Reference and Computational Models of Illocutionary Acts

Amichai Kronfeld
Microsoft
amik@microsoft.com

May 24, 1995

1 Introduction

My general goal in this paper is to explore possible relations between a cognitive analysis of the notion of reference on the one hand, and computational models of speech acts (illocutionary acts, to be precise) on the other. On the cognitive side, we have the following problem: how can thoughts (and sentences that articulate them) be about objects? The deceptive simplicity of the question masks a century of lively philosophical debate but it is not the debate that I wish to focus on here. Instead, I will assume a framework of what I have called the descriptive approach to the problem of reference. The most important single principle of this framework is that the descriptive content of an individuating representation denoting an object is both necessary and sufficient for a belief (or any other propositional attitude) to be about that object. I have argued for this principle elsewhere and it will simply be assumed here.

On the computational side, we have various attempts to show how illocutionary acts (that is, "complete" and independent speech acts such as asserting, promising, requesting, congratulating, etc.) can be viewed as special kinds of plans which in turn can be formalized and coded using various logical systems and other tools of the AI trade.

The bridge I propose in this paper between the cognitive and the computational, in this paper, will be constructed through the speech act of referring. If the problem of reference is: "how can thoughts be about objects," the problem of referring can be encapsulated as "how does a speaker let a hearer know what object is being talked about." Surely, a solution to the problem of referring presupposes a solution to the problem of reference. On the other hand, as we shall see, the computational approach to speech acts has a lot to contribute to the solution of the problem of referring as well.

The focus of this paper, then, will be a computational model of the speech act of referring. After discussing several fundamental concepts that are needed to get us off the ground, I will suggest what amounts to a preliminary analysis of the speech act of referring, based on the descriptive approach to the problem of reference. Then I will present one important attempt (by Cohen and Levesque) to provide axioms for illocutionary acts, thereby showing how intended effects of illocutionary acts can be derived (and therefore

\footnote{Kronfeld 1990, pp. 17-44.}

\footnote{op. cit. pp. 116-140.}
planned). This attempt will be further motivated by the notion of a literal goal which I will introduce shortly.

Cohen and Levesque's model will be followed by an outline of a computational model of referring which is based on their approach on the one hand, and the cognitive analysis of reference on the other. I will then point out a major weakness in this approach and offer a new direction, based on recent work on the formal properties of illocutionary acts. This work (Venderveken 1990; Venderveken 1991) is not free of problems itself, but as we shall see, the cognitive analysis of reference may itself provide the required solutions. So on the one hand, formal work on the logic and semantics of illocutionary acts may provide tools for advancing the computational model of referring. On the other hand, the cognitive analysis of reference may solve problems in the theory of illocutionary acts itself.

2 Preliminary Concepts

Throughout this paper, I will make use of four fundamental concepts. They are the discourse purpose of a speech act and its literal goal, the notion of identification (of an object as the referent in a speech act) and the notion of an individuating set. In this section I explain what I mean by these concepts.

2.1 Discourse Purpose and Literal Goals

When performing speech acts, people may have many different goals and intentions. For example, in making an assertion my goal may be to inform you, as well as to impress or insult you, try to alleviate boredom or simply wish to exercise my vocal chords. However, not all intentions that may motivate a speech act are relevant to this discussion. What I am interested in are communication intentions and goals.

Roughly, communication intentions and goals are those intentions that are intended to be recognized or, more precisely, intended to be satisfied (at least in part) through their recognition. For example, the intention to deceive is not a communication goal because it is not (and could not be) intended to be recognized. The intention to congratulate, on the other hand, is a communication intention par excellence: it will never be achieved unless recognized.

Sometimes the recognition of a goal suffices for its satisfaction. For example, if my objective is to congratulate you, I can succeed if (and in this particular case, only if) you recognize my intention to do so. Once you have recognized my intention, you are thereby congratulated. I call such goals, wherein recognition is sufficient for success, the literal goal of the speech act. In addition, we have discourse purposes, which are the goals underlying both the choice to engage in discourse in the first place (rather than in a nonlinguistic activity) and the choice of a particular propositional content to be expressed (Gross and

---

3 This view of communication intentions originates with Grice's analysis of the concept of meaning (Grice 1957). But much research in computational linguistics, though obviously influenced by Grice, has nevertheless stressed the role of intention and goal recognition in discourse, quite independently of a theory of meaning. Allen's dissertation (Allen 1978) and his subsequent work with Perrault (Allen and Perrault 1978; Allen and Perrault 1980), for example, emphasize the importance of goal recognition for inferring the speaker's plans. So do Gross and Sidner (Gross and Sidner 1986), as well as Sidner herself in her own work (Sidner 1983; Sidner 1985). These authors show how the recognition of what the speaker is "up to" contributes to coherence and comprehensibility of the discourse (see also Perrault et al. 1978), and is essential for the hearer's generation of an appropriate response.

4 The term "literal goal" is taken from Kasher (1977), where literal purposes are introduced. Our use of the two terms is virtually the same, except that Kasher wishes to explain what literal purposes are in a way that is independent of Gricean intentions.
Sidner 1986). In the case of congratulating, the literal goal and the discourse purpose are one and the same — but this is the exception rather than the rule. For example, if I ask you to close the door, the literal goal of the request is essentially to make you realize what you are asked to do. The discourse purpose is to make you actually do it. Note that, unlike the case of literal goals, recognition of the discourse purpose, albeit vital for the success of the speech act, is neither necessary nor sufficient for the objective to be achieved. You may recognize my wish that the door be closed without being inclined in the least to cooperate and you may close the door just because you feel like it even though you had no idea what I was talking about. Rather, given certain assumptions about the disposition of discourse participants, it is rational to expect that the hearer’s recognition of a discourse purpose will enhance the speaker’s chances of achieving his or her goal.

The discourse purpose of the speech act of referring is straightforward: it is to make the hearer identify the referent. The literal goal of referring, however, is trickier. What we are after is an intention that is unique to the speech act of referring and whose recognition is enough to satisfy it! What could such an intention be?

Once we adopt the descriptive approach to reference, however, the literal goal of referring is not hard to find. Remember that the main principle of the descriptive approach is that the descriptive content of an individuating representation denoting an object is both necessary and sufficient for a belief to be about that object. So an intuitive account of referring along these lines is simply this: A speaker has a mental representation denoting an object; by using a noun phrase that is intended to be interpreted as a linguistic representation of the object, the speaker intends to invoke in the hearer a mental representation denoting that very same object. But it should be noted that such an intention has precisely the quality we are looking for. Once the hearer recognizes the intention that he have a mental representation denoting the same object that the speaker has in mind, the hearer does have such a representation. For example, if I recognize your intention to convey to me a representation of whatever object you are talking about, then I do have a representation of that object, namely: the object the speaker is talking about. In other words, the mere recognition of the intention to represent an object suffices to produce a representation of that object. A central referring intention, therefore, is the intention to invoke a representation of a particular object in the hearer by means of his recognition of this intention. The goal of satisfying this intention is the literal goal of referring.

2.2 Identification as a Discourse Purpose

As I mentioned above, the discourse purpose of the speech act of referring is identification of the referent. But the concept of identification itself is so vague and ambiguous that it is of little help to a computational approach. Let us make, therefore, two important distinctions.

First, the speaker’s sense of “identify” is quite different from that of the hearer. A speaker is said to identify an object for or to a hearer, while the hearer is said to identify that object if the referring act is successful. Standard speech act theory focuses on the speaker’s sense since the main aim there is to characterize institutional rules of behavior (Searle 1969). But since our objective is to characterize a computational model within a plan-based approach, the actual effect on the hearer’s mental state is really what counts. If no intended effects are formulated, there is little point in planning. Thus, for the purpose of this paper, identification as the discourse purpose of referring is understood in the hearer’s sense, taking the speaker’s sense simply to be his plan to achieve the hearer’s identification of the referent.
Second, identification, as the discourse purpose of the referring act, should be carefully distinguished from identification in the sense of knowing who someone is. The former may be considered a pragmatic notion of identification. The latter is an epistemic one. Although the pragmatic notion of identification is also connected with knowledge (in the sense of knowing who the speaker is talking about), the two are quite distinct. To illustrate the difference, consider a situation in which two detectives are discussing a murder case that they are trying to solve. Neither speaker nor hearer in this situation know who the murderer is and they cannot “identify” him if identification is interpreted epistemically. From a pragmatic point of view, however, there is a clear distinction: if the hearer understands the usage of the phrase “the murderer” in this context, he has identified what the speaker is talking about. Otherwise he has not. Obviously, I take the pragmatic sense of identification to be the discourse purpose of referring.

2.3 Individuating Sets

As mentioned above, mental representations and their descriptive content are crucial for reference and referring. I emphasize mental because I take such representations to be constituents in the contents of propositional attitudes and not necessarily in the contents of utterances. Indeed, I do not believe that the meaning of all referring expressions is to be spelled out in terms of descriptive content. Some referring expressions (definite descriptions) obviously have descriptive content, but such content may or may not be relevant for either the belief that is being expressed or for the referring act to succeed. Other referring expressions (proper names, demonstratives and indexicals) have no descriptive content associated with them and the success or failure of referring acts in which they are used is to a large extent independent of their meaning, whatever such meaning is taken to be.

So when I talk of about mental, individuating representations and their descriptive content, I mean the kind of stuff that plays a role in forming the contents of our propositional attitudes, quite independently of any speech act that we may perform. Such representations, I claim, are grouped into what I call individuating sets. An individuating set is an exhaustive list of mental representations, all believed by the agent to denote the same object. I regard individuating sets, rather than isolated mental presentations, as essential to a referring model. The reasons for this claim are rather complex and I will only mention briefly two of them:

First, the notion of pragmatic identification can be interpreted as the process of applying various constraints on individuating sets. Here is a list of such constraints which illustrate different modes of pragmatic identification:

- A requirement that the relevant individuating set should contain a perceptual representation, to be acquired now or later.
- A requirement that the hearer be able to merge a newly generated individuating set with a pre-existing one.
- A requirement that the relevant individuating set contain one or more representation that are privileged with respect to the goals of the speaker.\(^7\)

\(^6\)See Donnellan's referential use of definite descriptions (Donnellan [1966]1971) and in general cases in which speaker's meaning is different from sentence meaning.

\(^7\)For example, suppose A asks B: "Do you think Ronald Reagan already had Alzheimer's disease when he testified about the Iran/Contra affair?" B may have a very rich individuating set for Reagan. He may know,
and so on.

Thus, the notion of an individuating set makes a unified account of identification as the discourse purpose of referring possible, and there is no need to confound pragmatics and epistemology while insisting that "real" referring takes place only when it is known who or what the referent is.

Second, the notion of an individuating set has enabled us to represent clearly a basic distinction between two types of beliefs about objects. Roughly, an agent can think of an object either qua having a certain property, or in itself. For example, my belief that Kripke was right in claiming that Donnellan did not refute Russell's theory of descriptions is about Kripke qua being the author of "Speaker's Reference and Semantic Reference". If it turns out that someone else wrote the article my belief will no longer be about Kripke. On the other hand, my belief that the author of this article lives in Princeton is also about Kripke, but not qua being the author or qua anything else. If it turns out that Kripke is not the author I would still believe that he is the one who lives in Princeton, whether he wrote the article or not. This dichotomy manifests itself both in thought and in speech and the concept of an individuating set provides a very convenient way of representing it entirely within a descriptive view: either a particular presentation mode is selected out of the relevant individuating set and is taken as crucial for the identity of the referent, or no such presentation mode is chosen and the referent is determined by the individuating set as a whole. Such a mechanism solves a host of related problems. First, it reconciles the view that mental representations denoting the referent are necessary for thinking about objects with rigid designators (such as proper names and demonstratives) which lack any descriptive content conventionally associated with them. Second, by making the individuating set the fundamental unit that determines what a belief or a statement is about, we see how referring can succeed even if the actual definite description used is not rich enough in content to denote the particular object or even if it is entirely wrong. Indeed, the ability to handle such cases is one of the major advantages of the computational model that I am advocating, as we shall see later on.

3 What is the Speech Act of Referring?

Armed with the notions of literal goal, discourse purpose, pragmatic identification and individuating sets, we can begin by spelling out what, roughly, the referring act is taken to be in this framework. As I mentioned earlier, my starting point is that the speech act of referring is a transaction between a speaker and a hearer in which the speaker begins with one mental representation and the hearer ends up with another. In other words, the speaker has a mental representation denoting what he believes to be a particular object, and he intends the hearer to come to have a mental representation denoting the same object, at least in part through the use of a noun phrase that is intended to be a linguistic representation of that object. More precisely (and including the literal goal and discourse purpose) we get the following:

Referring: When a speaker performs the speech act of referring, he has in mind an individuating set which he believes determines a particular object, and he uses a noun phrase with the following intentions:

say, each and every detail of Regan's Hollywood career. But if this individuating set does not contain a representation such as President of the United States between 1980 and 1988, pragmatic identification has not been accomplished.

*(Kripke 1977).*
1. **Literal goal:** that as a result of the hearer’s recognition of the noun phrase as a referring expression, the hearer will generate an individuating set that will determine this very same object.

2. **Discourse purpose:** that the hearer will apply various operations to the newly created individuating set so that it will meet the appropriate identification constraints. 

This, in a nutshell, is what I take to be the cognitive foundation for a computational model of referring. Let us now turn to the computational analysis.

**4 Axioms for Illocutionary Acts**

One approach to modeling speech acts as a planned activity is based on two related ideas:

1. Mental representations can be modeled by logical formulas, and

2. The process of planning to achieve a goal can be seen as a search for a derivation proceeding from axioms representing the world as it is, to theorems that represent a world in which the goal is satisfied.

The method, roughly, is to look for a formal system that is rich enough to capture central concepts of communication. Once such a formal system is available, there are two things that need to be done. First, the goal structure that typically motivates a speech act, and the likely effects of such an act on the hearer’s mental state, must be formalized. Second, the relation between goals and effects must be spelled out. That is, it must be shown why given these effects, it is rational for the speaker to expect that his goals will be satisfied.

The connection between the achievement of communication goals and the assumption of rationality is not a new idea, of course. Grice’s principle of conversational cooperation provides a theoretical framework in which discourse is seen as a means toward an end, and the project of grounding conversational maxims in general rationality constraints is developed further by Kasher. Within computational linguistics, Cohen and Levesque have developed a detailed formal theory in which illocutionary acts are characterized in terms of what the speaker and the hearer mutually believe about each other’s mental states as the result of a speech act. Moreover, Cohen and Levesque are able to show why utterances are expected to be successful given reasonable assumptions about the rationality of discourse participants. Their approach provides an excellent starting point within which principles of referring can be formalized.

The foundation of Cohen and Levesque’s theory of rational communication is a formal system which models (idealized versions of) beliefs, goals and actions. All that is needed for our purposes is the notation for five of the concepts in their logical apparatus:

---

8There is a need to distinguish between *local* individuating sets and *quasi-permanent* ones. Local individuating sets are generated, modified, dropped etc. during discourse. Quasi-permanent ones are the mental structures that allow us to think of objects independently of discourse. The two kinds are related to each other along the lines of short-term/long-term memory.

Note also that identification of the referent *qua* having a particular property is a special case of the satisfaction of identification constraints.

10(Grice 1975).


BELa(P): Agent A believes that P.

GOALa(P): Agent A has the goal of making P true.

BMBb(P): Agent A believes that A and B mutually believe that P.

This is the unilateral mutual belief from A's perspective — that is, the infinite conjunction of:

- BELa(P),
- BELaBELb(P),
- BELaBELbBELa(P),
- and so on.

DONEa(E): E is an action-event that has just happened, and whose only agent is A.

AFTER(a, P): P is true in all courses of events that follow an occurrence of act a.13

We will also need a partial characterization of rational action and cooperation in terms of sincerity, helpfulness, and competence. Although Cohen and Levesque provide formal definitions for these concepts, it would be convenient to treat them here as derivation rules, as follows:

Sincerity: From GOALaBELb(P) infer BELa(P), provided that A is sincere (with respect to P). That is, if you assume that A is sincere, then from A's attempt to make B believe that P you may conclude that A believes P himself.

Helpfulness: From BELbGOALaGOALb(P) infer GOALb(P), provided that B is helpful (with respect to P). That is, if you assume that B is helpful, then from B's believing that A is trying to make him do something, you may conclude that B will indeed try to do it.

Competence: From EELa(P) infer P and from GOALa(Q) infer eventually Q, provided that A is competent (with respect to P and Q). That is, if you assume that A is competent, then from his belief that P you may conclude that P is indeed true, and from the fact that he has the goal of making Q true you may conclude that Q will eventually take place14.

13Cohen and Levesque first introduce AFTER as a primitive, branching-time concept (Cohen and Levesque 1985). In later work they use HAPPENS(<a>) (action a happens next) as a linear-time primitive (Cohen and Levesque 1987). The semantics of Cohen and Levesque's logic is based on possible-world interpretation. Each possible world consists of a sequence of primitive events that indicates what has happened and what will happen in that world. As usual, accessibility relations among possible worlds characterize the operators representing propositional attitudes. For a complete specification of the syntax and semantics of this logic, as well as its development, see the references cited above.

14The logic that includes BEL, GOAL, BMB, DONE, and AFTER constitutes a basic, "atomic" layer which Cohen and Levesque then use to construct more complex concepts. In particular, they are able to define the notion of a persistent goal (a goal that an agent would not give up unless certain conditions are met), which is then used both in solving difficult problems in automatic planning (the "Little Nell" problem, McDermott 1982) and in formalising important aspects of intending. A formal characterization of intending, in turn, plays an important role in Cohen and Levesque's account of illocutionary acts.

I think that the concept of a persistent goal is an important one, and that Cohen and Levesque's approach to the formalisation of a theory of action provides an excellent tool for modelling speech acts, including the speech act of referring. However, delving into details would take us too far afield. Therefore, in illustrating Cohen and Levesque's approach (and in modeling referring along the same lines) I will use the primitive concept of GOAL rather than their P-GOAL (persistent goal) or INTEND. This is, of course, a simplification but it enables us to see the general outline more clearly.
Let us turn, then, to Cohen and Levesque's account of rational communication. Following Grice, they view communication first and foremost as an expression of attitudes that are intended to affect desired changes in the hearer's mental state. According to their initial approach, modeling this process consists of two parts. First, the model specifies the context-independent effects on the hearer's attitudes when a sentence with certain features (e.g., a certain grammatical mood) is uttered under normal conditions. Second, the model shows how the discourse purpose of the speaker is derived from these core effects by means of an elaborate theory of rational action and interaction (including cooperation).

Let us focus our discussion on declarative and imperative sentences in English. Typical illocutionary acts that can be performed by means of such sentences are inform and request, respectively. Their intended effects are that the hearer adopt a belief (in the case of inform), or do something (in the case of request). These are the discourse purposes of these illocutionary acts, and depending on circumstances, they are either achieved or not. What we are after, however, are the context-independent effects of declaratives and imperatives. To uncover these effects, it is best to treat them as those changes in the hearer's mental state that occur when the literal goal of the utterance is satisfied. This yields a criterion for specifying the context-independent effects of a speech act: they should be those effects on the hearer's mental state that take place as a result of the hearer's recognition of the attempt to achieve them.

Take the literal goal of inform. Suppose the speaker tells the hearer that $P$. The discourse purpose of the speech act is to make the hearer believe that $P$, that is

\begin{equation}
\text{GOAL}_{h} \text{BEL}_{h}(P).
\end{equation}

But this obviously cannot be the literal goal. The hearer may recognize that he is intended to believe $P$, but he may question the speaker's competence in making his judgment. Thus, a weaker goal suggests itself, namely, to make the hearer believe that the speaker believes that $P$:

\begin{equation}
\text{GOAL}_{h} \text{BEL}_{h} \text{BEL}_{a}(P)
\end{equation}

This still cannot be the literal goal. The hearer may recognize that the speaker wants him to believe that he (the speaker) believes that $P$, but the hearer may doubt the speaker's sincerity.

However, if the hearer understood the speaker at all, he cannot fail to see that the speaker has the goal that (4.2) expresses (let's ignore exceptions, such as irony for now). Hence, once the hearer recognizes the goal of making him believe that (4.2) is true, this goal is thereby satisfied. The literal goal of informing, therefore, is:

\begin{equation}
\text{GOAL}_{a} \text{BEL}_{h} \text{GOAL}_{a} \text{BEL}_{h} \text{BEL}_{a}(P)
\end{equation}

Given the literal goal, we derive the satisfaction of the discourse purpose as follows. The recognition of the literal goal satisfies it, and hence from (4.3) we get:

\begin{equation}
\text{BEL}_{h} \text{GOAL}_{a} \text{BEL}_{h} \text{BEL}_{a}(P).
\end{equation}

\footnote{(Cohen and Levesque 1985)}

\footnote{In a later publication (Cohen and Levesque 1988), Cohen and Levesque come to the conclusion that the specification of such context-independent consequences of utterances is impossible, and they change their account considerably. They were too hasty to do so, in my opinion (see Kronfeld 1990 pp. 164-166).}
From (4.4), on the assumption that the speaker is sincere, we get:

\[(4.5) \text{BEL}_h \text{BEL}_a \text{BEL}_s(P).\]

Moreover, since \(\text{BEL}_a \text{BEL}_s(P)\) implies \(\text{BEL}_s(P)\), we get:

\[(4.6) \text{BEL}_h \text{BEL}_s(P).\]

Finally, if the hearer takes the speaker to be competent with respect to \(P\), we get:

\[(4.7) \text{BEL}_h(P),\]

and the discourse purpose of inform is satisfied.

The case of request is somewhat more complex, but the reasoning is very similar. Take, for example, an utterance of the imperative "Close the door!" We can get the literal goal of the request by systematically weakening the discourse purpose until we arrive at a goal whose mere recognition is sufficient to satisfy it. Let \(P\) be the act of closing the door. Then the discourse purpose of the request is to make the hearer close the door, that is,

\[(4.8) \text{GOAL}_a \text{DONE}_h(P).\]

This, obviously, cannot be the literal goal. Recognizing (4.8), in itself, won't make the door shut. So perhaps the literal goal is that the hearer at least attempt to do it:

\[(4.9) \text{GOAL}_a \text{GOAL}_h \text{DONE}_h(P).\]

But this won't do either. The hearer may understand the request very well, but still refuse to cooperate. This leads to a weaker requirement: that the hearer believe that the speaker has a goal of making him (the hearer) adopt the goal of closing the door:

\[(4.10) \text{GOAL}_a \text{BEL}_h \text{GOAL}_a \text{GOAL}_h \text{DONE}_h(P).\]

But (4.10) still cannot express the literal goal, since the hearer may simply not believe that the speaker is sincere in his request. However, as was the case with informing, if the hearer understands the speaker, he cannot fail to see that the speaker has the goal expressed by (4.10). Hence, the goal of making the hearer believe that (4.10) is true is satisfied by its recognition, and thus the literal goal of the request is:

\[(4.11) \text{GOAL}_a \text{BEL}_h \text{GOAL}_a \text{BEL}_a \text{GOAL}_h \text{DONE}_h(P).\]

The reasoning that leads from the recognition of the literal goal to the satisfaction of the request is similar to the case of inform. Once the hearer recognizes (4.11), the literal goal is satisfied and we get:

\[(4.12) \text{BEL}_h \text{GOAL}_a \text{BEL}_h \text{GOAL}_a \text{GOAL}_h \text{DONE}_h(P).\]

From this, by assuming the speaker's sincerity, we get:

\[(4.13) \text{BEL}_h \text{BEL}_a \text{GOAL}_a \text{GOAL}_h \text{DONE}_h(P).\]
and since once a speaker believes he has a goal, he actually has it, it follows that

\[(4.14) \text{BEL}_h \text{GOAL}_h \text{GOAL}_h \text{DONE}_h(P).\]

From (4.14), by assuming that the hearer is helpful, we get:

\[(4.15) \text{GOAL}_h \text{DONE}_h(P),\]

and by assuming that he is also competent we can conclude that the request will (eventually) be satisfied:

\[(4.16) \text{DONE}_h(P).\]

Although Cohen and Levesque do not use the concept of a literal goal, their initial account has formulas similar to (4.3) and (4.11), representing the context-independent effects of declarative and imperative sentences on the hearer's mental state. These effects are incorporated into "feature" axioms, which in general have the following form:

\[(4.17) \text{If conditions } C \text{ hold, then after action } a \text{ effect } E \text{ obtains}\]

where conditions \(C\) specify central elements of a normal communication situation, \(a\) is the action of intentionally uttering a sentence with certain features, and \(E\) is a description of the hearer's mental state as a result of the utterance. Using this formulation, I now present simplified versions of Cohen and Levesque's declarative and imperative axioms

The declarative axiom states that if the hearer \(H\) believes that it is mutually believed by the speaker \(S\) and himself that (1) \(S\) has produced an utterance \(u\), (2) \(u\) is a declarative sentence whose conditions of satisfaction are expressed by \(P\), and (3) \(H\) is attending to \(S\), then after the act of producing \(u\), \(H\) believes that it is mutually believed by \(S\) and himself that \(S\) has the literal goal expressed by (4.3). The axiom itself is shown in Figure 1.

Similarly, the imperative axiom (Figure 2) states that if the hearer \(H\) believes that it is mutually believed by the speaker \(S\) and himself that (1) \(S\) has produced an utterance \(u\), (2) \(u\) is an imperative sentence whose conditions of satisfaction are expressed by \(P\), and (3) \(H\) is attending to \(S\), then after the act of producing \(u\), \(H\) believes that it is mutually believed by \(S\) and himself that \(S\) has the literal goal expressed by (4.11).

Given the declarative and imperative axioms, the discourse purposes of inform and request can be derived as follows: First, the mutually believed antecedent in each axiom is established\(^{16}\) As a result, the hearer can conclude that after the utterance, it is mutually believed that he has recognized the literal goal of the speech act. The discourse purposes are then derived, as we have seen, using the concepts of sincerity, helpfulness, and competence.

\(^{17}\)Ignoring, for the sake of simplicity, the important differences between goals, persistent goals, and intentions.

If $BMB^h_{\text{utter}}(u) & \text{imperative}(u, P) & \text{attend}_{h}(S)]$
then: $\text{AFTER}(\text{utter}_{h}(u),$

$\text{BMB}^h_{\text{GOAL}_{h} \text{BEL}_{h} \text{GOAL}_{h} \text{BEL}_{h} \text{GOAL}_{h} \text{DONE}_{h}(P)]})$

Figure 2: Imperative axiom.

5 The Referring Axiom\textsuperscript{19}

Given the Cohen and Levesque's approach, augmented with the notions of individuating sets and literal goals, we can now show how intended referring effects are achieved. But first, we need some additional notation:

- \text{content}(np): An open sentence with one free variable which represents the descriptive content of np (a noun phrase).
- \text{holds}(F, a): F is an open sentence with one free variable. When F is interpreted as a one-argument predicate, it is true of a.

As individuating sets of presentation modes are at the heart of the view of referring presented here, we also need a mechanism that will allow us to represent formally how an agent might reason about such entities. The following is merely a convenient notation, as a formal theory of individuating sets is yet to be developed.

- \Delta(Agt, u, t): A function whose arguments are an agent Agt, a referring expression u, and the time of utterance t, and whose value is a discursive presentation mode that represents for Agt whatever individual the utterer of u had in mind. Discursive presentation modes are mental representations of objects that result from referring acts occurring in discourse (note that Agt can be either the speaker or the hearer of u).
- \Pi(Agt, p, t): A function whose arguments are an agent Agt, and a perceptual image p that was recorded at time t, and whose value is a perceptual presentation mode that results from a perceptual act (e.g., looking), and denotes the object perceived.\textsuperscript{20}
- \text{res}(m): A function that takes a mode of presentation m as an argument, and returns the object that is determined by the individuating set that includes m (res is in fact a composition of two functions: the first maps presentation modes onto individuating sets that contain them, and the second maps individuating sets onto the objects that they represent).\textsuperscript{21}

Figure 3 shows the referring axiom, which describes the effects on the hearer's mental state when a referring expression is used under normal conditions. The similarity in structure between this axiom and the imperative and declarative ones (Figures 1 and 2 above)

\textsuperscript{19}A much earlier version of this section, based on the principles outlined in section 2.3, was published in collaboration with Doug Appelt (Appelt and Kronfeld 1987).

\textsuperscript{20}Some representations can combine discursive and perceptual modes (e.g. the presentation mode that is acquired after hearing the description: "the painter who made this picture"), and these can be accounted for as the values of functions that take other representations as arguments.

\textsuperscript{21}The name of the function res (from the Latin res, "thing") is an acronym for REferent of an individuating Set.)
should be obvious. All three specify the conditions under which a literal goal is recognized. Thus, the usual preconditions that are necessary for the performance of a speech act apply here as well, namely, that the speaker should utter a sentence, and that the hearer should listen (line 1). The other preconditions, however, distinguish the referring axiom from the other two: the sentence uttered should contain a noun phrase that is interpreted as a referential expression (line 2). Being a referential expression is a "feature" of the noun phrase in the same way that being in the imperative mood is a feature of an imperative sentence. Thus, the referring axiom is as much a "feature" axiom as the axioms in Cohen and Levesque's system. There is, however, a significant difference. The "features" in Cohen and Levesque's axioms are syntactic, that is, they can be inferred by looking at the form of the sentence alone. Being a referential expression, by contrast, is not a syntactic property that the noun phrase wears on its sleeve, so to speak. How a referential expression is recognized as such is a difficult question which I do not address here.

If indeed the hearer assumes that he and the speaker mutually believe that all the preconditions are satisfied, then after the utterance (line 3), the speaker also assumes that it is now mutually believed that the literal goals of the utterance are recognized. The literal goal of referring is expressed in line (4): that the hearer should focus on the same object that the speaker has in mind. The literal goal of uttering a referential expression with certain descriptive content is expressed in lines (5-6): the hearer should think that the speaker intends him to believe that the speaker believes that the descriptive content of the noun phrase is true of the intended referent. The difference in the structure of propositional attitudes between the two literal goals is significant. The mere recognition that a noun phrase is intended to be interpreted as a referential expression provides the hearer with a representation of the same object that the speaker has in mind. But the literal goal of using a referential expression with a certain descriptive content is very similar to the literal goal of using a declarative sentence for the purpose of informing. Since the speaker may be either insincere or incompetent, the hearer may not believe that the descriptive content of the noun phrase is indeed true of the intended referent. As in the case of assertions, however, the speaker will then simply adopt a weaker belief about the hearer's mental state.

Once the literal goals are recognized, they are satisfied. But this, obviously, is only the first step. Now the hearer is expected to discover and satisfy identification criteria. Such criteria are derived from a variety of sources, including the syntactic and semantic properties of the utterance, the discourse, general world knowledge, and so on. Within the model,

22The axiom in Figure 3 modifies and combines together the Referring Schema and Activation Axiom presented in (Appelt and Kronfeld 1987).
however, such criteria can be represented as constraints on the relevant individuating set. For example, if visual identification is required, the identification constraint is that the individuating set should contain a perceptual presentation mode that will be acquired after the utterance. If we take \( np \) and \( t_u \) to be the noun phrase and the time of the utterance, respectively, this requirement can be expressed as follows:

\[
(5.1) \quad (\exists t)(\exists p)[\text{res}(\Pi(H, p, t)) = \text{res}(\Delta(H, np, t_u)) \land (t \geq t_u)]
\]

If (5.1) obtains, identification is complete, and the referring act is successful. If not, then the hearer has to devise an appropriate plan to bring about a situation in which (5.1) is true.

Of course, mere perception of the referent is not enough. The hearer must believe that the thing he sees is the same as the one the speaker has referred to, and the descriptive content of the noun phrase often plays a major role in establishing this. It is certainly possible, however, for the hearer to figure out the correct referent even when the description used is entirely wrong. This is simply the result of interpreting the descriptive content of the referring expression as merely a tool for identification, which the hearer is free to use at his own discretion. This is reflected in the referring axiom, since the satisfaction of the literal goal of referring depends neither on the descriptive content of the referring expression, nor on the hearer’s attitudes toward its correctness.

The difference between the literal goal of referring and the literal goal of using a noun phrase with a certain descriptive content also explains why referring can be interpreted as an act of informing. As Appelt observes, a referring expression can be used to inform the hearer that some property holds of the intended referent. For example, a speaker may point to a tool on the table and say: “Use the wheelpuller to remove the flywheel.” The speaker can rely upon the pointing action to enable the hearer to identify the referent of the wheelpuller, while the descriptive content of the referring expression serves to inform the hearer what kind of tool it is. Given the referring axiom, this ability to inform while referring should hardly be surprising. Not only can identification succeed independently of descriptive content, but that part of the referring axiom which expresses the bearer’s goals and beliefs concerning the descriptive content of the referring expression (lines 5–6) has the structure of the literal goal of inform (4.3 above). Thus, as a result of the referring act, the hearer believes that

\[
(5.2) \quad \text{GOAL} \subseteq \text{BEL}_{h}\text{BEL}_{s}(\text{holds}(\text{cont}(v), \text{res}(\Delta(S, v, t))))
\]

If the hearer believes that the speaker is sincere and competent, he may indeed be convinced that the referent has the property expressed by the noun phrase; if so, the speaker has not only referred, but informed as well.

Another aspect of referring is that it is supposed to contribute to the success of illocutionary acts. Referring acts are rarely performed in isolation. Typically, referring occurs as a step in the performance of a larger speech act, and the objectives that the referring act is expected to accomplish play a role in the agent’s more general intentions and goals. Thus, it is not enough to provide the referring axiom; the model should also characterize formally the contribution of referring to illocutionary acts.

A full theoretical treatment of this problem is well beyond the scope of this paper, but an example may illustrate how the effects of referring contribute to the desired effects

---

\[24\text{(Appelt 1985a; Appelt 1985b).}\]

\[25\text{Adapted from (Appelt and Kronfeld 1987)}\]
of the request. Suppose $S$ tells $H$ under appropriate circumstances: “Please replace the 300-ohm resistor.” Let $u$ be this utterance, let ‘the 300ΩR’ be an abbreviation for the noun phrase “the 300-ohm resistor,” and let $\lambda x.300ΩR(x)$ be the descriptive content of that noun phrase. A natural way to combine the imperative and the referring axioms together is shown in Figure 4.

The formula in Figure 4 is not as complex as it seems. Lines (1) through (7) are the same preconditions as for both the referring and the imperative axioms. If these preconditions are satisfied, then after the utterance (line 8) the hearer assumes that it is mutually believed that he has recognized three literal goals: (a) the literal goal of referring (lines 9–10), (b) the literal goal of using a referring expression with certain descriptive content (lines 11–13), and (c) the literal goal of the request (lines 11, 14, 15).

Let us suppose that all the preconditions are mutually believed to be satisfied. Given this assumption, we can conclude that after the utterance, the following holds:

$$\frac{\text{If: } \text{BMB}_H^{\text{GOAL}}(\text{utter}_h(u)) \&}{\text{then: } \text{AFTER}(\text{utter}_h(u))}$$

The literal goals are thus recognized (and hence satisfied). Using sincerity, competence, and helpfulness, together with basic properties of BMB and BEL, it is easy to show that the following formulas are derivable from (5.3):
(5.4) \[ BMB^R_0[GOAL_h(res(\Delta(S, 'the 300\Omega R', t))) = res(\Delta(H, 'the 300\Omega R', t)))] \]

(5.5) \[ BMB^R_0[holds(\lambda x.300\Omega R(x), res(\Delta(S, 'the 300\Omega R', t)))] \]

(5.6) \[ BMB^R_0[GOAL_h DONE_h (replace_h(res(\Delta(H, 'the 300\Omega R', t))))] \]

(5.4) indicates that both speaker and hearer are focusing on the same thing; (5.5) provides the information that helps the hearer identify which thing it is; and (5.6) expresses what the hearer is supposed to do once he finds it.

6 Limitations and Extensions

Cohen and Levesque's framework, though very useful as we have seen, is fairly limited. Consider, for example, the Imperative axiom again. I used it to derive the effects of a request, given normal input/output conditions. But there is nothing in the axiom which distinguishes a request from, say, a command. The same problem occurs when we turn to the declarative axiom. It works for assertions, but there is simply no way the axiom can capture the important differences among the speech acts of suggesting, guessing, predicting, criticizing, dissenting, admitting or insinuating. The problem gets even worse when we turn to other types of illocutionary acts. Assertions and directives have neat syntactic counterparts: declarative and imperative sentences. But other than the optional usage of performatives, English sentences do not have general syntactic markers to distinguish, say, a promise from a prediction.

Moreover, we have used the notions of sincerity, helpfulness and competence in showing how the discourse purposes of speech acts are achieved. But the choice of these three concepts seems quite arbitrary. Sure, within the framework of Cohen and Levesque's logic, they fulfill very nicely the derivation function that was assigned to them. But how did we get them? Why these three and not others? Are there any others? We need a theoretical framework that not only justify the usage of these concepts as derivation rules in our logic, but also provides us with other derivation rules that our logic needs in order to express the full versatility of communication goals and effects.

In the remainder of this paper, I will outline possible directions where I'd like to go in the search for solutions to these problems. These directions seem extremely promising to me, but most of the work is still to be done.

Part of the problem in the framework I have presented so far is that it presupposes a rather limited theory of speech acts. Following Searle and Vanderveken's Foundations of Illocutionary Logic and Vanderveken's own two volume Meaning and Speech Acts, we can now be much more precise in the specification of the necessary and sufficient conditions that define each illocutionary act, and thanks to Vanderveken, we have a powerful logic to represent these conditions and reason about them.

In a nutshell, and skipping all the formal details, we know that there are five and only five types of illocutionary acts: assertives, commissives, directives, declaratives and

\( ^{26}(\text{Searle and Vanderveken 1985; Vanderveken 1990; Vanderveken 1991}) \)
expressives\textsuperscript{27}; that the general structure of an illocutionary act consists of propositional content and illocutionary force; and that the illocutionary force consists of six components that together uniquely define each and every illocutionary act.\textsuperscript{28} These components are:

1. The illocutionary point which in turns determines the direction of fit.
2. The mode of achievement.
3. The propositional content conditions.
4. The preparatory conditions.
5. The sincerity conditions.
6. The degree of strength.

Given this structure, there are six, and only six types of operations on illocutionary forces\textsuperscript{29}:

1. Restricting the mode of achievement.
2. Increasing the degree of strength of the sincerity condition.
3. Decreasing the degree of strength of the sincerity condition.
4. Adding propositional content conditions.
5. Adding preparatory conditions.
6. Adding sincerity conditions.

Moreover, corresponding to the five type of illocutionary acts, we have five primitive illocutionary forces of utterances, all sharing the following basic structure\textsuperscript{30}:

1. They have an illocutionary point.
2. They have no special mode of achievement.
3. They have the neutral degree of strength.
4. Their propositional content, preparatory and sincerity conditions are derived solely from their illocutionary point.

All the complex illocutionary acts are derivable from one of these primitive forces, using the aforementioned operations on illocutionary forces. For example, a request is derived from the primitive directive illocutionary force by adding a special mode of achievement according to which the hearer is given an option of refusal, while a testimony is derived from the primitive assertive illocutionary force by first adding the institution of the court as part of the preparatory condition, and second, by increasing significantly the degree of strength of the achievement of the illocutionary point and the degree of strength of the sincerity condition (see pp. 12–20). I have chosen the latest version, presented in (Vanderveken 1990, pp. 103–121).

\textsuperscript{27}(Searle 1979)

\textsuperscript{28}These components were defined somewhat differently in different stages of the theory development. Thus, in (Searle 1969) we have only preparatory, sincerity and essential conditions (see p. 66), while in (Searle and Vanderveken 1985) we have seven components, including a distinction between the degree of strength of the achievement of the illocutionary point and the degree of strength of the sincerity condition (see pp. 12–20). I have chosen the latest version, presented in (Vanderveken 1990, pp. 103–121).

\textsuperscript{29}(Vanderveken 1990, pp. 127–8).

\textsuperscript{30}op. cit. p. 125.
strength of the sincerity condition, given the fact that the speaker has taken the oath to
tell the truth, the whole truth, and nothing but the truth.

Armed with this rather elaborate structure (complete with Vanderveken's logic of illo-
cutionary acts and formal semantics of success and satisfaction) we can view our feature
axioms as a rough draft for the primitive illocutionary force of assertives and directives,
respectively. For example, the primitive assertive force is characterized as follows:

1. It has no special mode of achievement.
2. It has a neutral degree of strength (in contrast with a testimony which has a positive
degree of assertive strength or with a guess, which has a negative degree of assertive
strength).
3. It has the illocutionary point of representing an independently existing state of affairs
in the world.
4. It has propositional content, preparatory and sincerity conditions which are derived
from the illocutionary point alone. That is, since the illocutionary point is to represent
the world as it is, the speaker is expected to have evidence for what he says (preparatory
condition), is expected to believe it (sincerity condition), and to have no constraints
on the type of proposition he may assert (the empty propositional content condition).

Note that the declarative axiom does not mention either the mode of achievement or the
degree of strength. So by sheer omission it satisfies the first and second characteristics of
the primitive assertive illocutionary force. Moreover, since the discourse purpose implied
by the axiom is to make the hearer believe the propositional content, it follows that the
point of the assertion is to represent an independently existing state of affairs. Finally,
the Competence and Sincerity derivation rules used in achieving the discourse purpose of
the assertion are obvious manifestations of the preparatory and sincerity conditions of the
illocutionary force, respectively. Thus, the declarative axiom captures the essence of the
primitive illocutionary force of assertions. The imperative axiom lends itself to a similar
analysis and this is why we can view these feature axioms as a sort of finger exercise, before
Vanderveken's complex formal semantics of success and satisfaction kicks in.

The correspondence between our special derivation rules and components of illocutionary
force, hinted at in the last paragraph, is not accidental, of course. In fact, it is quite
general. Take, for example, the typical commissive speech act, namely, a promise. Part
of the definition of a successful promise is that the speaker is expected to intend to fulfill
it (sincerity), is expected to be able to fulfill it (competence), and is assumed to believe
that the promised act is beneficial to the hearer (helpfulness). Our special derivation rules,
therefore, can be placed (and justified) within a much larger theory. Moreover, the very
same theory can provide additional axioms and/or derivation rules, some general, some
specific to an illocutionary type or even to a particular illocutionary act. For example,
treating the primitive illocutionary forces as axiom schemas, we may be able to generate a
specific Request axiom (perhaps triggered by the occurrence of phrases such as “please” or
“could you”) which may allow the hearer to infer “I may safely refuse to do X” from “the
speaker has a goal of making me do X.”

Note, however, that the program of modeling speech acts as planned activities com-
plicates whatever derivation rules we can extract from Searle's Speech Act theory. Since
Searle's approach is speaker oriented (unlike Grice's), the concepts in his theory that corre-
spend to the special derivation rules are supposed to characterize properties of the speaker.
It is the speaker who is expected to be sincere, competent and helpful. But since the whole point of planning a speech act is to achieve effects on the hearer’s state of mind, these concepts must also characterize properties of the hearer as well. In other words, within a planning framework, we need to characterize speaker’s mental properties in order to explain how the hearer makes the right derivation and reach the appropriate mental state. But we also need the hearer’s mental properties to explain why it is rational for the speaker to expect his plan to succeed. Indeed, one of my goals using this approach is to show how Searle’s speech act theory must be welded with Grice’s program of communication intentions and implicatures precisely because we need both the speaker and the hearer’s point of view.

7 Reference and the Logic of Propositional Content

Let us return now to reference and referring. How does the cognitive analysis of these concepts fit within a generalized theory of speech acts?

One obvious possibility is first to combine Cohen and Levesque’s logic of propositional attitudes with Vanderveken’s logic of illocutionary acts, and then to apply the result to a computational model of referring. After all, like all speech acts, referring acts can be: (a) successful or unsuccessful and (b) their discourse purpose can either be satisfied or not. So we may very well be able to apply Vanderveken’s logic of illocutionary acts with its double sets of semantic values (success values and satisfaction values) to the act of referring as well.81

More significantly, however, the principles underlying a view about reference can (and should) be used to modify and augment the general theory of speech acts itself. After all, a theory of illocutionary acts is incomplete without a theory of propositional content, which in turn must incorporate the notion of reference as an essential relationship that determines what the content of illocutionary acts is about. Vanderveken’s logic of illocutionary force is indeed exciting and will prove to be very productive, I think. But his account of the logical form of propositions and his formalization of the semantics of propositional content is weak in my opinion, precisely because he does not pay enough attention to crucial problems of reference and referring.82 So the relationship between a formal theory of speech acts and an account of reference is entirely reciprocal. Not only can we use results from general speech act theory in our formalization of reference and referring, we also can (and should) use results from the analysis of reference to modify and strengthen weak areas in general speech act theory. Let me end this paper with a brief outline of what I have in mind.

My own analysis of reference, if correct, has two significant consequences for a theory of propositional content:

1. There is an important difference in structure between propositions as contents of utterances and propositions as content of propositional attitudes.

2. What determines the referent of a particular thought is an individuating set of presentation modes.

Once you examine closely the function and use of referring expressions in natural language, it becomes obvious that in many cases the speaker does not intend any particular

---

81 We should not be too hasty in extending Vanderveken’s logic this way. There is still a great difference between illocutionary acts and the non-illocutionary act of referring.

82 See (Kronfeld 1993).
representation to be recognized by the hearer. This is the case not only in most uses of proper names, demonstratives and indexicals, but in many uses of definite descriptions as well. In such cases the speaker does not mean any particular representation to be recognized and consequently, no representation can be part of the proposition that is the content of the utterance and the proposition, therefore, has the referent itself as a constituent (a so called singular proposition). Vanderveken's insistence (following Searle) that singular propositions are not possible contents of illocutionary acts is simply wrong. On the other hand, the idea that singular propositions can be the contents of thoughts, that is, that we can have beliefs about ordinary objects without having any individuating representation for them also does not work, as I have argued elsewhere. This means that the proposition which is a content of a belief must have an individuating representation of the object it is about as a constituent, unlike a proposition that serves as the content of an utterance. The only way out of this impasse, as far as I can see, is to drive a wedge between the semantics of utterances and the semantics of propositional attitudes. This, as one can imagine, has radical consequences for any theory that attempts to correlate propositional attitudes with the meaning of utterances that are used to express these attitudes.

Second, if I am right and the fundamental data structure by which an object is represented in thought is a set of representations (an individuating set) rather than a single representation in itself, then the semantics of propositional attitudes must itself be revised even further: in many cases there will not be any particular representation that will serve as a constituent of the proposition. In other words, the notion of an individuating set should be incorporated into a theory of the logical form of propositions (as contents of propositional attitudes). But this is yet to be done.

References

Choen, P. R. and H. Levesque. 1990. Intention is choice with commitment. Artificial Intelligence, 42(3).

33 (Vanderveken 1990, p. 97).
34 (Kronfeld 1990, pp. 33–44).


