Believable Advice

Jim Davis
Design Research Institute
502 Theory Center
Cornell University, Ithaca NY 14853
davis@dri.cornell.edu
phone 607-255-1134

2 Feb 1994

The Back Seat Driver (BSD) is an computer program that gives driving instructions in real time to the driver of a car[1]. BSD makes no pretense of artistic merit, and it need not convince the driver that it has feelings, but it does need to inspire the driver with confidence (it also needs to be worthy of confidence, but that's another story.). What does it mean for the driver to have confidence in the program? It means attibuting some kind of "mental state" to the program, thinking of it as aware of the driver's position and progress, and "seeing" the world in the same way the driver does. We can call this property "co-presence", suggesting that the program and the person are inhabiting the same world. Virtual reality systems achieve this effect by creating a synthetic world for the person to enter. The BSD instead must come to the driver's world.

Maintaining the illusion of co-presence required substantial effort in the design of the Back Seat Driver. BSD communicates with the driver through exactly one medium, synthetic speech. (It speaks to the driver, but it does not listen - speech recognition in a car being beyond the state of the art. The driver communicates to the program by pushing buttons on a keyboard, and by driving.) Spoken language is a subtle and flexible means of communication, and well suited to this task. Co-presence is better communicated by nuance than by explicit statement. (People would laugh if the system said "I'm right here with you." It sounds like something a therapist would say.)

The Back Seat Driver uses a number of linguistic features to provide the illusion of co-presence. One way we indicate co-presence through nuance is by using deictic pronouns. Deictics are words that "point" at something, for example "this", "that", or "these". The difference between "this" and "that" (and "these" and "those") is that "this" refers to something close. We use this in referring to landmarks. When the landmark is close, we use the proximal form (e.g. "these lights"); when distant, we use a brief noun phrase (e.g. "the next set of lights"). This is important. When a driver is stopped 30 meters back from a stop light, it may be literally true to that that is "the next set of lights", but it's not the way people are accustomed to speaking, and if the program spoke this way people would either lose confidence in the program, or worse yet misunderstand it.

A second means of conveying co-presence is to acknowledge the driver's actions. After the driver carries out an instruction BSD briefly acknowledges the act if there is time, and if the act was not so simple (e.g. continuing straight) as to need no acknowledgment. This acknowledgment is a short phrase like "Okay". Some drivers dislike acknowledgments, so they can be disabled, but most find the confirmation comforting. The timing of the acknowledgment does much to confirm the driver's sense that the program really knows where the car is. Another source of acknowledgment is the use of cue words in the instructions. It will often be the case that the route calls for the driver to do the same thing twice (e.g. make two left turns). The speech synthesizer we use (Dectalk) has completely predictable pronunciation, and if it were to speak the same words twice it would sound exactly the same, which might give the impression that the system is repeating itself because it is in error (like a record skipping). (Have you heard the joke where a completely automated airplane welcomes the passengers aboard, boasts of its reliability, and concludes "Nothing can go wrong. Nothing can go wrong. Nothing can go wrong..."?) The acknowledgments help to dispel this, but BSD also employs cue words such as "another" to indicate that it is aware of its earlier speech and the driver's previous actions.

Yet another means of conveying co-presence is to make occasional remarks about the road and the route. These remarks indicate that BSD is correctly oriented. As an example, when the road makes a sweeping bend to one side, BSD speaks of this as if it were an instruction ("Follow the road as it bends to the right.") even though the driver has no choice in what to do. BSD also warns the driver about potentially hazardous situations, such the road changing from one-way to two-way, or a decrease in the number of lanes. As with acknowledgments, these warnings can be disabled if the driver dislikes them. Other remarks have less to do with the route. We justify these by the maxims of cooperative conversations formulated by philosopher H. P. Grice[2], whose maxim of Quantity (part 1) says: "Make your contribution as informative as is required." Grice explains that one can convey information by appearing to flout the maxim. In this case, a driver can reason as follow: "The program said nothing of great significance, but it, like all cooperative agencies, obeys the maxim of quantity, and so if it had something important to say, it would say it. Therefore there is nothing urgently requiring my attention. So everything is well."

A final means of co-presence is that BSD notices when the driver does not follow its instructions. This is a case which pushes the limits of the illusion. BSD has no way of knowing why the driver did not take the requested action - it may be that the driver did not understand the request, or it may be that the driver has a good reason not to do it (e.g. a police officer was blocking the road.) BSD It **must** show that it is aware of the trouble, but it lacks the perceptual resources to discover the source of the problem, or the linguistic resources to simply inquire. It can't chastise the driver, because it does not know what went wrong, or why. What it does is make a carefully worded hedging remark that neither gives nor takes blame, but merely offers to find another route and resume giving instructions.

The illusion of co-presence is, in fact, an illusion. The program can't see. It "knows where it is" through a built in navigation system, and it "sees" by consulting a heavily annotated internal map that tells it what it would see, assuming both map and position are correct. Sometimes it was wrong, and usually, but not always, people noticed right away, probably because I was always in the car, ready to "break character" if need be. But I also saw drivers follow BSD's instructions even when they made no sense. I think it was partially just good luck that no one was ever injured.

To me, the issues BSD raises are that natural language offers a variety of means to convey subtle messages about the relation between a software agent and the client; and that if we going to build illusions we had better be very careful to establish safe bounds around them.

References

[1] James R. Davis. *Back Seat Driver: voice assisted automobile navigation.* PhD thesis, Massachusetts Institute of Technology, September 1989.

[2] H. P. Grice. Logic and conversation. In Cole and Morgan, editors, *Syntax and Semantics: Speech Acts*, volume 3, pages 41–58. Academic Press, 1975.