Reference and Computation: An Essay in Applied Philosophy of Language

Amichai Kronfeld
(Natural Language Incorporated)

Cambridge: Cambridge University
Press (Studies in Natural Language
Processing), 1990, xxi + 185 pp.
Hardbound, ISBN 0-521-36636-4, $49.50;
Paperbound, 0-521-39982-3, $14.95

Reviewed by
John Barnden
New Mexico State University

1. Summary of the Book

Kronfeld's book attacks the problem of *referring*: How do speakers reveal what entities they are talking about? How do they choose particular referring expressions in their utterances? These questions are asked from the perspective of a plan-based theory of speech acts and communication. Utterances are viewed as acts that are intended by speakers to have certain effects on the addressees, and on the world more generally. Utterances are therefore to be planned in broadly the same way as other types of acts are.

Kronfeld's embedding of referring in speech act theory rests partly on a Gricean formulation of the *literal goal* of a referring expression. The relationship of this goal to the looser notion of the *discourse purpose* of the expression is discussed. Another aspect of the speech act theme is the discussion of two ways in which a referring expression can be intended by a speaker to be *relevant*: functionally relevant or conversationally relevant. Functionally relevant expressions are used primarily to lead the addressee to identify an object. Conversationally relevant descriptions are those that are intended to focus the addressee on a specific aspect of an object. Such descriptions are related to a specific type of Gricean conversational implicature.

Kronfeld carefully draws a distinction between the problem of referring and the philosophical problem of *reference*. The latter problem is cast as the question, "How can thoughts (and sentences that articulate them) be *about* objects?" (p. 13). Kronfeld covers quite a lot of ground on the philosophical problem of reference, as the problem of referring is somewhat dependent on it. Kronfeld concentrates throughout on reference to physical objects, and on referring expressions that are noun phrases. Furthermore, almost all the book is about definite descriptions, names and other expressions that refer to *single*, specific, physical objects, rather than to indefinite objects, several objects, or sets of objects. Kronfeld also stresses that he is not directly concerned with the anaphoric linking of pronouns to other noun phrases. He is directly concerned only with links between noun phrases and the world.

In fact, much of the book is a defense of the *descriptive* approach to the philosophical problem of reference in thought and in language. In this approach, to refer to an object is essentially to have or invoke a mental representation of that object. The relationship between a sentence or thought and the objects it is about is that of denotation, which in turn is a function of descriptive content. The descriptive approach

is contrasted to the *causal* approach to reference, whereby reference rests on a causal chain in the world leading to the referring thought or natural language expression. Kronfeld does not directly seek to argue against the causal approach. His defense of the descriptive approach is centered on countering objections to it that arise out of Donnellan's well-known distinction between attributive and referential uses of expressions (Donnellan 1971). Kronfeld's fairly lengthy analysis of the distinction, arguing that in fact it conflates three distinctions, leads to a framework for the computational approach that is developed in the book.

Kronfeld's approach to the question of what it is for a thought to be about an object is based on individuating sets (Appelt and Kronfeld 1987). Such a set is a set of mental representations, belonging to an agent, that are all believed by the agent to denote the same object. Kronfeld includes in his defense of the descriptive approach the claim that it can countenance mental *indexicals*—mental representations can be relative to the *I* and the *now*, for instance. This stance is adopted to defuse some of the criticisms of the descriptive approach, which have taken it to exclude such relativity.

The book also devotes a lot of time to the question of an addressee's interpretation of, and a speaker's choice of, referring expressions in sentences that report beliefs of agents. The question of speaker choice is related to the distinction between functional and conversational relevance mentioned above. The purpose of the discussion of belief reports is to show that certain types of report that have been taken as problems for the descriptive approach are not in fact so, but are just examples of the use of misleading, pragmatically inappropriate, and therefore ill-chosen, referring expressions.

Kronfeld is explicitly interested not just in the philosophical aspects of the referring problem, but also in the computational aspects. Quite understandably, he does not seek to provide a fully worked out computational account of referring. Rather, he provides a general framework and some useful first steps toward the construction of such an account. The last chapter of the book briefly describes a Prolog question-answering program, BERTRAND, that captures some basic computational features of Kronfeld's approach to referring. The system is admitted to be very limited, but it manipulates individuating sets to some degree, links referring expressions from different sentences to each other, and chooses reasonable referring expressions in answers. The bulk of the chapter is, however, devoted to formalizing the literal goals of referring expressions and showing how discourse purposes are inferentially related to those goals. Here Kronfeld presents a modified version of the formal treatment in Appelt and Kronfeld (1987). The treatment is based on the speech act formalization of Cohen and Levesque (1985). Kronfeld explicitly repudiates a later, modified formalization by those authors (Cohen and Levesque 1988) in favor of the earlier one. The treatment makes heavy use of the notion of *mutual belief*, encapsulated in a modal-logic operator, as well as using various other modal operators.

2. Overall Evaluation

The book should be of interest not only to computational linguists but also, to some extent, to linguists *tout court*, philosophers of mind or language, and knowledge representation researchers. The book should be useful as a recommended text in graduate courses.

I have objections to some fairly central precepts or stances in the book, and I feel that it could and should have gone further in focusing on computational issues. Having said this, my overall impression of the book is positive, in the sense that it is worth reading for some interesting and useful ideas. It is inevitable that a (partly) philosophical book will be met with vociferous complaints, so I would not wish my

objections to put someone off reading the book. Nevertheless, I set them out below to make it easier for prospective readers to assess the relevance of the work to their interests.

3. Main Positive Points

Kronfeld casts the discourse purpose that a speaker has for a referring expression as a matter of getting the addressee to understand what *identification constraints* are operative, and to make him or her generate an individuating set obeying those constraints. For instance, the speaker may use the phrase *George Bush* with the intention of getting the addressee's individuating set to contain the mental description "president of the U.S." On another occasion, however, the constraint might be that the individuating set contain a physical description of George Bush, to allow the addressee to pick him out of a crowd. Other possible constraints are listed. I found this pragmatic, context-dependent approach to referring an appealing one.

Kronfeld's approach promises to make good sense of what happens when a speaker has a mistaken belief. Suppose the following is the case. Sister Angelica holds the world record for spaghetti-eating, but Ralph mistakenly believes himself to be the holder of that record. Ralph (correctly!) believes that he has had sex with his wife. Hence, it is plausible that Ralph has come to believe that the spaghetti-eating record holder has had sex with his (Ralph's) wife. The problem is to block the inference that Ralph believes of Sister Angelica that she has had sex with his wife. According to Kronfeld, however, what determines the object of a belief is not a single mental representation, but the entire individuating set of which it is a member. Thus, in Ralph's belief that the record holder has had sex with his wife, the mental representation "the record holder," although itself actually denoting Sister Angelica, is a member of an individuating set, in Ralph's mind, that altogether denotes Ralph himself because it contains rich, correct descriptions of himself.

The discussion of Donnellan's distinction in Chapter 3 is interesting and useful. The original distinction is between the attributive use of a referring expression, such as when *Smith's murderer* is intended to mean *Smith's murderer, whoever he is*, and the referential use, as when *Smith's murderer* is intended to make the addressee think of a specific person, even though it may happen that that person is not in fact Smith's murderer (and *even* in some cases in which the speaker knows that). Kronfeld argues that this view conflates three conceptually independent distinctions. The first concerns how much the speaker knows about the referent. The second is to do with whether the speaker intends to focus the addressee's mind on a specific aspect of the referent. The third is to do with whether the speaker intends the referent of a definite description to actually satisfy the description. Kronfeld uses these distinctions to structure a computational view of referring. He postulates that the first is to do with the speaker's knowledge base, the second with the speaker's planner, and the third with the speaker's utterance generator.

4. Main Reservations

Kronfeld concentrates throughout on reference to *physical objects*. The stated reason is that the relationship between a referring phrase and a more intangible entity, such as the presidency, is a more difficult matter. But a technical problem arises from undue concentration on physical object reference: it blinds one to certain issues that really need to be dealt with at present rather than later. For instance, in Chapter 6 Kronfeld provides a complex explication of Ralph's having a *de re* belief about an object *o* that it is

F. The explication is in terms, partly, of *o* being the only object that has some property ϕ and Ralph's believing that the object, *qua* being the ϕ is *F*. However, Kronfeld nowhere addresses the issue of whether *this* belief is about ϕ. If it is, what does it mean for a belief to be about a *property*, and why do we not have to have a similarly complex explication of this aboutness? In other words, Kronfeld has supposedly answered the question of belief-aboutness in the case of (physical) objects only on the implicit assumption that other aboutness issues do not exist or are resolved. The very existence and implicitness of the assumption is a side effect of the physical-object bias.

In this connection, I found it strange that there was so much concentration in the book on what the conditions under which a belief is, *in fact*, about an object. From the point of view of someone interested in the psychology or engineering of communication, I would think that the primary point of interest is the conditions under which *people believe* beliefs (and phrases) to be about objects. That is: the primary point is the psychology of reference attribution (cf. Hornstein's 1984 discussion of reference). Indeed, one needs an argument that there is any scientifically coherent, objective notion of aboutness in the first place, as opposed to commonsensical views held by speakers and addressees. The omission of these considerations is puzzling in view of Kronfeld's emphasis on embedding of the referring issue in an account of communication based very heavily on plans and beliefs.

There is some implicit prevarication on whether beliefs (of the sort central in the book) involve explicit mental representations or not, as opposed to being explicit things that *could* arise in the mind as a result of current mental representations. Although most of the book seems strongly to imply that each belief of interest is a matter of holding certain explicit mental representations (certainly the talk of individuating sets being involved in beliefs suggests this), some parts seem to go the other way. For instance, in a footnote (p. 38) Kronfeld subscribes to the *positive introspection* of belief. That is, if you believe something, then you believe that you believe it. So you are always in the state of believing infinitely many different things. Doesn't this populate the mind with an infinite number of mental representations? Surely Kronfeld must have it in mind, in fact, that not all the beliefs he talks about are based on explicit mental representations. But the issue is nowhere discussed.

Kronfeld may have fallen into a certain well-camouflaged trap, albeit one already occupied by many other explorers. This is the trap of ascribing to ordinary people quite arcane beliefs (Barnden 1986, 1989). For example, the formulation of the above-mentioned second distinction in the analysis of Donnellan's distinction includes the phrase "[the speaker] intending the referring expression to be interpreted as a rigid designator." Does Kronfeld really mean here that *ordinary people* (from which I exclude philosophers!) are aware of the notion of rigid designator? Well, they certainly aren't, consciously, though I suppose they might be unconsciously—but such a possibility would need some argument! Now, of course, Kronfeld might actually have in mind, in the second distinction, something more on the lines of the more commonsensical explication "[the speaker] intending the referring expression to be interpreted as picking out a specific entity without implying importance for any particular property." However, if this is what Kronfeld really meant, he would have done well to say so.

There are many other places where Kronfeld does not guard against giving the appearance of ascribing arcane propositional attitudes to people. For instance:

- "The literal goal of the referring act is to make the hearer generate a local individuating set that..." (p. 75). (Does an ordinary speaker really think about individuating sets?)

- $\mathbf{BEL}_{\mathbf{ralph}}(\exists z)(\phi(z) \wedge (\forall w)(\phi(w) \rightarrow w = z) \wedge F(z))$ (p. 122). (Do people really and commonly have beliefs of this form and complexity?)

- The (seeming) implication on page 157 that when agent A believes that A and B mutually believe that P, A actually believes infinitely many things: that P, that B believes P, that B believes A believes P and so on. (The issue here is linked to my earlier comment about infinite sets of beliefs.)

- Several places in Chapter 7 have the addressee believing things such as this: it is mutually believed by speaker and addressee that the speaker has the goal that the addressee believe that the speaker has the goal that the addressee believe that the speaker believes that P.

As regards the addressee belief in this last observation, the feeling that Kronfeld really does mean to impute such highly complex beliefs to addressees is backed up by the statement on page 161. There he says that if an addressee, let's say Adolph, understands the sentence *Close the door!* then Adolph *cannot fail to see* that the speaker has the goal of making the addressee believe that the speaker has the goal of making the addressee have the goal of closing the door. This seems entirely wrong. Even if one were to grant that the average Adolph was intellectually *capable* of seeing it—that is, that he was, perhaps unconsciously, a speech-act theory expert—it does not follow that he actually *does* see it on any given occasion. But it's certainly not obvious that the average addressee is even capable of seeing it.

The elaborate belief in the last item of the list just displayed is one of the *simpler* beliefs in the speech-act part of Chapter 7, and, moreover, Kronfeld has already applied certain simplifications for expository purposes. Now, Kronfeld is well aware of this elaborateness, and says that it is "daunting" from the point of view of implementation. He attempts to excuse himself by stating that "complexity of statements in theory does not necessarily dictate a similar complexity in practice"; and he continues by saying that under certain default assumptions the formalization can be simplified considerably. Unfortunately, he fails to detail the simplifications, which in my view are crucial. Although he does not say so very clearly, I gather that the main point here is that real communicants need not conjure with mental representations as complex as the formulae he presents. (Notice that Grice (1957) says that he disclaims "any intention of peopling all our talking lives with armies of complicated psychological occurrences," although Kronfeld himself does not mention this.) Given that this is the case, surely it is precisely the question of what mental representations communicants conjure with, and how they do so, that should be of most interest in a computational treatment. That is, it is the *practice*, not the theoretical idealizations, that needs to be attended to. The lack of attention to actual practice in the speech-act part of Chapter 7 is the main respect in which I feel the book is computationally emaciated.

Part of the badly needed computational flesh is some explanation of the sense in which people have beliefs about mutual beliefs. Does Kronfeld take a thought about mutual belief to be some sort of infinite collection of thoughts (as in the penultimate item in the above list)? Or does the thought incorporate an algorithm that can generate those thoughts on demand? Or do people think about mutual belief via some simple internal symbol, analogous to a modal operator? In this last case, does the symbol stand for the official notion of mutual belief, or is that notion just an idealization of a much more commonsensical notion that people think in terms of?

5. Miscellaneous Points

The citations into relevant recent work in philosophy are rather sparse (despite the subtitle of the book). In particular, I was disappointed not to see any mention of highly relevant works such as Fauconnier (1985), Heny (1981), Hornstein (1984), and Schiffer (1987). Another highly relevant philosophy reference is Richard (1990), especially as it uses discourse context in a crucial way in a pragmatic approach to belief reports. Since that book is contemporaneous with Kronfeld's book he cannot be held to task for not citing it, but the reader would be advised to consult Richard's book to get a fuller impression of mental-representation approaches to, and some pragmatic aspects of, belief reports.

The index looks quite good, although I did not need to use it much. I did happen to notice that early mentions of mutual beliefs are not indexed, and that mentions of possible worlds are not indexed at all, despite the importance of these two notions in the exposition.

The book is very readable, being written in a clear style, not being too long (174 pages of main text), and having its formalizations confined mainly to the last chapter. The end-of-chapter summaries are extremely useful. I was, however, puzzled that there was no concluding chapter summarizing the whole book and gathering together future research possibilities. There is a foreword by John Searle, which conveys the flavor of the work well. A prospective reader could get a decent impression of many of the detailed claims of the book by spending an hour reading the foreword and the chapter summaries.

References

Appelt, Douglas E., and Kronfeld, Amichai (1987). "A computational model of referring." *Proceedings, 10th International Joint Conference on Artificial Intelligence.* Milan. 640–647.

Barnden, John A. (1986). "Imputations and explications: Representational problems in treatments of propositional attitudes." *Cognitive Science*, **10**(3), 319–364.

Barnden, John A. (1989). "Towards a paradigm shift in belief representation methodology." *Journal of Experimental and Theoretical Artificial Intelligence*, **2**, 133–161.

Cohen, Philip, and Levesque, Hector (1985). "Speech acts and rationality." *Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics.* Chicago. 49–59.

Cohen, Philip, and Levesque, Hector (1988). "Rational interaction as the basis for communication." Technical Report 433, Artificial Intelligence Center, SRI International.

Donnellan, Keith S. (1971). "Reference and definite descriptions." In *Readings in the Philosophy of Language*, edited by J. F. Rosenberg and G. Travis. Englewood Cliffs, NJ: Prentice Hall. 195–211.

Fauconnier, Gilles (1985). *Mental Spaces: Aspects of Meaning Construction in Natural Language.* Cambridge, MA: The MIT Press.

Grice, H. P. (1957). "Meaning." *Philosophical Review*, **66**(3), 377–388.

Heny, Frank (ed.) (1981). *Ambiguities in Intensional Contexts*, Dordrecht: D. Reidel.

Hornstein, Norbert (1984). *Logic as Grammar.* Cambridge, MA: The MIT Press.

Richard, M. (1990). *Propositional Attitudes: An Essay on Thoughts and How we Ascribe Them.* Cambridge, England: Cambridge University Press.

Schiffer, Stephen (1987). *Remnants of Meaning.* Cambridge, MA: The MIT Press.

John Barnden is an associate professor in Computer Science at New Mexico State University. He is conducting research in belief representation, mental-state metaphor, and connectionism. He received his M.A. in mathematics and his Diploma in computer science from University of Cambridge. Barnden's address is: Computing Research Laboratory, New Mexico State University, Box 30001/3CRL, Las Cruces, NM 88003; e-mail: jbarnden@nmsu.edu