]])])) $\supset$

(DONE {REQUEST spkr addr e α t_1})

Assume the antecedent is true. We first apply the Domain Axiom 1A to the utterance event, and note that if spkr has just done an utterance event, he believes he has done it (see [2]), meaning the speaker is also an observer of the event. Therefore, we have (at level n=1 of the ABEL)

(1) (DONE spkr t_1?;e) $\wedge$ (BEL spkr (BEFORE e (BEL spkr (TRUE s e)))) is true.

By Domain Axiom 1A, this implies

(2) (DONE spkr spkr t_1?;e) $\wedge$

      (BEL spkr [BEFORE spkr e (BEL spkr (DONE t_1?;{REQUEST spkr addr e α}))]),

That is, an event e is being performed by spkr, and spkr thinks spkr believed he has just done a direct request to fulfill the embedded imperative sentence. By expanding out REQUEST into an ATTEMPT, we see that (2) entails

(3) (BEL spkr (BEFORE spkr e

        [BEL spkr (DONE spkr

            t_1?;{ATTEMPT spkr addr e (BMB addr spkr (GOAL spkr β)) α t_1}],

where

      $\beta \stackrel{\text{def}}{=} \Diamond\alpha \wedge$ [AFTER spkr e (INTEND$_2$ addr α [(GOAL spkr $\Diamond\alpha$) $\wedge$

                              (HELPFUL addr spkr)])], and

      $\alpha \stackrel{\text{def}}{=} \exists e_2$(DONE addr e_2) $\wedge$ (FULFILL-CONDS s e_2 e)

Next, by the memory and introspectibility properties of BEL and by expanding the definition of ATTEMPT in (3), we get (4)

(4) (BEFORE e [BEL spkr

      (DONE t_1?;[(BEL spkr $\sim \beta \wedge \sim \alpha$) $\wedge$

      [INTEND$_1$ spkr t_1?;e_2;(BMB addr spkr (GOAL spkr β))?

               (GOAL spkr $\Diamond\alpha$)]

17

$$\text{(GOAL spkr } \Diamond\alpha) \text{]; e)])}$$

Now, in virtue of the introspection properties of GOAL, and noting that an agent who thinks at time t_1 that he intends at time t_1 to do an action at time t_1 INTEND$_1$ in fact has that intention, spkr knew what his intentions were as well. Therefore, we have:

(5) (DONE t_1?;[(BEL spkr $\sim \beta \land \sim \alpha$) $\land$

$\qquad\qquad$ [INTEND$_1$ spkr t_1?;e_2;(BMB addr spkr (GOAL spkr β))?

$\qquad\qquad\qquad\qquad$ (GOAL spkr $\Diamond\alpha$)]

$\qquad\qquad$ (GOAL spkr $\Diamond\alpha$)]; e)])

So, we have

(5) (DONE spkr t_1?;{REQUEST spkr addr e α t_1};)

Q.E.D.

All these conclusions are, of course, true right after the performative sentence has been uttered.

B.2 Lying is not performative

Let s be "I lie to you that s'".

Theorem 3 *Lies are not performative*

$\models\forall$ spkr, addr, e, n

(MK spkr addr [(DONE spkr t_1?;e) $\land$ (UTTER spkr addr e s)]) $\supset$

$\quad\sim$(DONE {LIE spkr addr e (TRUE s e) t_1})

Proof:[11]

Let $\Psi \overset{\text{def}}{=}$ (BEL addr (BEFORE spkr e (BEL spkr (TRUE s e)))) $\land$

$\qquad\quad$ [BEFORE e (BEL spkr $\sim$ (TRUE s e))].[12]

To begin, assume the antecedent to Domain Axiom 1A holds at level n = 1, meaning that the addressee does not suspect the speaker of insincerity. That is, assume

(*) $\sim$(BEL n addr spkr [BEFORE spkr e

$\qquad\qquad\qquad$ $\sim$(GOAL spkr [AFTER spkr e

---

[11]The part of the formula to be operated upon next will appear in italics.

[12]Notice that Ψ entails $\sim$(KNOW addr [BEFORE spkr e (BEL spkr (TRUE s e))])

$$\text{(KNOW addr [BEFORE spkr e}$$
$$\text{(BEL spkr (TRUE s e))]}\Big]$$

So, at level n = 1, by Domain Axiom 1A

(1) (BEL addr [BEFORE spkr e

(GOAL spkr [AFTER spkr e

(KNOW addr

(BEFORE spkr e [BEL spkr *(True s e)*]))$\Big]$ holds.

That is, the addressee thinks the speaker wanted him to come to know that the speaker believes the "I lie ..." sentence is true. In virtue of the supposed truth conditions of performative lying, he wanted the addressee to know that he lied.

(2) (BEL addr [BEFORE spkr e

(GOAL spkr

[AFTER spkr e

(KNOW addr

(BEFORE spkr e

[BEL spkr

(DONE spkr t_1 *?;{LIE spkr addr e (TRUE s' e)}*$\Big]$,

So, by the definition of LIE, the expansion of the INTEND$_1$ argument to the ATTEMPT in its definition, and the first expansion of BMB, and we conclude

(3) (BEL addr

[BEFORE spkr e

(GOAL spkr

[AFTER spkr e

(KNOW addr

[BEFORE spkr e

(BEL spkr

(BEL spkr $\sim$*(TRUE s' e))* $\wedge$

[GOAL spkr (AFTER spkr e Ψ) $\Big]$

19

By Assumptions 2 and 3, if spkr thinks he had those BEL and GOALs, then he did. This entails

(4) (BEL addr

        (BEFORE spkr e

             (GOAL spkr

                 [AFTER spkr e

                    ($KNOW$ addr

                        [BEFORE spkr e

                          (BEL spkr $\sim$(TRUE s' e)) $\wedge$

                          (GOAL spkr (AFTER spkr e Ψ)) $\Big]$

Therefore, by the definition of KNOW, we have

(5) (BEL addr

        [BEFORE spkr e

             (GOAL spkr

                 $[AFTER\ spkr\ e\ [BEFORE\ spkr\ e$

                        (BEL spkr $\sim$(TRUE s' e)) $\wedge$

                        (GOAL spkr (AFTER spkr e Ψ)) $\Big]$

Now, (AFTER spkr e (BEFORE e α)) entails α, so we have,

(6) (BEL addr [BEFORE spkr e $(GOAL\ spkr\ (GOAL\ spkr$ (AFTER spkr e Ψ) $\Big]$

and (GOAL spkr (GOAL spkr α)) entails (GOAL spkr α), so we have

(7) (BEL addr

        [BEFORE spkr e

$$(\text{BEL spkr} \sim(\text{TRUE s' e})) \wedge$$
$$(\text{GOAL spkr} \ (\text{AFTER spkr e} \ \Psi))])$$

This entails:

(8) (BEL addr

        [BEFORE spkr e

          (GOAL spkr (AFTER spkr e

                $\sim$(KNOW addr [BEFORE spkr e (BEL spkr (TRUE s' e))])]

This contradicts the assumption in Domain Axiom 1A. Thus, we can conclude that the assumed antecedent (*) is false, i.e.,

(BEL addr spkr

        [BEFORE spkr e

          $\sim$(GOAL spkr [AFTER spkr e

               (KNOW addr (BEL spkr (TRUE s e))]].

But this contradicts the intention constitutive of asserting, and hence lying.

In the same fashion, if the second conjunct of the antecedent of Domain Axiom 1A is false, we immediately arrive at the same contradiction. Hence, the utterance is not a lie. Q.E.D.