1 The Challenge of Dialogue

Let's start with an extract from a dialogue in which two players work together using a computerized maze game (Garrod & Anderson, 1987). In this extract, one player *A* is trying to describe her position to her partner *B* who is viewing the same maze on a computer screen in another room.[1]

Example 1.1

1———**B:** . . . Tell me where you are?
2———**A**: Ehm : Oh God (*laughs*)
3———**B:** (*laughs*)
4———**A:** Right : two along from the bottom one up:
5———**B:** Two along from the bottom, which side?
6———**A:** The left : going from left to right in the second box.
7———**B:** You're in the second box.
8———**A:** One up :(*1 sec.*) I take it we've got identical mazes?
9———**B**: Yeah well : right, starting from the left, you're one along:
10———**A:** Uh-huh:
11———**B:** and one up?
12———**A:** Yeah, and I'm trying to get to . . .

Up to (7), it is clear that *A* and *B* have different understandings of *A*'s location. But after *A*'s concerned question in (8), they start to appreciate and address their confusion. And by (12), they appear to have the same understanding of *A*'s location and to use the same language to describe that location. We say that they have started to become *aligned*, both with respect to the understanding of the situation and with respect to their use of language to describe that situation.

Alignment is the extent to which individuals represent things in the same way as each other. We argue that alignment is fundamental to communicative

[1] Here, *:* refers to a brief pause, *1 sec.* to a longer pause. Overlapping speech is not annotated. Throughout the book, *A* is treated as female and *B* as male.

https://doi.org/10.1017/9781108610728.003 Published online by Cambridge University Press

success. For an interaction to be successful, the interlocutors represent relevant aspects of the world (such as the maze position) in the same way as each other. To do this, we argue that they also represent relevant aspects of the language (such as the meaning of *one along*) in the same way. They also seek to be as efficient as possible – not using unnecessary words, not leaving long pauses and not having to repair mistake after mistake.

In everyday dialogue, interlocutors constantly make rapid, short contributions, so that they share the workload. In our example, both interlocutors have to process language extremely rapidly and efficiently, a situation which is quite typical in dialogue. Without appreciable delay, they interpret each other's utterances, complete them, repeat their choices of expressions, and realize when it is their turn to speak and when to stay silent. Many of these decisions clearly take no more than a few hundred milliseconds. It is the aim of this book to explain how people can routinely engage in successful dialogue using mechanisms of the sort that are basic to human cognition.

Now imagine the maze game dialogue as though one interlocutor were contributing without the other. If so, *A* would say:

2———**A**: Ehm : Oh God (*laughs*)
4———**A:** Right : two along from the bottom one up:
6———**A:** The left : going from left to right in the second box.
8———**A:** One up :(*1 sec.*) I take it we've got identical mazes?
10———**A:** Uh-huh:
12———**A:** Yeah, and I'm trying to get to . . .

This of course makes no sense. An overhearer could not fully determine what *A* is trying to convey, and a researcher could not explicate the processes that underlie *A*'s utterances. Why does *A* describe a position in the maze (4 or 6)? What is the relevance of *One up* (8) or *Uh-huh* (10)? Why does *A* produce the extended query in (8)? Why does *A* produce a disfluency in (2) and a pause in (8) but no pauses elsewhere?

B's contributions would be equally unintelligible and uninterpretable on their own.

1———**B:** . . . Tell me where you are?
3———**B:** (*laughs*)
5———**B:** Two along from the bottom, which side?
7———**B:** You're in the second box.
9———**B**: Yeah well : right, starting from the left, you're one along:
11———**B:** and one up?

Why does *B* use the rather odd construction *Two along from the bottom, which side?* or the words *two* and *along* (5)? Similarly we cannot interpret why *B* laughs (3) or, perhaps more interestingly, why *B* splits a complete utterance over two contributions (9 and 11).

In this format, we cannot appeal to the fact that *A* has used these expressions immediately before. In order to understand what the interlocutors are saying, we have to consider what they both say, so that we can determine the interdependencies between the contributions. This means that we cannot adequately understand the underlying cognitive mechanisms – how they select words and grammatical constructions, plan contributions and speak in a reasonably fluent and intelligible manner – without analysing both interlocutors at the same time.

1.1 Individuals in the Dialogue System

Dialogue is a form of joint activity. In the most general sense, a joint activity is simply an activity in which two (or more) individuals contribute. But dialogue involves a much more precise notion. The individuals have to commit to taking part in the dialogue, and they have to interact to work towards a common goal with their contributions affecting their partner's contributions. The individuals are therefore tightly linked as part of a system. To understand dialogue, we need to understand both the structure of the system and the roles of the individuals-within-the-system.

Specifically, we interpret dialogue as a *cooperative joint activity*, in which the individuals are committed to the activity, respond to the intentions and actions of each other, and support the activity to ensure its success (notably, by compensating for a partner's mistakes in order to keep the activity on track). This notion is derived from Bratman's (1992) notion of shared cooperative activity, which occurs when duettists sing, workmen build something together or footballers perform a move together. But we extend it to include joint activities with a competitive component (e.g. chess players cooperate on following the rules but compete to win). Hence we can include rational arguments as well as friendly conversations, because the interlocutors are still committed to promoting the success of the dialogue. They do so by exerting distributed control over the progress of the dialogue, for example by providing feedback to their partners when necessary. Underlying the interlocutors' behaviour is what Bratman (1992) calls a joint intention – an intention to perform the activity as a whole, even though each individual does not intend to perform it alone.

Cooperative joint activities are a major challenge for cognitive science. Almost always, it analyses thought and behaviour in the individual – how individuals perceive, remember or reason, or indeed how they use language. Even when cognitive science concerns itself with social cognition, the focus is on how one individual represents and processes socially relevant information. In other words, the constructs are defined with respect to an individual – what happens in one person's mind. And a theory cannot make claims about the relationship between two minds – for example, whether those minds are

aligned. We refer to this dominant approach as *monadic* cognitive science. Our book goes beyond monadic cognitive science and treats the dyad rather than the monad as the basic unit of analysis.

1.2 The Shared Workspace Framework

Imagine two individuals preparing to construct a piece of flat-packed furniture. They take the relevant pieces from the boxes and lay them out on a rug between them. For example, they might separate different types of screws into piles, keeping them apart from the hinges and the handles, and then lay out the parts of the drawer – the component they have to construct first. At this point, they plan to join up the front with one side of the drawer (with one individual holding the pieces of wood while the other screws them together), and they have to select the right screws to do so. They therefore both focus on some of the pieces in front of them – those pieces that are in their upcoming action. They then perform this cooperative joint action by manipulating just these pieces, and they do so in what we call a *shared workspace*.

The shared workspace is highly restricted in scope. It contains what is relevant for the here-and-now – in other words, the pieces needed for the current action (e.g. the wood and screws), not for potential future actions (e.g. not the back of the drawer). And it is the physical basis of alignment – that is, it contains the objects that the individuals represent. These contents are typically manifest to both individuals – that is, they both attend to its contents, and both assume that their partner does so too. However, these contents do not have to be manifest to both individuals (or indeed to either). For example, the individual holding the wood might believe that the individual holding the screws is aware of the wood (i.e. the wood is manifest), but the individual holding the screws might not believe that the individual holding the wood is aware of the screws (i.e. the screws are not manifest).

But it contains more than just objects. It also contains events, typically the individuals' behaviours that contribute to the cooperative joint activity. Moreover, the individuals communicate about construction, for example discussing who will hold the pieces of wood, or indicating which screw to use. Such communicative acts change the contents of the shared workspace for two reasons. They lead to some objects (or perhaps events) entering the workspace at the expense of others. But in addition, the workspace now contains the communicative acts themselves – whether it is the utterance *Use the largest screws* or a pointing gesture towards those screws. These acts are likely to be manifest to both individuals, as they are likely to assume that they both attend to the words in a communicative context. Each individual can add objects and events, communicative acts, or their combination to the workspace, and can manipulate them or combine them with their partner's contributions.

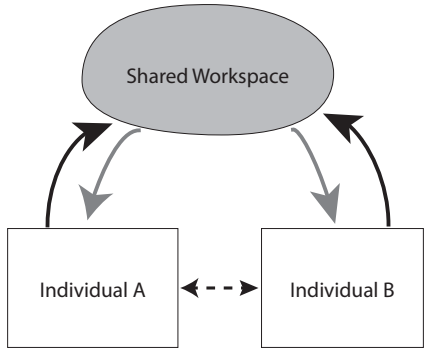

Figure 1.1 The shared workspace framework for cooperative joint activities.

Our framework involves a system in which two individuals interact via the shared workspace (see Figure 1.1). They can both manipulate the contents of the shared workspace (via the upward black arrows) according to need, either one after the other or both at the same time. And they can monitor or predict any aspects of the workspace (via the downward grey arrows) in order to assess the progress of the joint activity and determine its likelihood of success. These arrows therefore capture information flowing in both directions between the workspace and each individual. The individuals therefore link up to the shared workspace as a system. We can (monadically) analyse the individuals-within-the-system, for example considering the nature of their private cognitive representations that they use in interaction (e.g. the linguistic representations underlying an utterance). But we can also (non-monadically) analyse the system as a whole, for example considering the relationships between each individual's representations and the relationships between each individual's behaviour.

1.3 Alignment via Interaction

We assume that alignment underlies many successful cooperative joint activities, and that it corresponds to a well-functioning shared workspace. But alignment is a relationship between the individuals' cognitive representations, and does not relate to the workspace directly. In Figure 1.1, alignment is illustrated using the dashed horizontal arrow between the individuals. Unlike the other arrows, it does not illustrate information flow (as the individuals' representations are private). The information goes via the shared workspace, but its effects are (often) to promote alignment. The investigation of alignment (e.g. determining what leads to individuals having similar representations) is not monadic, because alignment can be understood only by comparing different individuals.

https://doi.org/10.1017/9781108610728.003 Published online by Cambridge University Press

Our focus is on alignment as a consequence of interaction. Interlocutors communicate with each other and as a result tend to develop many similar representations, both with respect to language and to their broader understanding of the topic under discussion. For example, if *A* calls someone a chef, then *B* is likely to use the word *chef* as well, perhaps pronounce it similarly and refer to the same person. If *A* then uses an unusual grammatical construction (e.g. *the chef that's smiling*), then (as we shall see) *B* often uses that construction in a subsequent utterance (e.g. *the waiter that's surly* instead of the more common *the surly waiter*). Interlocutors are typically unaware of such alignment.

We argue that alignment is a consequence of parity (equivalence) between representations used in production and comprehension. When the upward arrows reflect language they correspond to production, and when the downward arrows reflect language they correspond to comprehension. When information flows up from one interlocutor to the shared workspace and then down to the other interlocutor, it leads to alignment.

Alignment does not merely relate to language and its interpretation but rather to models of the situation under discussion. Our furniture builders have models of the physical basis of the action – in this case, the drawer under construction. In successful cooperative joint activity, their models are aligned with each other, and also correspond closely to the contents of the shared workspace. We propose that they continue constructing, and continue discussing their construction, and as a result align on a more extensive situation model, which incorporates representations corresponding to the current shared workspace and aspects of the history of the interaction.

Each interlocutor also has their own representations for planning and carrying out the dialogue, but these are joint representations in the sense that they take into account the part played by the interlocutor. So when *A* answers *The left* to *B*'s question *Two along from the bottom, which side?*, *A* represents something like 'Two along from the bottom on the left'. The representation takes into account both *B*'s and *A*'s contributions. It is a (perhaps imperfect) model of the situation under discussion – and *A* and *B* are aligned in this respect if they have the same model as each other. In addition, their representations (such as their models of the maze) reflect the contents of the shared workspace (such as the configuration of the maze).

The combination of a need to develop extensive situation models and a limited capacity shared workspace means that interlocutors attempt to align efficiently. For example, they try to align on succinct referring expressions with fixed interpretations – which we call *routines*. More generally, they attempt to say just enough to communicate efficiently. In addition, they attempt to time their contributions to the shared workspace optimally, so that they can access and act on those contributions when they need to – without

exceeding the capacity of the workspace. But it is also possible to enhance the shared workspace by highlighting aspects of it to the interlocutors, for example by having a physical whiteboard that represents some aspects of shared workspace or by technologies such as social media.

1.4 Organization of the Book

The book is organized into four main parts. Part I presents the shared workspace framework. We introduce it in relation to cooperative joint activities in general (Chapters 2 and 3) and establish the general properties of the framework, which we then interpret in relation to dialogue (Chapters 4 and 5). We describe the dyadic system itself and the way that the joint actors control the system.

Part II contains two chapters that are concerned with the interlocutors. Chapter 6 presents the theory of interactive alignment and primarily focuses on the alignment of linguistic representations underlying the utterances that each speaker contributes to the dialogue. Such representations enable interlocutors to formulate phrases, words or gestures that move the dialogue forward. Chapter 7 is concerned with the alignment of non-linguistic representations which we call dialogue models. These models capture two important aspects of a dialogue: the situation model that represents what is being discussed and the dialogue game model that represents the conversational moves by the interlocutors (e.g. information-probing).

Part III explains how interlocutors use the shared workspace to achieve alignment of cognitive representations. The shared workspace offers a limited window on that part of the world that joint actors can manipulate and observe. Because it is severely limited in scope, they need to use it as efficiently as possible. Chapter 8 describes how interlocutors put just enough information into the shared workspace to achieve alignment and distribute the information between them as efficiently as possible. Chapter 9 discusses how interlocutors time their contributions to fit in with each other's contributions fluently. By saying just enough and in good time, interlocutors optimize the interactive alignment process.

Part IV extends the shared workspace framework beyond the 'minimal dyad'. Chapter 10 considers multi-party conversation in relation to conversational roles (such as game players, audiences and collectives), with respect to both alignment and control. It then applies the framework to restricted dialogue and, importantly, to monologue. Chapter 11 relates the framework to culture, and considers both cultural institutions (such as conventional activities) and cultural artefacts – that is, technologies that can enhance (and change) communication such as illustrations, writing and electronic devices.