Syntactic Alignment Between Computers and People: The Role of Belief about Mental States

Holly P. Branigan (Holly.Branigan@ed.ac.uk)
Psychology, University of Edinburgh
Edinburgh, EH8 9JZ, UK

Martin J. Pickering (Martin.Pickering@ed.ac.uk)
Psychology, University of Edinburgh
Edinburgh, EH8 9JZ, UK

Jamie Pearson (Jamie.Pearson@ed.ac.uk)
Psychology, University of Edinburgh
Edinburgh, EH8 9JZ, UK

Janet F. McLean (Janet.McLean@ed.ac.uk)
Psychology, University of Edinburgh
Edinburgh, EH8 9JZ, UK

Clifford I. Nass (nass@stanford.edu)
Department of Communication, Stanford University
Stanford, CA 94305-2050

Abstract

People tend to mirror the syntax used by their interlocutors in dialogue. Given that people treat computers as "social actors" in many ways, we might expect them to mirror computers' syntax as well. We report an experiment in which naïve participants played a dialogue game in which they believed that they were interacting with either a person or a computer. In fact, in both cases their "interlocutor" was a computer program that produced pre-scripted utterances. Participants demonstrated a very strong tendency to repeat the syntactic form of their "interlocutor's" immediately preceding utterance in both conditions. It does not appear that beliefs about the mental states of one's interlocutor mediate between perception and production.

Introduction

In recent research on the cognitive aspects of interaction, there has been increasing interest in the nature of interactions between humans and artificial media. This is of great theoretical interest because it allows us to determine whether beliefs about the "mind" of an interlocutor affect communicative behavior.

It is clear that people do not generally believe that computers have human minds, but they behave towards them in ways that suggest that they are attributing human characteristics to them (Reeves & Nass, 1996). Reeves and Nass demonstrated that there is a clear distinction between people's self-reports of their beliefs about computers, and their actual behavior directed towards computers. For example, users report that they would not be polite to computers, but their actual behavior does not bear this out: Just as people avoid criticizing other people to their face in order to be polite, users provide more positive comments about the computer's performance when the computer asks about itself than when a different computer asks about it (Nass, Moon, & Carney, 1999).

A critical question in dialogue is the extent to which beliefs about the mental state of an interlocutor affect a speaker's behavior. The work of Nass and colleagues suggests strongly that at least some aspects of social cognition are not dependent on beliefs about mental states. However, it is not yet known whether the same is true of linguistic behavior. In this paper we consider the question of whether the linguistic content is the same in human interactions with computers as in those with other humans.

Alignment in dialogue

Some aspects of dialogue are likely to be intentional (e.g., deciding what to talk about, in broad terms), but other aspects must be fairly automatic, as there would simply not be time or resources available for interlocutors to make explicit decisions about every aspect of the process. However, theories of dialogue make different assumptions about whether the

automatic component is larger (Pickering & Garrod, in press) or smaller (Clark, 1996).

For example, Brennan and Clark (1996) found evidence that speakers in a series of dialogues came to use the same referring expressions to refer to particular objects. For example, they might call a shoe a *pennyloafer* or a *docksider*. Brennan and Clark suggested that this tendency towards lexical entrainment or alignment arose because the interlocutors established a shared conceptualization, or *conceptual pact*. Such pact-making has the aim of ensuring that interlocutors successfully achieve mutual understanding and is not considered to be an automatic process. We might expect non-automatic processes to be relatively susceptible to beliefs about the mental state of an interlocutor.

However, the tendency towards lexical alignment could also be explained in terms of automatic processing. Garrod and Anderson (1987) found the same tendency in a series of dialogues where pairs of speakers negotiated their way through mazes. They interpreted their findings in terms of what they called *input-output coordination*, whereby lexical and semantic rules used to interpret input are subsequently reapplied to formulate output. Garrod and Anderson also found little evidence that lexical alignment is the product of explicit negotiation. As one would expect if it is the result of an automatic coordination process, convergence in common expressions seemed to emerge naturally (see also Garrod & Clark, 1993). Thus, although lexical alignment may be in principle open to conscious manipulation, it does not usually appear to be so in practice. In fact, alignment appears to be a basic organizing principle of dialogue, and also occurs at other levels such as phonology, speech rate and syntax (Pickering & Garrod, in press), where conscious awareness of what individuals are aligning to appears to be particularly unlikely.

Early studies found evidence for syntactic repetition between interlocutors, but their effects might have been the result of priming at the lexical (or other) level (Levelt & Kelter, 1982; Schenkein, 1980). Branigan, Pickering, & Cleland (2000) excluded a lexical explanation in their demonstration of syntactic alignment between interlocutors. They used a dialogue "game" in which pairs of speakers took turns describing pictures to one another, and matching pictures to their partner's description. The pictures used three entities, and employed "dative" verbs like *hand* or *send*. Figure 1 shows an example of a picture used in the experiment. One speaker was in fact a confederate of the experimenter, and followed a predetermined script. On experimental trials, the confederate was scripted to produce a prime description with either a *prepositional-object* (PO) structure (e.g., *the pirate handing the cake to the sailor*) or a *double-object* structure (DO) (e.g., *the pirate handing the sailor the cake*)). Branigan et al. found that when naïve participants subsequently described a target picture that could also be described using these structures, they showed a strong tendency to use the same structure as the utterance they had just heard. The priming effects were extremely strong, with naïve participants repeating the confederate's sentence form 78% of the time when the verb was repeated between prime and target, and 63% of the time when it was not (the entities were always different).

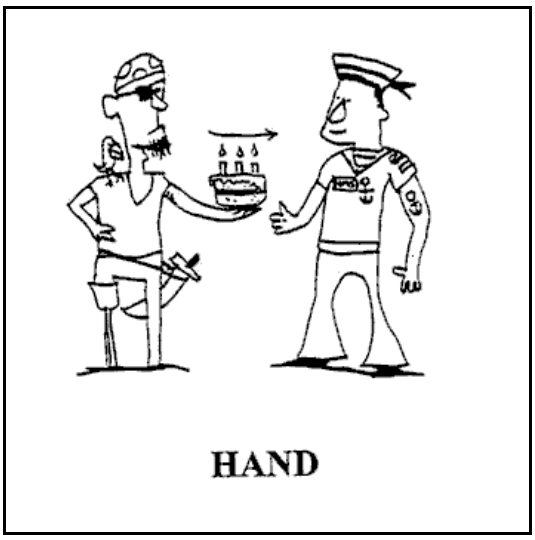

Figure 1: An example of an experimental picture depicting the ditransitive action "hand"

There is no evidence that such syntactic alignment is intentional. Speakers tend to repeat syntactic form without any indication that they are aware that they are doing so, although we do not know whether priming effects are enhanced by such awareness (see Pickering & Branigan, 1999 for a review).

Branigan et al. (2000) interpreted their results in terms of the activation of syntactic information: Comprehending a particular structure activates associated syntactic rules which are used in both production and comprehension. Residual activation of those rules raises the likelihood of their application in subsequent speech.

If syntactic alignment is a largely automatic process, then we would expect it to be relatively impervious to beliefs about the mental state of an interlocutor. That is, an utterance with particular syntactic characteristics will bring about the same effect on the addressee, regardless of the identity of the producer. For example, comprehending a PO sentence will automatically activate the syntactic rule(s) associated with the PO structure.

Thus, we can contrast two accounts of this process of alignment. According to one account, people tend to repeat syntactic form to an extent determined by "objective" features of the environment, such as the precise form of utterance used as prime, by the time between prime and target, and so on. But views about the interlocutor would not be relevant – it would not matter whether naïve participants believed they were interacting with another person or not. According to the second account, alignment is a strategy that people use because they believe it is beneficial in helping both interlocutors to reach mutual understanding, and which is, to at least some degree, under their control. Although they are not aware of exactly what they are aligning on, speakers can "switch on" a process whereby they adopt an aligned response where possible. However, this process is brought into play only when they believe they are interacting with another agent who will benefit from it.

In order to test between these accounts, we set up a situation in which a naïve participant played a version of Branigan et al.'s (2000) dialogue game on a computer, where they could not see their interlocutor. The experimental task was presented as a picture-matching and –describing game, in which they would interact with an unseen interlocutor by typing. The game involved alternate 'picture-matching' and 'picture-describing' turns. On a matching (prime) turn, the naïve participant read a picture description that they were told was sent to them by their "interlocutor"; they then had to decide whether it matched a picture displayed on-screen. On a describing (target) turn, they typed a description for a picture displayed on-screen; they were told that this description would be sent to their "interlocutor", so that their "interlocutor" could make a matching decision. Half the participants were told they were interacting with another person in another room, the other half that they were interacting with a computer. In both cases, the interlocutor's responses were pre-programmed. Thus, all aspects of the experiment were the same apart from the participants' belief about the interlocutor with which they were interacting. We manipulated the syntactic structure of the description that participants received on a matching turn, and examined how this affected the syntactic structure that they produced for the immediately subsequent describing turn. If alignment is intentional with respect to participants' beliefs about interlocutor identity, we would predict that participants would syntactically align to a greater extent when they believe that they are interacting with another person than when they believe they are interacting with a computer.

Experiment

Participants

Thirty-two members of the University of Edinburgh community were paid to participate.

Method

Items Items consisted of a picture and its written description followed by a second picture to be described. They were derived from the items originally used in Branigan et al. (2000). The picture together with its description was termed the *prime*. The picture to be described was termed the *target*. For each picture, the appropriate verb was printed beneath the picture. There were 96 items in total, 24 of which were experimental items. The experimental pictures depicted ditransitive actions involving an agent, a patient, and a beneficiary. Six ditransitive verbs were used (*give*, *hand*, *offer*, *sell*, *show*, *throw*), with each verb being used four times as a prime picture and four times as a target picture. Each prime picture and its description was paired with a second experimental picture to make one experimental item. The set of items in the *same-verb condition* consisted of prime and target pictures that involved the same verb. The set of items in the *different-verb condition* consisted of prime and target pictures that involved different verbs.

The remaining 72 items were fillers. The filler pictures depicted transitive actions involving an agent and a patient (e.g., a nun following a soldier). Eighteen transitive verbs were used, with each verb being used eight times as a prime and eight times as a target. Each filler picture and its description was paired with a second filler picture to make one filler item.

We prepared two sets of descriptions for each of the two verb conditions. In each set, half of the experimental prime picture were assigned PO descriptions of the form the *X verbing the Y to the Z*; half were assigned DO descriptions of the form the *X verbing the Z the Y*. Half of the descriptions matched exactly the prime picture. However, the remaining half of the descriptions did not match the picture. In these cases, there was a mismatch between one of the entities depicted in the picture and one of the entities mentioned in the description. For example, the picture of a nun following a soldier had the description *the nun following the clown*.

An item comprised a prime picture with a description followed by a target picture (and was defined by the target picture, with respect to the verb manipulation). We constructed two lists containing 24 experimental items and 72 fillers. Each list contained 12 experimental items with PO target descriptions and with 12 with DO target descriptions. Exactly one version of

each item appeared in each list. For each participant, we generated a randomized order of the materials.

The experiment thus had three factors: Prime Type of the prime description that the participant received (PO vs. DO); Interlocutor (computer vs. person interlocutor); Verb in the prime description and the target picture (same vs. different verb).

Procedure Participants were randomly assigned to one of the two conditions. In the *computer-interlocutor condition*, participants were told that they were to play a picture-description and -matching game with a computer that could 'understand' descriptions of pictures. In the *human-interlocutor condition*, participants were told that they were to play the game via a computer terminal with another person who was in another room. All other details of the experimental procedure were identical.

Participants sat in front of a PC, which presented the pictures and descriptions on-screen. Participants were told that they would be referred to as *B*, and that their task was to take it in turns with their interlocutor, who would be referred to as *A*, alternately describing pictures and deciding whether descriptions of pictures matched pictures that they saw.

Each trial comprised a 'matching' turn and a 'describing' turn. Participants were told that A would go first and produce a description to a picture that they 'saw', which the participant would match. On 'matching' turns, the screen initially displayed an empty text-box marked A along the bottom of the screen. A typed description 'from the interlocutor' then appeared in the box after a set period of time. This was the prime description. The time for the description to appear was always 5500 ms for experimental items, and varied between 3500ms and 5500ms for filler items. A picture then appeared on-screen above the description after 2000ms. Participants responded 'Y' or 'N' for whether or not the description and the picture matched. The screen then cleared for 1000ms.

On 'describing' turns, an empty text-box marked B appeared along the bottom of the screen. After 1000ms, the target picture appeared. Participants typed a description of the picture, which appeared in the box, and then pressed <ENTER>. Participants were told that this sent the description to the interlocutor, who had to decide whether the participant's description matched a new picture of their own. (This was supposedly the interlocutor's matching turn.) The screen then cleared for 1000ms and then a box marked A appeared along the bottom of the screen. Participants received simulated feedback on whether their interlocutor responded Y or N. Half of the time Y appeared in the box; half of the time N appeared. The Y or the N appeared after a time varying between 1000ms and 2500ms, and remained on-screen for 2000ms. The screen cleared for 2000ms before the next trial began.

There was a practice session consisting of six trails before the experiment proper. The practice had the same format as the experiment, but only items with transitive actions were included.

Results

Scoring We coded descriptions as either PO or DO. Responses were coded PO if the (inanimate) patient of the action immediately followed the verb and was followed by the preposition *to* and the (animate) beneficiary of the action; and DO if the (animate) beneficiary of the action immediately followed the verb, and was followed by the (inanimate) patient of the action. We excluded trials if the participant did not produce a description or if the description did not use the verb that was printed on the picture. In all, we excluded 7 responses (1% of the data). The proportions of PO target responses for the *same-verb* condition and the *different-verb* condition are reported in Table 1 and Table 2, respectively.

Table 1: Proportion of PO target responses in the *same-verb* condition

Interloc.:	Computer		Person	
Prime:	PO	DO	PO	DO
mean	.940	.093	.750	.177
priming	84.7%		57.3%	

Table 2: Proportion of PO target responses in the *different-verb* condition

Interloc.:	Computer		Person	
Prime:	PO	DO	PO	DO
mean	.861	.545	.781	.347
priming	31.6%		43.4%	

We conducted 2x2x2 ANOVAs treating participants (F_1) and items (F_2) as random factors. Prime Type (PO vs. DO prime) was within-participants and -items; Interlocutor (computer vs. person interlocutor) and Verb (same vs. different verb) were between-participants but within-items. ANOVAs on the proportions of PO target responses produced in each condition revealed a main effect of Prime Type (F_1 (1,28) = 186.568, $p<0.001$, MSe = 0.025; F_2 (1,23) = 525.159, $p<0.001$, MSe = 0.059) and an interaction between Prime Type and Verb ($F_1(1,28) = 17.092$, $p<0.001$, MSe = 0.025; $F_2(1,23) = 28.499$, $p<0.001$, MSe = 0.059). The three-way interaction between Prime Type, Verb and Interlocutor was also significant ($F_1(1,28) = 6.355$, $p<0.02$, MSe = 0.025; $F_2(1,23) = 10.095$, $p<0.005$, MSe = 0.059).

Importantly, 2x2 ANOVAs conducted for the computer interlocutor condition revealed significant effect of Prime, as did ANOVAs conducted for the

person interlocutor condition (all ps <0.001). There was also a significant effect of Prime in the 2x2 ANOVAs conducted for the same verb condition and different verb conditions (all ps <0.001).

Discussion

Our experiment demonstrated syntactic alignment in typed communication via a computer. Naïve participants taking part in what they believed to be a dialogue with an unseen interlocutor tended to produce typed descriptions that had the same syntactic structure as their "interlocutor's" previous description. Alignment occurred whether naïve participants believed they were interacting with another human participant in another room or with an unintelligent computer. As in earlier studies (Branigan et al., 2000), alignment occurred whether the prime and target employed the same verb or not, but priming was significantly stronger if the verb was repeated than if it was not.

The occurrence of priming of a similar magnitude to earlier face-to-face studies of syntactic alignment (Branigan et al., 2000), coupled with the increased priming effect when the verb was repeated (cf. Pickering & Branigan, 1998), strongly suggests that alignment processes in typed dialogue involving no other visible interlocutor is broadly similar to alignment in dialogue between co-present interlocutors who use speech to communicate.

The main purpose of this research was to contrast two accounts of alignment, one where interlocutors align so long as appropriate environmental conditions are met (e.g., that turn-taking is obeyed), the other where interlocutors only align with an interlocutor that they believe to have similar mental states to their own. In our experiment, we kept the environmental conditions constant but manipulated whether the naïve participant believed that the interlocutor was a person or a computer.

The results provided striking evidence that interlocutors align linguistically with what they believe to be a computer, and that the strength of this alignment is broadly comparable with the alignment that occurs when they believe themselves to be communicating with another person. This is consistent with Reeves and Nass's (1996) claim that people treat computers (and other new media) as social actors. It extends this work into the domain of language, and into a domain where participants are not normally aware of the nature of their responses (they do not generally appreciate the grammatical difference between the PO and the DO construction).

They also provide evidence for the claim that syntactic alignment is largely an automatic process. An account under which interlocutors make a "syntactic pact" akin to Brennan and Clark's (1996) "conceptual pact" when they mirror the syntax just used by their interlocutor becomes less considerably less likely as a result of the finding that syntactic alignment is very strong with a computer that they believe to be unintelligent.

More generally, the results suggest that the tendency toward imitation in language use is largely an automatic process that is unmediated by consideration of the mental states of the interlocutor. They are therefore compatible with accounts of alignment or imitation that presume no strategic activation of a decision component (Pickering & Garrod, in press). This fits with accounts of imitation that proceed via a perception-behavior "expressway" (Bargh & Chartrand, 1999; Dijsterhuis & Bargh, 2001).

Acknowledgments

This research was supported by Scottish Enterprise, ESRC grant no. R000239363, and a British Academy Postdoctoral Fellowship.

References

Bargh, J.A., & Chartrand, T.L. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462-479.

Branigan, H.P., Pickering, M.J., & Cleland, A.A. (2000). Syntactic coordination in dialogue. *Cognition*, *75*, B13-B25.

Brennan, S.E., & Clark, H.H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 1482-1493.

Clark, H.H. (1996). *Using Language*. Cambridge, England: Cambridge University Press.

Dijksterhuis, A., & Bargh, J.A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology*. San Diego: Academic Press

Garrod, S.C., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*, 181-218.

Garrod, S.C., & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes*, *8,* 101-126.

Levelt, W.J.M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14,* 78-106.

Nass, C., Moon, Y., & Carney, P. (1993). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology, 29,* 1093-1110.

Pickering, M.J., & Branigan, H.P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, *3*, 136-141.

Pickering, M.J., & Garrod, S.C. (in press). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*.

Reeves, B., & Nass, C..I. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Stanford, CA: Cambridge University Press.

Schenkein, J. (1980). A taxonomy for repeating action sequences in natural conversation. In B. Butterworth, *Language Production*, Vol. 1. London, England: Academic Press.