Computer Rules, Conversational Rules

David Chapman*
Arris Pharmaceutical Corporation

There has been much controversy recently as to whether the rules of interaction discovered by conversation analysts are amenable to use by computers (Gilbert 1990; Hirst 1991; Luff, Gilbert, and Frohlich 1990). Button (1990) has argued that the rules of conversation are of a different ontological category than the rules used by computers, and that this means computers cannot be programmed to engage in conversation. Others (Fraser and Wooffitt 1990; Frohlich and Luff 1990; Gilbert, Wooffitt, and Fraser 1990) have argued to the contrary that the rules of conversation can be captured in a program, and indeed that some have been. I will argue for a third position. Button is right in his critique of existing attempts to import conversation analysis into computational linguistics and in his argument that there is a rule type mismatch. His arguments do not, however, show that computers cannot in principle be programmed to engage in conversation. I will argue by analogy to computer network protocols that an interactionist *computational interpretation of the conversation analytical rules is possible, and that Button's critique can thereby be bypassed.*

Button (1990) has argued that computers cannot engage in conversation because the rules of computation are of a different sort than the rules of conversation. The rules (or programs) that govern computers

- are explicitly represented,
- are causally efficacious, directly engendering the activities they describe, so that
- they cannot be violated, and thus
- have the force of mathematical laws.

The rules of conversation that have been discovered by conversation analysts (Heritage 1984), on the other hand,

- are typically not represented by their users,[1]
- are not causally efficacious, but nevertheless
- apply uniformly, even when they are violated, and
- have the force of social norms.

If these properties seem odd, let us consider some examples. On the first point, most of us are unaware of the role that gaze direction plays in the selection of the next

---

* Arris Pharmaceutical Corporation, 385 Oyster Point Boulevard, Suite 12, South San Francisco, CA 94080.
1 Once discovered, they may be represented by conversation analysts, who also of course use them.

speaker in three-way conversations (Goodwin 1980). A conversation analyst would argue that though it is logically possible that we unconsciously represent rules about gaze direction, there is no evidence for that. The other three points may be illustrated by the rule that you should shake the hand of a person you are introduced to. This rule is not a physical law; you are able to violate it at will. Nevertheless, the rule applies even when you have violated it: the person you have been introduced to is liable on the basis of the rule to consider you rude. You are, then, liable to be held to account for the violation; you may get an unfortunate reputation. Thus the normative character. These four properties are summarized by conversation analysts in two terms: people are said to *orient to* rules, rather than being governed by them, and rules are viewed as a *resource* in conversation, not a determining factor.[2]

Button argues that the incompatibility between these sorts of rules implies (1) that recent attempts to import rules from conversation analysis into computational linguistics are misguided, and (2) that computers cannot, in principle, participate in conversation. I believe he is right on the first count and wrong on the second.

The book *Computers and Conversation* (Luff, Gilbert, and Frohlich 1990) describes several systems (Frohlich and Luff 1990; Gilbert, Wooffitt, and Fraser 1990) that attempt to incorporate rules taken from conversation analysis into natural language interface systems. In these systems, conversational rules, such as those of turn taking, are explicitly represented as data structures in a grammar or formal rule system and are used as the basis of a natural language processing program. Since these rules are explicit, causally govern action, and cannot be violated, they are of a quite different character than the conversation analytical rules that inspired them.

Does this matter? It depends on one's purposes. For building better human/computer interfaces, this transformation in rule type will be perfectly justified if indeed it results in interfaces that are easier to use than ones not inspired by conversation analysis. This is an engineering question, not a foundational one, and it can only be answered empirically, not analytically. If, however, one wishes to better understand human interaction by computational modeling, the transformation is indeed troubling. The four properties of conversational rules seem to be central characteristics of human action more generally (Dreyfus 1979). We are, as Button says, "going up a blind alley" if we continue to ignore this mismatch in rule type.

The easiest response to this difficulty is to suggest, as Hirst (1991) has, that

> Button seems to be saying nothing more than that [conversation analytical] rules must be represented declaratively, not procedurally, so that they can be deliberately and explicitly invoked with a view to their effect. But far from precluding their use by a computer, this argument suggests that they fit very nicely into present-day AI reasoning systems!

This approach has been pursued by Fraser and Wooffitt (1990). They propose that conversational rules are explicitly represented and manipulated by metarules (Davis 1980). The metarules can choose to violate base-level rules under appropriate circumstances. Thus the base-level rules are not causally efficacious and do not directly determine action. They can, however, be used when violated to explain another agent's actions.

However, this valiant attempt fails to capture the conversation analytical notion of rule. First, Fraser and Wooffitt's rules are still representations. The conversation analytical perspective is not that rules should be represented declaratively, but that they

---

2 For one attempt to explicate these ideas in an AI context, see Agre and Chapman (1990).

should not be represented *at all*.[3] Second, though individual rules in the implemented system are not causally efficacious, the set of them is; the logic of the group of them cannot be violated. Third, there is no account of the sense in which the rules have the force of social norms.

We have thus far considered Button's first claim, that the type mismatch between conversational and computer rules means that direct translation of the former into the latter falsifies their nature. Let us now consider his second claim, that this implies that computers cannot engage in conversation.

Button hasn't shown that computers *can't* orient to rules, just that in current AI practice they *don't*. To see why Button's objections need not be fatal, we need to understand the interactionist perspective of conversation analysis, and to see how this perspective might be interpreted computationally.

Although the subject matter of conversation analysis is roughly the same as that of computational linguistics, the goals of the two fields are fundamentally different. Conversation analysis does not seek explanations of linguistic behavior. It is concerned rather with describing patterns of interaction between people. Because it is not looking for causal explanations, and because it is concerned with inter-individual phenomena, it is not concerned with things-in-the-head such as representations. Conversation analysis does not deny that there are things-in-the-head; it is simply uninterested because they are seen as irrelevant to its goals. It is, thus, not part of cognitive science, and what counts as progress in each discipline does not look like progress to the other.

These ideas may be easiest to understand by way of an analogy. Consider a computer workstation running a network file system that lets you access files on a remote file server. The client and server communicate via a network file protocol such as FTP or NFS. This protocol is a set of rules that specify how the client and server are to interact. However, the protocol does not appear in the network file system implementation: it isn't a data structure, procedure, set of procedures, or any other thing-in-the-computer. It is merely a specification. In fact, the computer does not represent the protocol it uses. That's probably just as well, because representing and manipulating the protocol—as a set of first-order axioms in a theorem prover, for instance—would be a difficult and computationally inefficient way to build a network file system.

Having concluded that the protocol is not in the computer, should we suppose instead that it is in the environment? Or, by analogy, having abandoned the mentalist supposition that patterns of action result from representations of those patterns, must we accept the behaviorist supposition that action is patterned by stimuli in the environment?

The environment of a computer on the net is another computer on the net. But if the protocol is not in the one, it is not in the other either, so that doesn't help any. The protocol *is* written down in a natural language document called an RFP; but that doesn't play any role in the actual operation of the file system. The protocol may also be represented in the head of the file system's writer. These representations do play a causal (because historical) role in the operation of the network code; but not in the usual sense in which representations play a role in action in AI. The representations in the designer's head can change (he may forget the protocol) without affecting the operation of the network code.

To return to the conversational rules case, the observation that the network protocol is in the head of the programmer is irrelevant, because there's no programmer in whose head the conversation analytical rules would live. Similarly, although network

---

3 Hirst's confusion on this point is understandable; Button's exposition does not make the point explicit.

protocols are typically written down on paper, the rules of conversation mostly aren't because they mostly haven't been discovered yet.

Thus, the interactionist perspective of conversation analysis sidesteps the mentalism/behaviorism debate.[4] For conversation analysis, the phenomena of interest can be located neither in the head nor in the environment. Like network protocols, they are interactional. To understand *how* people do what they do, one has to know about things-in-the-head; but conversation analysis refuses to speculate about these, because it is interested only in *what* people do. Thus, for conversation analysis, rules are not causal agents, but descriptions of regularities in interaction. By analogy, one might observe network traffic and induce the structure of a protocol without any access to the programs that use the protocol. Indeed, in some cases this access might not help much; it is notorious that the uucp protocol is undocumented and very hard to induce by reading the convoluted code that uses it.

We can now diagnose the problems with existing computational interpretations of conversation analysis as symptomatic of a deeper problem: the systems retain a mentalist orientation, with their designers seeking to locate explanations of action in mental structures. This orientation is endemic in AI generally; but nothing precludes interactionist AI in principle (Agre and Chapman 1990). In fact, the network protocol analogy suggests starting points for a different computational approach to interpreting conversation analysis. We'll see now, in another example, how a rule governing computer network communication has three of the four properties of conversation analytical rules cited earlier: it is not represented, it is not causally efficacious, and it applies even when it is violated.

The fundamental rule of communication on an Ethernet local network is that *only one computer may transmit at a time*. If two computers try to talk at once, there is a "collision" and the messages are scrambled. However, such collisions are unavoidable and occur regularly. When a collision occurs, the transmitting computers detect it and engage in a "retransmit protocol" to rectify the problem.

This rule has all the cited properties of conversation analytical rules except being a social norm. First, as with the file protocol, the Ethernet controller does not anywhere represent or otherwise include the rule. The rule is too pervasive and fundamental a feature of the situation to require representation.

Second, representing the rule wouldn't be useful, in any case, because it is not causally efficacious; it *cannot* engender the constraint it imposes. The problem is that the rule is a constraint on global interaction, not on individual action. A computer does not know when another computer is about to transmit, so it can't avoid collisions. The rule "there shall be no collisions" could be enforced by a complex protocol that gave machines information about when other machines might start transmitting. But such protocols require significant design and don't just fall out of the interactional rule. Thus we see that representing an interactional rule is not always a help in conforming to it.

As for the third property, the Ethernet collision rule applies even when it is violated. When two computers do transmit simultaneously, the rule is used to interpret the resulting garble on the network, and the retransmit protocol is used to repair the trouble.

Thus the protocol is an interactionist (rather than mentalist) form of rule use, but it is undeniably computational. What, then, about conversation analysis and AI?

---

4 For a clear exposition of how a third alternative to mentalism and behaviorism is possible, see Preston (in press).

Button is right that the conversation analytical rules should not be represented as expert-system-like rules. But the fact that computers are *governed* by one sort of rules (programs) does not preclude their *orienting to* another sort (such as those of conversation analysis). Does the fact that the rules of conversation are not represented mean that we must eschew Lisp and use holistic neural networks? No. There's nothing mystical about the guts of a network file system: procedures manipulate data structures representing packets and connections and host addresses. Yet the program uses a protocol it does not represent.

Of course network communication is in almost all other respects unlike human conversation; it would be wrong to suggest that Ethernet controllers orient to the no-collisions rule. But this example suggests that if the fourth issue—the normative character of rules—were addressed, Button's argument may not hold water. I think this, and not the representational issue, is the hard and interesting challenge of conversation analysis for computational linguistics.

What does it mean that the rules of conversation have the force of social norms? I doubt that there can be a general answer to this question. Conversation analysts, following Garfinkel's ethnomethodological critique of the appropriation of common-sense categories like "social norm" as theoretical terms (Garfinkel 1991; Heritage 1984), would not even attempt to answer it. However, some elementary observations may point in the right direction. First, social action is *accountable* in the sense that one may be required to produce an account of the rationality of one's action. This requirement is relatively unproblematic; it could be argued that some existing AI systems produce such accounts. Second, when social interaction runs into trouble, as it regularly does, the participants are required to make an effort to find the location of difficulty, to determine which participant is responsible for the problem, and to take steps to repair it. Third, this process of trouble location and repair is not a mechanical one; it requires interactive work and a commitment to negotiating the specifics of concrete marginal cases.

I believe it is possible to build a natural language system whose rule use satisfies the first three criteria in the same way the Ethernet controller does, and whose action is arguably subject to social norms in virtue of producing accounts, repairing misunderstandings, and negotiating assignment of the location of difficulties. This would not show that computers can engage in conversation; there are many other obstacles. It would, however, demonstrate that the particular problems Button raises are not the stumbling blocks.

References
Agre, Philip D., and Chapman, David (1990). "What are plans for?" *Robotics and Automation*, 6, 17–34.

Button, Graham (1990). "Going up a blind alley: Conflating conversation analysis and computer modeling." In *Computers and Conversation*, edited by Paul Luff, Nigel Gilbert, and David Frohlich, 67–90. Academic Press.

Davis, Randall (1980). "Meta-rules: Reasoning about control." *Artificial Intelligence*, 15, 179–222.

Dreyfus, Hubert (1979). *What Computers Can't Do*. Harper and Row.

Fraser, N. M., and Wooffitt, R. C. (1990). "Orienting to rules." In *Proceedings, AAAI Workshop on Complex Systems, Ethnomethodology and Interaction*, edited by Nigel Gilbert, Boston, July 1990, 69–80.

Frohlich, David, and Luff, Paul (1990). "Applying the technology of conversation to the technology for conversation." In *Computers and Conversation*, edited by Paul

Luff, Nigel Gilbert, and David Frohlich, 187–220. Academic Press.

Garfinkel, Harold (1991). "Respecification: Evidence for locally produced, naturally accountable phenomena of order*, logic, reason, meaning, method, etc. in and as of the essential haecceity of immortal ordinary society, (I)—an announcement of studies." In *Ethnomethodology and the Human Sciences*, edited by Graham Button, 10–19. Cambridge University Press.

Gilbert, Nigel, editor. (1990). *Proceedings, AAAI Workshop on Complex Systems, Ethnomethodology and Interaction*, Boston, July 1990.

Gilbert, Nigel; Wooffitt, Robin; and Fraser, Norman (1990). "Organizing computer talk." In *Computers and Conversation*, edited by Paul Luff, Nigel Gilbert, and

David Frohlich, 235–257. Academic Press.

Goodwin, Charles (1980). "Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning." In *Language and Social Interaction, Sociological Inquiry*, edited by Don Zimmerman and Candace West, 50(3,4), 272–302.

Heritage, John (1984). *Garfinkel and Ethnomethodology*. Polity Press.

Hirst, Graeme (1991). "Does conversation analysis have a role in computational linguistics?" *Computational Linguistics*, 17(2), 211–227.

Luff, Paul; Gilbert, Nigel; and Frohlich, David, editors. (1990). *Computers and Conversation*, Academic Press.

Preston, Beth (in press). "Behaviorism and mentalism: Is there a third alternative?" *Synthese*.