Why is conversation so easy?

Simon Garrod[1] and Martin J. Pickering[2]

[1]University of Glasgow, Department of Psychology, Glasgow, UK, G12 8QT
[2]University of Edinburgh, Department of Psychology, Edinburgh, UK, EH8 9JZ

Traditional accounts of language processing suggest that monologue – presenting and listening to speeches – should be more straightforward than dialogue – holding a conversation. This is clearly not the case. We argue that conversation is easy because of an interactive processing mechanism that leads to the alignment of linguistic representations between partners. Interactive alignment occurs via automatic alignment channels that are functionally similar to the automatic links between perception and behaviour (the so-called perception–behaviour expressway) proposed in recent accounts of social interaction. We conclude that humans are 'designed' for dialogue rather than monologue.

Whereas many people find it difficult to present a speech or even listen to one, we are all very good at talking to each other. This might seem a rather obvious and banal observation, but from a cognitive point of view the apparent ease of conversation is paradoxical. The range and complexity of the information that is required in monologue (preparing and listening to speeches) is much less than is required in dialogue (holding a conversation). In this article we suggest that dialogue processing is easy because it takes advantage of a processing mechanism that we call 'interactive alignment'. We argue that interactive alignment is automatic and reflects the fact that humans are designed for dialogue rather than monologue. We show how research in social cognition points to other similar automatic alignment mechanisms.

Problems posed by dialogue

There are several reasons why language processing *should* be difficult in dialogue. Take speaking. First, there is the problem that conversational utterances tend to be elliptical and fragmentary. Assuming, as most accounts of language processing do, that complete utterances are 'basic' (because all information is included in them), then ellipsis should present difficulty. Second, there is the problem of opportunistic planning. Because you cannot predict how the conversation will unfold (your addressee might suddenly ask you an unexpected question that you have to answer), you cannot plan what you are going to say far in advance. Instead, you have to do it on the spot. Third, there is the problem of making what you say appropriate to the addressee. The appropriateness of referring to someone as '*my next-door neighbour Bill*', *Bill*, or just *him* depends on how much information you share with your addressee at that point in the conversation. Does she know

Corresponding author: Simon Garrod (simon@psy.gla.ac.uk).

who Bill might be? Does she know more than one Bill? Is it obvious to both of you that there is only one male person who is relevant? Similarly, in listening, you have to guess the missing information in elliptical and fragmentary utterances, and also have to make sure that you interpret what the speaker says in the way he intends.

If this were not enough, conversation presents a whole range of interface problems. These include deciding when it is socially appropriate to speak, being ready to come in at just the right moment (on average you start speaking about 0.5 s before your partner finishes [1]), planning what you are going to say while still listening to your partner, and, in multi-party conversations, deciding who to address. To do this, you have to keep task-switching (one moment speaking, the next moment listening). Yet, we know that in general multi-tasking and task switching are really challenging [2]. Try writing a letter while listening to someone talking to you!

So why is conversation easy?

Part of the explanation is that conversation is a joint activity [3]. Interlocutors (conversational partners) work together to establish a joint understanding of what they are talking about. Clearly, having a common goal goes some way towards solving the problem of opportunistic planning, because it makes your partner's contributions more predictable (see Box 1). However, having a common goal does not in itself solve many of the problems of speaking and listening alluded to above. For instance, it does not ensure that your contributions will be appropriate for your addressee, alleviate the problems of dealing with fragmentary and elliptical utterances, or prevent interface problems.

One aspect of joint action that is important concerns what we call 'alignment'. To come to a common understanding, interlocutors need to align their situation models, which are multi-dimensional representations containing information about space, time, causality, intentionality and currently relevant individuals [4–6]. The success of conversations depends considerably on the extent to which the interlocutors represent the same elements within their situation models (e.g. they should refer to the same individual when using the same name). Notice that even if interlocutors are arguing with each other or are lying, they have to understand each other, so presumably alignment is not limited to cases where interlocutors are in agreement.

But how do interlocutors achieve alignment of situation models? We argue that they do not do this by explicit negotiation. Nor do they model and dynamically update every aspect of their interlocutors' mental states. Instead,

---

Box 1. Conversation as a joint activity

Both modern and traditional theories of dialogue argue that conversation can only be understood as joint activity [3,21]. In other words conversation necessarily involves cooperation between interlocutors in a way that allows them to understand sufficiently the meaning of the dialogue as a whole; and this meaning results from joint processes. Take, for example, the dialogue in the example below, which was recorded from two players (*A* and *B*) engaged in a collaborative maze task who are trying to communicate their positions on their different mazes [22]. Although it might look disorganised, the sequence of utterances is quite orderly as long as we assume that the dialogue is made up of a series of joint actions reflected in links across turns [21,23]. A question such as (12) calls for an answer such as (13) This means that production and comprehension processes become coupled. *B* produces a question and expects an answer of a particular type; *A* hears the question and has to produce an answer of that type. Furthermore, the meaning of what is being communicated depends on the interlocutors' agreement or consensus rather than on dictionary meanings [24] and is subject to negotiation [25]. This explains why overhearers not directly engaged in the dialogue have trouble understanding what is being said [26]. The coupling of production and comprehension processes in dialogue may go some way towards overcoming problems of opportunistic planning.

Example maze-game dialogue taken from Garrod and Anderson [22]. (Colons mark noticeable pauses of less than 1 s)

1   **B:** OK Stan, let's talk about this. Whereabouts – whereabouts are you?
2   **A**: Right: er: I'm: I'm extreme right.
3   **B:** Extreme right.
… …
8   **A:** You know the extreme right, there's one box.
9   **B:** Yeah right, the extreme right it's sticking out like a sore thumb.
10 **A:** That's where I am.
11 **B:** It's like a right indicator.
12 **A:** Yes, and where are you?
13 **B:** Well I'm er: that right indicator you've got.
14 **A:** Yes.
15 **B:** The right indicator above that.
16 **A:** Yes.
17 **B:** Now if you go along there. You know where the right indicator above yours is?
18 **A:** Yes.
19 **B:** If you go along to the left: I'm in that box which is like: one, two boxes down OK?

Box 2. Evidence for alignment in dialogue

Interlocutors become aligned at many different linguistic levels simultaneously, almost invariably without any explicit negotiation. At the level of the situation model, interlocutors align on spatial reference frames: if one speaker refers to objects egocentrically (e.g. *'on the left'* to mean on the speaker's left), then the other speaker tends to use an egocentric perspective as well [27]. More generally, they align on a characterization of the domain, for instance using coordinate systems (e.g. *A4*, *D3*) or figural descriptions (e.g. *T-shape*, *right indicator*) to refer to positions in a maze [22,28].
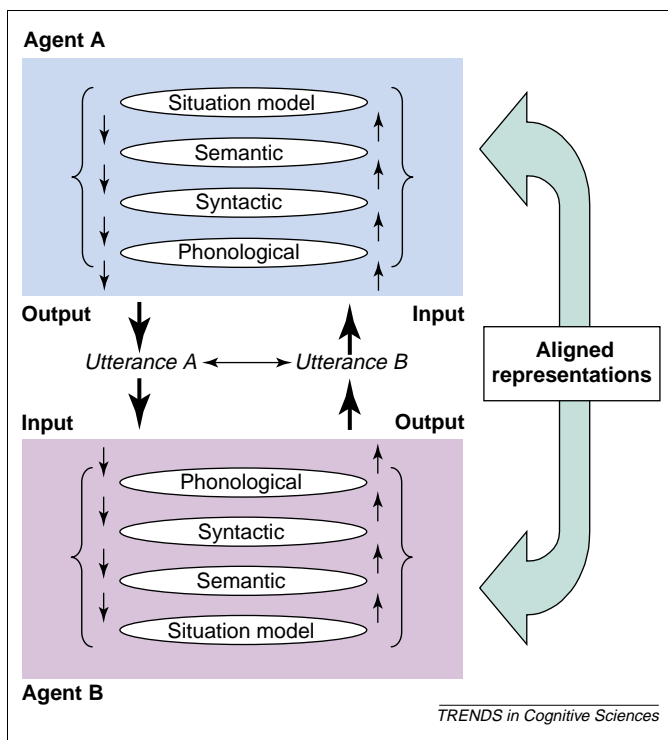
Dialogue transcripts are full of lexical repetition [29], and there are many experimental demonstrations of lexical alignment [22,30]. Interlocutors start to refer to particular objects using the same referring expressions (which gradually become shorter), but they tend to be modified if the interlocutor changes [24]. Syntactic alignment also occurs in dialogue, with speakers repeating the syntactic structure used by their interlocutors for cards describing events [11] (e.g. '*the diver giving the cake to the cricketer*') or objects [31], and repeating syntax or closed-class lexical items in question-answering [32]. They even repeat syntax between languages, when one interlocutor speaks English and the other speaks Spanish [33]. There is evidence for alignment of clarity of articulation [34], and of accent and speech rate [35]. Finally, alignment at one level increases alignment at other levels, with people being more likely to use an unusual form like '*the sheep that is red*' (rather than *the red sheep*) after they have just heard '*the goat that is red*' than after they heard '*the door that is red*' [31]. This is because *sheep* is semantically related to *goat* but not *door*.

---

words, when Nicola says something to Harry, the utterance activates linguistic representations in Harry. Because the same representations are used in producing and understanding, Harry then has those same representations activated when he comes to speak, and he will therefore tend to use them. So Nicola's productions influence Harry's productions and their internal representations become aligned. Crucially, alignment applies at all linguistic levels up to and including that of the situation model (see Figure 1).

The value of interactive alignment

How does interactive alignment help overcome the problems of dialogue? First, consider processing elliptical and fragmentary utterances. Interactive alignment ensures that interlocutors operate on common representations. So in speaking, each partner generates his utterance on the basis of what he has just heard from the other and can leave out redundant information without the risk of misunderstanding. Similarly in listening, aligned representations at the levels of the situation model, semantic interpretation, and syntactic form enable the listener to fill in the gaps at these levels.

Now consider opportunistic planning. Because conversation is a joint activity, much of the high-level planning (e.g. formulating speaker intentions) is distributed between interlocutors (see Box 1). For example, in producing a question, the speaker has already specified the high level goal for his addressee's next utterance, namely to answer that question. We also know that the form of the question constrains the form of the answer. For instance, '*Being called "your highness"*' is a well-formed reply to the question '*What does Tricia enjoy most?*', whereas '*That she be called "your highness"*' is not [12,13]. This is because you cannot say '*Tricia enjoys that she be called*

they use a largely unconscious process of 'interactive alignment' [7]. This is a process by which people align their representations at different linguistic levels at the same time. They do this by making use of each others' choices of words, sounds, grammatical forms, and meanings. Additionally, alignment at one level leads to more alignment at other levels. Hence, 'low-level' alignment (e.g. of words or syntax) leads to alignment at the critical level of the situation model (see Box 2). Conversations succeed, not because of complex reasoning, but rather because of alignment at seemingly disparate linguistic levels.

Interactive alignment comes about for two related reasons: (i) Parity of representations used in production and comprehension (i.e. when speaking and listening) [8–10]; and, (ii) Priming of representations between speakers and listeners [11]. Parity of primed representations leads to imitation, and imitation leads to alignment of those representations between interlocutors. In other

Figure 1. This diagram illustrates how parity of output and input leads to the alignment of internal representations between two agents (A and B). The horizontal arrow between utterances indicates the evidence we have for alignment of output and input. The small vertical arrows represent internal flow of information; the large vertical arrows represent the flow of information between the interlocutors. This scheme incorporates the channels of alignment at different linguistic levels.

"your highness"'. As interactive alignment predicts, speakers reuse the structures that they have just interpreted as listeners when formulating their response. This means that the low level planning of utterances is also distributed between interlocutors, thereby avoiding the problem of opportunistic planning.

What about the problem of making your utterances appropriate for your addressee? As a conversation proceeds, interactive alignment predicts that interlocutors build up a body of aligned representations, which we call the 'implicit common ground'. When this is sufficiently extensive, interlocutors do not have to infer each others' state of mind. What this means, crucially, is that people routinely have no need to construct separate representations for themselves and for their interlocutors, or to reason with such representations.

Finally, there is the problem of constant task switching between listening and speaking. With interactive alignment, production and comprehension become interdependent because they extensively draw on the implicit common ground. Hence, the interlocutors tend to use many of the same computations in producing their utterances, which therefore tend to be similar at many different linguistic levels at the same time (see Box 2). As the conversation proceeds, it will become increasingly common to use exactly the same set of computations. We call this process 'routinization'. So utterances involve an increasing proportion of expressions whose form and interpretation is partly or completely frozen for the purposes of the conversation [7], as is well illustrated by the expression *'right indicator'* in the example conversation given in Box 1.

Box 3. The perception–behaviour expressway

Dijksterhuis and Bargh argue that many social behaviours are automatically triggered by perception of action in others [16]. Such automatic perception–action links are well documented in the neurophysiological literature (e.g. motor imitation arising from the firing of mirror neurons in monkey premotor cortex [36,37]) and in the psychological literature [38]. There is evidence for automatic links in controlling facial expressions, movements and gestures, and speech. For example, when observing another person experiencing a painful injury and wincing, observers imitate the wince in their own expression [39]. Similarly, participants will mimic postures such as foot shaking and nose rubbing carried out by a person with whom they are conversing [40], and when they repeat another's speech they adopt the other's tone of voice as well [41]. Finally, it has recently been demonstrated that conversational partners dynamically align their posture [42]. We argue that the automatic alignment channels linking different levels of linguistic representation operate in essentially the same fashion.

Such routinized expressions are similar to stock phrases and idioms [14], except that they only 'live' for the particular interaction. Routinization greatly simplifies the production process [15] and gets around problems of ambiguity resolution in comprehension.

Automatic alignment channels and the perception–behaviour expressway

Although we have discussed interactive alignment in the context of language processing, similar alignment mechanisms appear to be present for other social activities. Dijksterhuis and Bargh argue that the majority of routine social behaviour reflects the operation of what they call a perception–behaviour expressway (see Box 3) [16]. Their basic argument is that we are 'wired' in such a way that there are direct links between perception and action across a wide range of social situations. Such links lead to imitation and imitation has the effect of aligning social representations between pairs of interacting individuals. For instance, it only takes one person to yawn in company and everyone else starts to yawn as well [17]. Not only do the others yawn, but they also come to feel more tired or bored [18]. We argue that interactive alignment operates through similar automatic alignment channels (not via reasoning) [7]. To be more explicit, we propose that the automaticity of alignment is post-conscious in Bargh's terms [19]. This means that interlocutors have to attend to what the other is saying in order for the automatic alignment to occur. We argue that alignment is also conditional to the extent that it can be inhibited when it conflicts with current goals and purposes, or promoted when it supports those goals. Behavioural mimicry is conditional in this way (e.g. people mimic other's incidental movements or gestures more when they intend to establish a rapport with the other person [20]). In the same way that the perception–behaviour expressway facilitates social interaction, automatic alignment channels facilitate language processing during conversation. (See Box 4 for other questions surrounding automatic alignment.)

Conclusion

So why is conversation easy? Our answer is that the interactive nature of dialogue supports interactive alignment of

linguistic representations. In turn the alignment of representations has the effect of distributing the processing load between the interlocutors because each reuses information computed by the other. Alignment comes about through automatic alignment channels similar to those in Dijksterhuis and Bargh's perception–behaviour expressway, which suggests that humans are 'designed' for dialogue rather than monologue. This to be expected because it is through dialogue that humans learn to speak in the first place.

References

1 Sellen, A.J. (1995) Remote conversations: the effect of mediating talk with technology. *Hum. Comput. Interact.* 10, 401–444
2 Allport, D.A. *et al.* (1972) On the division of attention: a disproof of the single channel hypothesis. *Q. J. Exp. Psychol.* 24, 225–235
3 Clark, H.H. (1996) Using Language, Cambridge University Press
4 Johnson-Laird, P.N. (1983) Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness, Harvard University Press
5 Sanford, A.J. and Garrod, S.C. (1981) Understanding Written Language, John Wiley & Sons
6 Zwaan, R.A. and Radvansky, G.A. (1998) Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185
7 Pickering, M.J. and Garrod, S. Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* (in press)
8 Prinz, W. (1990) A common coding approach to perception and action. In *Relationships Between Perception and Action: Current Approaches* (Neumann, O. and Prinz, W., eds), pp. 167–201, Springer
9 Fowler, C.A. *et al.* (2003) Rapid access to speech gestures in perception: evidence from choice and simple response time tasks. *J. Mem. Lang.* 49, 396–413
10 Liberman, A.M. and Whalen, D.H. (2000) On the relation of speech to language. *Trends Cogn. Sci.* 4, 187–196
11 Branigan, H.P. *et al.* (2000) Syntactic coordination in dialogue. *Cognition* 75, B13–B25
12 Ginzburg, J. and Sag, I.A. (2001) Interrogative Investigations, CSLI
13 Morgan, J.L. (1973) Sentence fragments and the notion 'sentence'. In *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane* (Kachru, B.B. *et al.*, eds), pp. 719–752, University of Illinois Press
14 Jackendoff, R. (2002) Foundations of Language, Oxford University Press
15 Kuiper, K. (1996) Smooth Talkers: The Linguistic Performance of Auctioneers and Sportscasters, Erlbaum
16 Dijksterhuis, A. and Bargh, J.A. (2001) The perception–behavior expressway: automatic effects of social perception on social behavior. In *Advances in Experimental Social Psychology* (Vol. 33) (Zanna, M.P., ed.), pp. 1–40, Academic Press
17 Provine, R.R. (1986) Yawning as a stereotypical action pattern and releasing stimulus. *Ethology* 71, 109–122
18 Hatfield, E. *et al.* (1994) Emotional Contagion, Cambridge University Press
19 Bargh, J.A. (1994) The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition. In *Handbook of Social Cognition* (Vol. 1) (Wyer, R.S. and Srull, T.K., eds), pp. 1–40, Erlbaum
20 Lakin, J.L. and Chartrand, T.L. (2003) Using non-conscious behavioral mimicry to create affiliation and rapport. *Psychol. Sci.* 14, 334–339
21 Schegloff, E.A. and Sacks, H. (1973) Opening up closings. *Semiotica* 8, 289–327
22 Garrod, S. and Anderson, A. (1987) Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition* 27, 181–218
23 Sacks, H. *et al.* (1974) A simplest systematics for the organization of turn-taking in dialogue. *Language* 50, 696–735
24 Brennan, S.E. and Clark, H.H. (1996) Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1482–1493
25 Linnell, P. (1998) Approaching Dialogue: Talk Interaction and Contexts in Dialogical Perspectives, John Benjamins
26 Schober, M.F. and Clark, H.H. (1989) Understanding by addressees and over-hearers. *Cogn. Psychol.* 21, 211–232
27 Schober, M.F. (1993) Spatial perspective-taking in conversation. *Cognition* 47, 1–24
28 Garrod, S. and Doherty, G. (1994) Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition* 53, 181–215
29 Tannen, D. (1989) Talking Voices: Repetition, Dialogue and Imagery in Conversational Discourse, Cambridge University Press
30 Garrod, S. and Clark, A. (1993) The development of dialogue co-ordination skills in schoolchildren. *Lang. Cogn. Process.* 8, 101–126
31 Cleland, S. and Pickering, M. (2003) The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *J. Mem. Lang.* 49, 214–230
32 Levelt, W.J.M. and Kelter, S. (1982) Surface form and memory in question answering. *Cogn. Psychol.* 14, 78–106
33 Hartsuiker, R.J. *et al.* Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish/English bilinguals. *Psychol. Sci.* (in press)
34 Bard, E.G. *et al.* (2000) Controlling the intelligibility of referring expressions in dialogue. *J. Mem. Lang.* 42, 1–22
35 Giles, H. *et al.* (1992) Accomodation theory: communication, context and consequences. In *Contexts of Accomodation* (Giles, H. *et al.*, eds), pp. 1–68, Cambridge University Press
36 Rizzolatti, G. and Arbib, M.A. (1998) Language within our grasp. *Trends Neurosci.* 21, 188–194
37 Di Pellegrino, G. *et al.* (1992) Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180
38 Hommel, B. *et al.* (2001) The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878
39 Bavelas, J.B. *et al.* (1986) I show how you feel: motor mimicry as a communicative act. *J. Pers. Soc. Psychol.* 50, 322–329
40 Chartrand, T.L. and Bargh, J.A. (1999) The chameleon effect: the perception–behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910
41 Neumann, R. and Strack, F. (2000) Mood contagion: the automatic transfer of mood between persons. *J. Pers. Soc. Psychol.* 79, 211–223
42 Shockley, K. *et al.* (2003) Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 326–332