The University Of Sheffield.

Two Types of Conversation: Face-to-Face and Digital

by Patrick Connolly

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of the Arts and Humanities
Department of Philosophy

November 2020

List of contents

Acknowledgements

A special thanks is owed to Jenny Saul. This project quite literally owes its existence to Jenny's guidance. The dissertation I had originally started work on didn't address the topic of digital conversation but after reading a draft chapter I'd written for that project Jenny suggested that one of the footnotes might be worth developing into a short 4,000 word *Analysis* paper. The footnote was on the contrast between face-to-face and digital conversation and this, then, is my first draft of that *Analysis* paper - a mere 66,000 words longer than planned. This isn't really an isolated anecdote either. Jenny's interest, eye for what is interesting and the insightful comments and suggestions she has given me throughout this project have sustained and made it immeasurably better. It has been a pleasure and an honour to work with Jenny.

I would also like to *give kudos* to Dominic Gregory for his invaluable help and guidance throughout this project.

I am very grateful too to Emma Borg for the time we spent working together. During this period of time not only was I fortunate enough to have conversations with a philosopher I already admired, but Emma's patience and advice have been instrumental in my development as both a writer and a philosopher

I am very grateful, too, to the White Rose College of the Arts and Humanities for generously funding my research. This project certainly wouldn't have been possible without it.

Writing this post-viva, I would also like to thank my examiners Eliot Michaelson and Rosanna Keefe for the incredibly interesting questions and discussion in the examination. Being as it was during the Covid pandemic our conversation was, perhaps fittingly, at times broken up by digital technology, but the discussion was incredibly helpful.

There are a great number of people I would like to thank for comments, discussions and feedback on material that has ended up in here but I've decided not to list them for fear of missing someone out. I am indebted particularly to Will Hornett and Sean McIntosh though for discussions early into this project that helped me shape my thinking on how to approach it.

A very special thanks is also in order to the people who have supported me through this project with love and patience – in particular Anna and Rose. I particularly asked a lot of them in the weeks leading up to completing this and they did it willingly and supportively. Which is how they have been throughout. Without their continued support this simply wouldn't have been possible. And thanks to Agatha, Maeve and Martha – without you three I could have finished this *much* quicker.

I also would like to thank all the people I have had the privilege of attending the Sheffield Philosophy of Language Reading Group with. It has been a complete pleasure to be part of the discussions with the group and it is something I shall miss.

The satisfaction I feel now having completed a dissertation is tempered slightly by this also marking the end of an enjoyable period of my life in the philosophy department at Sheffield. This is made especially sad by the thought that the circumstances of 2020 mean I won't be able to wander around the labyrinth of Victoria Street to personally say goodbye to the many wonderful people I've met there.

Many thanks are due to Eric Olson, Steve Makin and Paul Faulkner for advice and guidance over the course of my time in Sheffield.

I'd like to thank Komarine Romdenh-Romluc and Niall Connolly too for your help and guidance, particularly during my first year as a postgraduate.

I'd finally like to thank Anne-Marie Frisby, Katie Owen, Joanne Renshaw, Siri Romare and Sally Weston in the departmental office for their patience and help with my no doubt very silly questions over the years.

Abstract

This thesis is a comparison of face-to-face and written digital conversation. I start with the intuition that face-to-face conversation can often appear more engrossing and satisfying than its digital counterpart. I argue that one of the most promising ways of understanding this difference can be seen when we consider the contrasting coordinative structures of these two types of conversation. In face-to-face conversation the task of communication is at all times spread between participants whereas in digital conversation the burden of communication is passed almost entirely from one to the other. One notable result of this is that it gives us good reason to think that communication in digital conversation is in many ways more difficult. I then argue that the difference in coordination in digital conversation has consequences for the nature of the cooperation we find in such interactions. I argue that these consequences of the different structures of face-to-face and digital conversation are what best explains the starting intuition.

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Some of the work presented in Chapter 5 has been conditionally accepted for *The Journal of Social Philosophy*.

Table of Contents

Introduction

In this thesis I explore the intuition that despite being in many ways similar to the conversations we have face-to-face, the way that we experience digitally mediated written conversations is somehow different. When I think of the great conversations I've had I can recall very few I've had online that that have absorbed me in the way face-to-face conversation often can. It seems to be the case that face-to-face conversation has the potential to be more engrossing and often more satisfying, somehow, than its digital counterpart. In some senses this might seem to have obvious explanations. We might think that looking at a screen rather than another person provides an obvious experiential difference in itself. Perhaps we might think the lack of simultaneity in digital conversations creates with it a different type of experience. Another thought might be that the environments digital conversations take place in are often less conducive to personable or intimate interaction; as mentioned above we have a screen rather than a person in front of us, there is often more anonymity online and many of the spaces we have these conversations in are much more public with utterances being potentially read by others. I think it is likely these types of difference are all contributory in some ways to how we experience these two types of conversation differently. What I argue here, though, is that when we focus specifically on the structural differences and the distinctive sets of requirements placed on interlocutors in the two types of conversation, we see there are some much richer contrasts to be drawn that relate to the nature of the communication that takes place in these two types of conversation. For when we look at what is required of us in these two types of conversation not only are interlocutors in a digital conversation geographically detached but they are communicatively detached too.

I start in Chapter 1 by providing an outline of what we might call a paradigm type of conversation. The cluster notion I propose entails that a paradigm type of conversation is a turn-taking exchange in which interlocutors each have some control over the direction of

the interaction, contributions made to the discussion are interdependent, and participants are engaged in processes of continued perspective sharing and prediction. In Chapter 2 I draw on work from psycholinguistics and psychology to fill in some of the details as to what a face-to-face version of a paradigm type of conversation looks like. Following Herbert Clark in distinguishing between process-level coordination and content-level coordination I suggest that in order to coordinate in the way required to sustain such interactions interlocutors must engage in a particular type of process coordination. This process coordination also has consequences for the type of cooperation required of the participants. Considering this in terms of cooperation can help us to understand the relationship between coordination at the process level and coordination at the content level that occurs in face-to-face conversation. In Chapter 3 I start the comparison of face-to-face and digital conversation. This chapter looks specifically at spoken and written language and the focus is on some of the differences in acquisition, their respective historical roles and the available communication channels. I argue that although these differences may be in part contributory, they each are, to different degrees, surmountable. And considered alone these surface-level differences don't seem to tell us much about why face-to-face and digital conversations differ. In Chapter 4 I argue the most striking difference between these two types of conversation is in the diverse nature of requirements placed on interlocutors in each. Recalling discussion of Chapter 2 I argue that whereas in face-to-face conversation the task of communication is at all times spread between participants, in digital conversation the burden of communication is passed in almost its entirety from one participant to the other. This, we should expect, makes communication in digital conversation a more difficult task. Not only this, though, the difference in coordination in digital conversation has consequences for the nature of the cooperation we find in these interactions. It is these two consequences - the difficulty of communication and the nature of the cooperation – that I take to be most relevant to considerations of why face-to-face conversation can often appear more engrossing and satisfying than its digital counterpart. In the final chapter I present a case study on the speech act of trolling. My purpose in this chapter is to suggest what I think to be the direction of research this type of study points us towards. There I examine in close detail a new type of linguistic behaviour that is made possible by the different nature of digital conversation.

The first chapter is, in part, an attempt to characterise a certain type of interaction which I call a paradigm type of conversation. There are, we might think, many different types of ways we can have face-to-face conversation. Roughly put, on an account such as

Mitchell Green's (1999, 2017) any type of interaction that has some pertinent line of inquiry can be considered to be a conversation. For my purposes of comparison of face-to-face and digital conversation, however, what I want to do is reduce the variables as much as possible to allow a closer comparison of like-for-like. So my primary motivation in this chapter is to suggest a narrow category of interaction that can be used later in the dissertation as a comparison class. Conversation as an object of study is motley and difficult to pin-down so what I propose here is a cluster notion of conversation in which some interaction might have more or less of some of the properties suggested. What I propose the paradigm type of conversation looks like is that it is a turn-taking exchange (which rules out from the paradigm case examples such as academic lectures, soliloquy or novels which Green accepts as types of conversation). This turn-taking exchange, I suggest, should also be much more *synocratic* in nature than it is *autocratic*. That is, the conversational participants must each have some degree of control over the direction of the conversation (in turn, this rules out from the paradigm some turn-taking activities such as cross-examinations or to a lesser extent job interviews). In Section 3 I then look at features that characterise more the experience of having a paradigm type of conversation. These are features such as the interdependence of contributions, the role of continued perspective sharing and of prediction. Having roughly characterised the type of conversation I wish to consider for the remainder of the dissertation, in the next chapter I begin the work of looking in more detail at some of the features of a face-to-face paradigm type of conversation.

In the second chapter I sharpen this notion of a paradigm conversation by exploring some of the structural features of such an activity. Here we consider the implications of some of the features of face-to-to-face conversation. Spoken and signed languages are generally instantaneous and interlocutors can see and/or hear each other. The medium used is generally evanescent, simultaneous and unrecorded. And the control of the discussion rests with the interlocutors who have self-expression and self-determination. Drawing upon recent work in psycholinguistics that looks at timings in face-to-face conversation we begin to see some of the consequences of such features. For it is the case that in order to sustain such interactions participants are required to perform a rich set of communicative tasks that involve providing feedback, initiating repair sequences and predicting speech act types and turn-duration. Drawing upon Herbert Clark's distinction between content and process coordination I argue that when we focus on the process coordination tasks involved in face-to-face conversation we begin to see a consistency of structure across different conversations. So although the content of two conversations

might differ vividly, the types of tasks required of the conversational participants in order for them to sustain their interactions we should expect stay roughly the same. I conclude this chapter by arguing that we can best understand how this coordination works by thinking about the type of cooperation required in order to meet these coordinative requirements. Using Paul Grice's cooperative principle I argue that we can understand the close relationship between the process and content levels of conversational coordination by considering the preconditions that must be met in order to meet something like Grice's cooperative principle. The result is that in order to sustain a face-to-face conversation conversational participants are continuously engaged in a form of cooperative behaviour.

The third chapter starts the comparison of face-to-face and digital conversation. One of the most obvious points of contrast between face-to-face and written digital conversations is modal. Face-to-face conversation is usually conducted using spoken or signed languages, whereas the type of digital conversation I am interested in here is primarily written. In this chapter then I contrast spoken and written language to examine whether there is something specific to these modes which might help explain the differences in the two types of conversation. I look at differences in acquisition, in the historical roles of spoken and written language and at some of the different channels of communication available in the two modes. I argue that although such differences may be contributory as to why as to why face-to-face and digital conversations differ, these such differences alone don't seem to be suitably explanatory. And indeed, we also have good reason to believe that to varying degrees each of these three differences are already in the process of being reconciled in the two modes.

In the fourth chapter I address what I take to be the most interesting difference between face-to-face and digital conversation. Here I argue that when we pay attention to the differences in the types of process tasks first discussed in Chapter 2 we see that the requirements placed on interlocutors in face-to-face and digital conversations are interestingly distinct. We see similarity in some features such as the self-determination and self-expression available to interlocutors. This I argue gives us good reason to consider face-to-face and digital conversations to be suitably similar for comparison. However, when we consider some of the different features of digital conversation we see that whereas the modes used in face-to-face conversation are generally evanescent and recordless and the interactions instantaneous, in digital conversation utterances are recorded and permanent and interactions take place at most quasi-instantaneously. A significant consequence of this

is that in digital conversation the need to carry out continuous communicative tasks is not as pressing as it is in face-to-face conversation. So whereas in face-to-face conversation the overall job of communication is continuously shared amongst interlocutors, in digital conversation the burden of communication is, roughly speaking, passed from one to the other. I argue that this has interesting consequences for the type of cooperation we might think is required for sustaining digital conversation. I will suggest that it is this underlying difference which best explains why we might think face-to-face conversation can often seem more engrossing or satisfying.

The final chapter serves both as a case study and an example of the type of research that I think follows from paying attention to some of the differences in face-to-face and digital conversations. I do this by looking specifically at one particular phenomenon we find primarily in digital environments. 'Trolling' has become a term to denote a wide range of behaviour we find in internet communication, ranging from what appear to be harmless japes through to bullying, abuse and hate speech. In this chapter I will argue that by using tools from the philosophy of language and by considering trolling as a type of speech act, we can start to see some of the structural similarities between these seemingly disparate acts. Once these similarities become clearer, we can then understand better what trolling is and why it has become such a pervasive feature of digital conversation.

1

A Paradigm Type of Conversation

Despite being a foundational area for much of our theorising on language, the notion of 'conversation' rarely has a concise definition attached when it occurs in the literature – the sense is, perhaps, that we all have a strong intuitive sense of what is meant by 'conversation'. So maybe the thought is that not much more needs to be said about what conversation is. We might think there are obvious paradigm cases of conversation in which interlocutors speak informally about what they did the previous evening, catch up with some gossip, discuss their plans for the summer, or argue about the merits of favourite films. These all seem to be uncontroversial examples of conversation. Though we might also think not all uses of language are what we would call conversations; we might wonder whether a novel or a TV series, for instance, are a type of conversation, or perhaps news bulletins, or the UK Weights and Measures Act 1985. It may be thought that these types of language use lack the interactivity we would traditionally expect of conversation. It may be that a novel or a TV series contains dialogue that appears conversational, but whether the novel itself, or a TV series itself is a conversation seems less clear. *Prima facie*, then, it might be thought that as these latter types of language use lack obvious interactivity (for example, when reading a novel the roles remain static, the author does the writing, the reader does the reading and in a TV series the roles are similarly static - actors and screenwriters do the talking and the viewer does the listening), then we have a simple method of delineating between conversational and non-conversational language use along the lines of the interactivity. On closer inspection, as will be discussed later, it isn't quite that simple even with these types of case. Further to this, there are also types of language use that *are* interactive, yet possibly sit somewhere on the border of what might be thought of as conversations, for example, a job interview, a cross-examination in court or a text message exchange.

My suspicion is that one of the reasons we don't have a settled definition of conversation of the type loved by philosophers – that is of a string of necessary and sufficient conditions that can help us delineate between conversational and non-conversational language uses – is that conversation is, by its nature, simply too wild and sprawling, simply put; conversation can come in many different forms.[1] As such, the balance between being too restrictive and overgenerating by introducing strict definitional criteria makes the point of distinction perhaps too fine to find. Conversation also traverses many different supposed disciplinary frontiers. As an activity it is linguistic, psychological and social. Even within philosophy one could analyse it within philosophies of (joint) action, of language, of mind or through social philosophy. So perhaps the interesting aspects of what we might call conversation are relative to our interests. Even though a precise definition of conversation may be out of reach, this needn't entail we can't at least attempt to consider it

[1] It would be remiss not to note here Mitchell Green's conception of conversation (1999, 2017) as he does offer a neat way of determining what conversation is and he would include in his class of conversation such activities as reading a novel or watching a TV series. For example, he says that:

> "[W]hile what I've elsewhere (Green, 1999) called exchanges will involve a desultory sharing of information among interlocutors, conversations often have a teleological dimension: instead of just chatting, we frequently aim to answer a particular question or set of questions about what to do or what is the case." (2017, p.1593)
>
> "I want to suggest that conversations are characteristically directed toward some end or other. From this teleological perspective, conversations emerge as projects that might be spread over many years, continents, or journal issues." (2017, p.1594)
>
> "[So] let us say that whereas a verbal exchange is any sequence of speech acts, a conversation is a sequence of such acts ostensibly aimed at answering either a theoretical or practical question." (2017, p.1595)

Note, then, that for Green whenever some exchange has some teleological dimension (primarily conceived by Green as it having a question answering dimension) it then becomes conversational in nature. The boundaries of this notion of conversation, then, allows that a lecture, a novel, a poem, or a series of journal articles can be classed as 'conversation'. This notion of conversation is far too broad for my purposes here, though I have no principled objection to Green defining conversation in such a way. Green's goals are different to my own here; he wishes to expand the work that can be done using Stalnaker's CG-context model and so it is an expansive project. My aim, on the other hand, is restrictive in that I aspire towards an easier way to contrast face-to-face and online interactions. What I also hope will later become apparent is that the fundamental difference between Green's notion of conversation and the conversation I discuss here and in subsequent chapters relates to the interdependent notions of content and process (much more on this in Ch.2). Green is content-focused – it is the contents of some interaction that he takes to be salient to their nature. Whereas my paradigm type of conversation takes similarity of process-structure to be the fundamentally important notion. In terms of a project looking to contrast face-to-face and digital conversation, if we compare the content alone there need not be any interesting general difference. It is perfectly conceivable that the content of some digital conversation could have content identical to that of a face-to-face conversation, whereas when we concentrate on the processes involved we see some intrinsic contrasts. Green's account is deserving of more attention than I have space for here and I would certainly have liked to say more to contrast my own notion of conversation and Green's. However, I have taken the decision that to do so would add an an extra level of complexity that distracts from the overall aims. I will keep note of some of the contrasts with Green in footnotes, however.

as a distinct type of activity, and this is why I take the approach of attempting to characterise it as a cluster notion.

My overall project here is an attempt to understand some of the differences between traditional face-to-face interaction and the newer type of text-based digital interaction we find in online environments (from hereon in I will use the shorthand *digital conversation* to refer to these types of interaction).[2] And working from the intuition that although much is similar between face-to-face and digital conversation, there is something different in the way we experience them. I want to set up a way of testing this intuition by way of a comparative study. In order to make this comparison, (and in particular make it a manageable task) I plan to restrict the types of interaction that will be looked at. What I consider here will be a narrow class of interactions and for the sake of this discussion I will call these 'conversations'. As noted above, a precise definition of conversation is elusive and so what I present in this chapter is a cluster concept which I will call a paradigm type of face-to-face conversation.[3] The general idea being that the more features of the cluster some exchange possesses, the closer to the paradigm it is. The resulting cluster notion will be described in greater detail in Chapter 2, and the task of Chapters 3 and 4 will be to contrast this paradigm case of conversation with what appear to be correlating text-based digital conversations.

This cluster account of face-to-face conversation characterises them as linguistic exchanges that have properties such as *turn-taking* (§2.2), *synocratic control* (§2.3), *interdependence of contribution* (§3.1) and *shared perspectives* (§3.2). To enable a

[2] It's worth emphasising at this point that I am contrasting face-to-face conversation with very specifically *text-based* digital conversation even if the shorthand 'digital conversation' is imprecise. Many of us living through the Covid-19 pandemic will have experienced a great deal of digital conversation that wasn't text-based, for example through the use of video calling software such as Zoom. As I shall explain in Ch.3, I regard such conversations to be versions of face-to-face conversation even though strictly speaking they would be classed as 'digital'.

[3] Perhaps it might be wondered at this stage why I don't simply avoid the difficulty of attempting to define 'conversation' by instead differentiating between face-to-face *language use* and written/online *language use*. My overall project, though, is a comparative reflection and so I'm guided here by the desire to restrict the objects of the comparison rather than deal with the whole complex gamut of language use. This might lead to a further question asking why then I don't restrict the objects of comparison further. The argument might go that perhaps by looking at smaller units of conversation such as individual speech acts and their uptake we might then have a yet more precise and much easier to define set of comparative objects. This is indeed a strategy I think worth pursuing and part of the work in Chapter 5 I take to be the beginning of a look at the types of speech act that arise in digital spaces and is the direction I envisage this whole thesis is directed towards. However, for present purposes I think there are some interesting contrasts that come from a broader look at the underlying differences in the nature of face to face and digital conversations.

conversation to have such properties requires that participants in a conversation will often, on one level, be engaged in some agreed direction of the content of the talk relating to some topic(s) or question(s).[4] On another level, however, these participants are also necessarily highly-engaged in a series of often unconscious meta-tasks some of which require them to pay close and continuous attention to their interlocutor (as well as to themselves) in order to sustain the interaction. These meta-tasks involve, amongst others, prediction and interactive alignment (more on these features in Ch.2 §3.1). One result of the requirement to attend to such meta-tasks is that it marks these types of exchange as different to many other types of linguistic behaviour. It is these features that I take to be worth further attention as the characterisation of conversation develops. Not all of this work is done in this chapter, however. In Chapter 2 I sharpen this characterisation and present a more detailed look at some of the coordinative and cooperative aspects that underpin conversational exchanges. I will show how this correlates with some observations from linguistics, psychology, psycholinguistics, and conversation analysis. The primary purpose of these two chapters, then, is to develop a characterisation of this most basic form of human linguistic interaction for the purposes of comparison (in Chapters 3 & 4) to a newer form of similar-seeming conversations that are afforded to us by recent technological developments.

The plan for this chapter is as follows; Section 1 is a brief outline of some thoughts on why I think such a restricted comparison class of conversation is required (§1.1) and I also give some of my reasoning as to why I think the universality and primacy of conversation should make us think it is an interesting topic in and of itself (§1.2). Section 2 is a consideration of an intuitive way in which we might start to draw a line between conversational and non-conversational linguistic activity – I do this by considering how we might delineate between different types of exchange by using a simple *infelicitous report test*. Using this I begin the task of separating a few different types of interaction into categories according to some of their notable features. Doing this allows us to think about some of the different features of different types of exchanges. In the final section I draw upon observations from three theorists - Maurice Merleau-Ponty, Erving Goffman and Sandford Goldberg – all of whom touch upon a similar rough conception of the type of activity I consider to be a paradigm type of conversation. What each of them suggests points towards a type of activity in which the participants enter into, or create, something which,

[4] And so in this sense they resemble the Greenian/Stalnakerian type of conversation (see fn.1).

to put it crudely, is greater than the mere sum of its parts (§3.1). The idea being that a paradigm type of conversation is something like a unified activity. I suggest that what makes it seem so can be considered to be the result of some of the phenomena we experience in conversation of this type such as the sharing of perspectives (§3.2) and the continuous attention required of us and the joint-commitment this entails (§3.3).

1 Why conversation?

1.1 *Development of a comparison class*

A tidy set of necessary and sufficient conditions to define conversation eludes me here and so what I present is a characterisation of the type of activity I have in mind. What I present here is best considered a cluster concept with a group of properties that might feature to different degrees in some activity. The idea being that the greater the degree of some property, the closer it is to a paradigm case of conversation (with respect to that property). This seems to be the most promising way to define such a broad activity. Anything more prescriptive risks being too restrictive because even some of the features we might think seem to be obvious properties of conversation, such as participants having a language in common, have counter-examples.[5] Avoiding a strict definition also allows us to take seriously one of the concerns in John Searle's critique of the field of 'conversation analysis.'[6]

One of Searle's arguments is that it is unlikely that conversation analysis could ever be as fruitful as, for example, speech act theory due to the generally vague purposes of conversation. So whereas we can isolate and detail quite precisely different types of speech act and apply analyses of a particular type to different occurrences of that same speech act, we have no such way of doing similar with conversation. Searle's point is ultimately a point about content. Whereas we can offer good explanations as to what the content of some particular speech act is or how it functions, stating what the content of conversation is in a similarly explanatory way is not possible. Generally conversations are composed of numerous different speech acts but equally there could be infinite possible contents. Whereas illocutionary acts such as asking a question, giving an order or asserting something have specific communicative roles which are generalisable, conversations are, in terms of content, heterogenous – considered as a whole they vary wildly from one to the next. And so in terms of content Searle appears to be correct about the limitations of a

[5] In that particular case the counterexample is cross-signing discussed in Chapter 2 Section 1.3.
[6] (1992)

theory of conversation. Nevertheless, I will argue that we can characterise conversation in other ways – more specifically by looking at the relationship between the content and processes involved in conversational interactions - and in doing so we *can* find some interesting generalisations. This is no straightforward task, however, and so it will occupy the remainder of this chapter and the following chapter. What I do in these two chapters is build a characterisation of a particular type of interaction. Once we develop some such way of isolating a particular type of interaction as being conversation, this then allows us space to directly contrast it with other types of interaction. My specific purpose here, then, is to use it as a basis for comparison with the type of exchanges that have become prominent in recent years – text-based electronically mediated interactions.

1.2 *The universality of conversation*

Having given my aim for this chapter within the scope of the wider thesis it would be remiss to end a section entitled 'Why conversation?' without briefly noting that one of the major motivations behind the research in this thesis: conversation is interesting in and of itself. It is difficult to think of a human activity not rooted in biological necessity that is quite so widespread.

Conversation is one of the universal aspects of human life; it is the cauldron of languages and central to how we acquire them. So even on this basis, it is worth paying attention to, as many have before; be it in the theorising of David Lewis, H. Paul Grice, or Robert Stalnaker who use conversation as the basis for some of their most profound insights, [7] to theorists such as linguist Charles J. Fillmore who wrote that "face to face

[7] Indeed, conversation as a topic has an indirectly rich history in recent philosophy of language. We need only think of the role conversation has to play in seminal work such as Lewis's 'Scorekeeping in a Language Game' (1979) Grice's 'Logic and Conversation' (1989c), as well as his 'Further notes on Logic and Conversation' (1989a), 'Presuppositions and Conversational Implicature' (1989d) and in sections of the 'Retrospective Epilogue' to his Studies in the Way of Words (1989e), or in Stalknaker's ((1999 [1970]), (1999 [1974]), (2002), (2014)) work on assertion and the common ground, and the considerable debates that have followed them. I say indirectly rich history because although the theories that have developed in response to Lewis, Grice and Stalnaker provide great insights into some of the aspects we find in conversation, what conversation actually *is* rarely receives dedicated attention. That 'conversation' is used as a backdrop to some of the most interesting developments in recent philosophy of language history and yet there seems to be little time spent developing what we mean by 'conversation', might lead us to question whether there is some gap in our theorising. To be clear, the above point isn't that theorists have been ignoring language uses that aren't conversational, there have been many uses of examples of non-conversational language use, such as written notes or voice recordings, to draw out some of the interesting aspects of language use (For example, (Predelli, 1998, 2011) (Carston, 2008, p.326) (Perry, 2003, p. 378) make use of written notes to make points about the referent of 'I', the literature on the so-called 'answering machine problem' (see, for example, Cohen, 2013; Cohen & Michaelson,

conversation is the basic and primary use of language, all others being best described in terms of their manner of deviation from that base".[8] Fillmore's point is developed by Herbert Clark over the course of his *Using Language*, for example he says "[f]ace-to-face conversation…is the principal setting that doesn't require any special skill…[and the] basic setting for children's acquisition of their first language."[9] And conversational language use is, as Stephen Levinson and Francisco Torreira note, "the prime ecological niche for language, the context in which language is learned, in which the cultural forms of language have evolved, and where the bulk of language use happens."[10] What I think unites these theorists is an understanding of the nature of language requires an understanding of conversation, and it is in that spirit that I proceed here. So regardless of the use to which I put the notion of conversation I work with later in the dissertation, and the wider theoretical positioning of such a study, this chapter and Chapter 2 are, I hope, interesting in and of themselves because they examine one of the most universal of human activities.

2 A tentative boundary between conversation and non-conversation

It seems natural to start a characterisation of conversation by considering some cases and thinking about whether they are conversations or not. So at this early stage I use a rough and intuitive tool to draw some lines marking out some guide as to what is and what isn't a conversation. Let's call this unsophisticated tool the *infelicitous report test.* The idea of *the infelicitous report test* is simple; if it seems infelicitous to report an activity as being a conversation, then let's pause to consider why that might be the case.[11] I start in Section 2.1 by providing an example I take to be uncontroversially a conversational exchange and for the remainder of the chapter I use this example as a counter-point to show some of the differences we find with other types of language use such as novels, lectures, cross-examinations, job interviews and scholarly exchanges through journal articles. Contrasting these types of exchange leads me to conclude that some of the important properties of the cluster notion of conversation are *turn-taking* (§2.2) and *synocratic control* (§2.3). In

2013; Connolly, 2017; Romdenh-Romluc, 2002; Sidelle, 1991) all centre on voice recordings, and so are also examples of specifically non-conversational language uses being put to work to make points about the referents of indexicals such as 'I', 'here' and 'now'.)

[8] (1981, p. 152)

[9] (1996, p.9)

[10] (Levinson & Torreira, 2015, p. 1)

[11] Of course, such appeals to ordinary language use have severe limitations, and there are good Gricean reasons to explain why it might sound infelicitous to describe, for example, a cross-examination as being 'a conversation', this might seem perhaps insufficiently informative and so it would be far more informative to say 'cross-examination' when describing such occurrences (and so it need not necessarily preclude such an exchange as being a type of conversation).

keeping with the idea that what is being described here is a cluster notion, these properties are best considered to be linear in nature; that is, particular exchanges may have more or less of the properties by matters of degree. So let's start with an exchange that appears to have all of these features and is close to an ideal form of conversation (were there ever to be such a thing).

2.1 *A typical interactive synchronous conversation*

There certainly seem to be some types of interaction that we might comfortably consider to be conversations. Sadio and Brigitte discussing the news in a cafe, Amelia and Emmeline talking in the pub about how they are going to spend their weekend, or two colleagues talking about their new boss all seem to be uncontroversially the type of occurrence we might consider to be typical conversations. There is a topic of discussion (which may be fluid and ever-changing) and the interlocutors talk about it. It certainly wouldn't seem infelicitous to call such interactions conversations, and although caution with such 'tests' is necessary, it would seem to me that these types of exchange seem to sit comfortably within the category of paradigm cases of conversation.

An example of a typical extract from a conversation of this type might be such as the following example, transcribed originally by Anita Pomerantz.[12]

(1)

01	A:	Just think of how many people would miss you. You would know who cared.
02	B:	Sure. I have a *lot* of friends who would come to the funeral and say what an intelligent, bright, witty, interesting person I was.
03	A:	They would*n't* say that you were *humble*
04	B:	No. Humble, I'm not.[13]

[12] (1978, p. 89)

[13] I've selected this particular example from the conversation analysis literature partly because the transcription is straightforward compared to many examples from the field. The only conventions worth noting here are that each line corresponds to approximately 1 second of speech, and, as you

On the surface of (1) it seems that A and B both take turns to speak, the topic appears to be fluid, moving perhaps from death at 01 to funerals at 02 on to B's self-aware lack of humility at 03 and 04. Without knowing anything about the context of (1) there seems no good *prima facie* reason to regard this extract as non-conversational. It is likely only an extract of a longer exchange, but it seems felicitous to say that (1) is a conversation between A and B. As such, I'll use (1) as a reference point and by contrasting (1) with other examples we will also start to understand a little more of what is happening in (1).

2.2 *Turn-taking interactivity*

When compared to examples such as (1) we might think that it would seem less felicitous to say that Salka reading the novel *Things Fall Apart* by Chinua Achebe is a paradigm type of conversation, or that Paul Grice giving his William James lectures was a paradigm type of conversation.[14] If Salka reported upon reading *Things Fall Apart* that she'd just had a conversation with Chinua Achebe one might think it would be infelicitous to describe reading a book as such, and so Salka's utterance might be taken to be metaphorical speech of some sort.[15] Similarly an attendee of one of Grice's lectures who reported that she had 'had a conversation' with Grice might be taken to mean she'd spoken directly with him. One reason for this seeming infelicity could, therefore, be that such statements are insufficiently informative, to describe a lecture as a conversation or the reading of a novel as such seems to miss some important details from the description. For some purposes though it may be that we can consider such activities to be conversations-of-sorts. For current purposes, however, what I take to be notable with these examples and cases such as (1) is the differences in the nature of the interactivity when contrasted with examples such as (1).

In (1), A and B both contribute directly to the discourse in a rapid to-and-fro. In the case of the novel or the lecture, however, Salka and Grice's audience member don't directly

probably expect, *italics* reflect emphasis in the speech. In many conversation analysis transcriptions the convention of using 01, 02, 03, is used for every line (as opposed to every utterance as here), this gives the analyst a picture of the duration, which isn't essential here, though I find the numbering as presented useful for referencing specific parts of the dialogue.

[14] Green's certainly wouldn't be quite so quick to disregard these types of activity from the class of conversation. For Green that these types of occurrence both have topics of discussion, and that topic is what is spoken (or written) about, and so these too might be considered conversations.

[15] Green (2017, p.1601) doesn't take such use to necessarily be metaphorical, for example he says; "Antonio Damasio in Descartes' Error reminds readers at numerous points in the text that he thinks of himself as engaging in conversation with them. We need not take him as speaking metaphorically."

make a contribution to the content of the exchange - Achebe and Grice 'do all the talking'.[16] In the terms used by conversation analysts, the difference here is that in examples such as (1) participants are engaged in *turn-taking* whereas in the example of a novel or a lecture no such turn-taking is present. So in (1) the participants take turns to perform different roles, at 01 A speaks and B listens, at 02 the reverse is true. Whereas in attending a lecture or reading a book, the roles remain almost consistently static.[17] In such activities one party almost permanently adopts the role of speaker/writer and another the role of listener/reader. So what I suggest as an initial property of the type of interaction I am interested in here is that it will generally be interactive – that is, it will generally be a turn-taking activity. Now of course we've all likely had conversations in which the great majority of one role is taken on by a particular participant – sometimes in a face-to-face exchange one person will very often 'do all the talking' as if they are delivering an academic lecture. When some particular conversation takes on such a dynamic, then perhaps it moves away from being a paradigm case of conversation and its nature becomes closer to that of a novel or academic lecture, however unlike a novel or academic lecture, that the opportunity for discussion to return to a more balanced turn-taking enterprise allows us to consider it to be closer on the scale to a 'conversation' as I am conceiving it. So it isn't necessarily the turn-taking by itself that makes our paradigm type of activity 'conversation'. As we shall see in §2.3 some linguistic activities are intrinsically turn-taking activities yet might not be best regarded as conversations (at least by using the infelicitous report test) and so what I suggest next is we look at who determines the shape and content of these turns. Specifically, I next want to suggest that an important characteristic of the type of activity I am regarding as a paradigm type of conversation is that the participants each have a degree of control over the direction of the discussion. That is, an ideal conversation operates like what we might call a *synocracy*.

2.3 *Synocracy*

In this section I want to highlight the role of mutual control between participants in a paradigm case of conversation. My sense is that in the case of a novel or a lecture the direction and topic of the activity is dictated almost entirely by the author or lecturer. If this

[16] Though perhaps it could be argued that upon reading a novel or listening to a lecture we are in a sense interacting with the writer/speaker, they have produced/are producing language for us to comprehend. And the interaction is in the comprehension. Further to this, it might be argued that any content that is audience-generated is a contribution to 'the content of the exchange'.
[17] Though as pointed out to me by Rosanna Keefe, it could well be the case that one could perhaps shout at the TV, or mutter something whilst reading a book.

is the case, we might therefore consider such activities to be generally quite autocratic in nature (with a novel often being more autocratic than a lecture due to its pre-determined state by the time a reader comes to it). Contrast this again with (1). In such turn-taking exchanges the direction of talk is more broadly dictated by the current speaker at any one time. As these roles switch during the course of such an exchange, we might expect that, broadly speaking, control over the direction of discussion is spread between the participants. In a sense, then, interactions such as (1) are *synocratic* in nature; they are collaboratively controlled. However, we shouldn't be misled into thinking that it is turn-taking itself which leads to such a synocracy. As we shall see in the remainder of this subsection, there are types of exchange which have similar turn-taking structures such as the above but seem to be much more autocratic in nature.

2.3.1 *Cross-examination*

It might seem odd to describe the experience of being cross-examined in court as having had 'a conversation with a lawyer', [18] but it is undoubtedly a turn-taking activity - perhaps even more rigidly so than examples such as (1).[19] Take the following exchange from the Oklahoma City Bombing trial on 12 November 1997.

(2)

01 Q. Did he tell you he had seen a pickup truck?

02 A. Yes.

03 Q. All right. Did he describe the pickup truck to you?

04 A. Yes. He said it was a dark-colored –

05 Q. Hold it a second. Did he describe it? Yes or no.[20]

When comparing (2) and (1) there are some obvious similarities. In both examples each participant takes turns to respond to the previous participant's contribution. In both examples the interaction proceeds synchronously; as one speaks the other listens and

[18] Though Green (1999) notes cross-examination as a type of conversation, he doesn't include them in his (2017) taxonomy.
[19] More rigidly turn-taking in the sense that it would generally be against the purposes of the court if a cross-examiner were to not allow the cross-examinee opportunity to speak, and it would be unlikely a cross-examinee would be given opportunity to talk at length interrupted.
[20] This and many other similar examples can be found in Maj Britt Mosegaard Hansen's (2008). This from (p. 1400)

responds to what has been said. And in both examples there are (tacitly, at least) agreed directions of talk in that all participants appear to be addressing similar or related topics. There are some notable differences too though, most pertinently there are notable differences in control of the exchange.

In (2) Questioner places strict restrictions on what Answerer can say. Take for example Answerer's response at 04 to the question posed at 03. Ordinarily we might expect a question such as 'Did he describe the truck to you?' to carry an implicature along the lines of 'if so, how did he describe it?', and we see Answerer at 04 respond as if it does carry such pragmatic content. At 05, though, Questioner cancels the implicature and makes clear it is merely the polar question that should be answered. The cross-examiner at such a point appears to exert control on the direction of the exchange even to the point of directly shaping how Answerer can respond.[21] [22] In cases such as cross-examination, then, we perhaps see much more autocratic interactions than in cases such as (1).

In (1) we might reasonably speculate that A and B have a much more evenly balanced control of the direction of conversation than Questioner and Answerer do in (2). It will of course be the case that in most conversations there are social pressures or expectations that bear upon what participants might feel they can contribute to an exchange, though when thinking of the idealised paradigm, we might expect that its nature is much more synocratic than autocratic. [23] The participants are each granted a degree of control of the direction of the conversation. In a cross-examination, however, one party to the exchange has a much greater degree of control of the content of the discussion than the other. That these interactions are initiated in such a way and conducted in the way that they are makes them intrinsically and profoundly authoritarian. There is no real symmetry of influence in cross-examination, the cross-examiner determines the direction of talk and the potential penalties of contempt of court and perjury place severe restrictions on how a cross-examinee can respond. What I suggest then is that exchanges such as cross-

[21] See (Borg & Connolly, forthcoming) for further discussion of how this type of case is best considered in terms of linguistic liability.
[22] There are also other interesting dynamics in such interactions though, because even though *prima facie* the cross-examiner generally has control over what she can ask of the cross-examinee, this authority is also subject to an even greater authority – that of the court. Indeed, the cross-examiner is also limited in what she can discuss, she wouldn't be given much chance to idly discuss the day's weather with the cross-examinee, for example, if it was superfluous to the case.
[23] There is an interesting consequence of considering degrees of synocracy as an important aspect of conversation, as it will be the case that in some conversations not bound by the conventions or rigid power structures of, for example, a court, the dynamic is much more autocratic, sometimes by consent, but also for other reasons. I'll pick this point up more in Chapter 2 (§4.2.2)

examinations drift further from the paradigm type of conversation on account of being autocratic in nature. In the paradigm type of conversation we should expect not only turn-taking to be notable, but also that participants have some degree of control over the direction of the talk.[24] Let's now consider another example of a turn-taking exchange, though one which we might think contains a higher-degree of synocracy than a cross-examination.

2.3.2 *Job interview*

Despite also being fundamentally a turn-taking exchange, we might also think it seems infelicitous to report that a job interview was a conversation (although perhaps to a lesser extent than we would a cross-examination). Generally speaking the interview room is less rigidly structured than the courtroom. And interviewees will have more opportunity to speak freely than those under oath and being cross-examined.[25] It is also the case that one could imagine a job interview in which parts of the exchange between the interviewer(s) and interviewee could even have a similar character to exchanges such as (1). For example, participants in a job interview may discover they have a similar hobby and potentially discuss that as equals, and there may be elements of the job being interviewed for that allow for a looser discussion. Indeed some interviewers may purposely decide to structure an interview as being closer to an informal chat than a formal interview.[26] However, even if the interviewer were to conduct the interview as such, and even if there are moments in an interview which seem to break from the hierarchical structures, we might think that it still doesn't entail that in general such interactions are not much more autocratic in nature than examples such as (1).

[24] Another difference worth noting is that in (2) the Questioner isn't speaking to Answerer alone, and nor is Answerer answering questions for Questioner alone. The proceedings of a court case are often public, and even if not public they are mostly played out in front of an audience of jurors, judges, legal representatives, defendants, plaintiff etc… So (2) isn't a private discussion between two interlocutors, it is a form of interactive public speech which is initiated for the formal purposes of the court. What I think marks this as significant is that it creates a conversational dynamic that often isn't present in paradigm face-to-face conversations in which the audience is the interlocutors.

[25] Indeed, unlike in the courtroom we might expect a polar question in a job interview to carry many of the conventional implicatures we find in day to day conversation. A question such as 'Have you had experience teaching?' would likely be given as a request for specific details of such experience rather than a mere 'yes' or 'no' response.

[26] Though as is pointed out to me by Rosanna Keefe, some of the most informative interviews are in many respects much less autocratic and the interviewee is given opportunity to lead the direction of the conversation.

We would probably expect that most often it will be the case that the interviewer in a job interview still has a position of power afforded to them by their role as interviewer and arbiter of whether or not to employ the interviewee. And although the penalties for the interviewee might not be thought to be as strong as in the case of a cross-examination, failure to satisfy the examiner could potentially carry the penalty of not getting the job. It appears then that what they can say is limited by what the interviewer expects or wants (or what the interviewee takes the interviewer to expect or want) [27] and so the stakes for the interviewee are high. For the interaction to be successful the interviewee will be required to submit to the expectations of the interviewer in a way that is not reciprocated.[28] As such even though a job interview has a turn-taking structure and a greater degree of synocracy than, for example, a cross-examination, the control of the direction of the talk is generally dictated by the interviewer. As such, on a continuum we might think a job interview to be less conversational than examples such as (1), yet more conversational than examples such as (2).

2.3.3 *Simulations?*

I have suggested that what makes examples such as cross-examinations and job interviews different from examples such as (1) is rooted in the different degrees of synocracy. Though it is worth noting that beyond some of the most obvious differences, there are other contrasts we might wish to draw which are also rooted in this notion of synocracy. For example, we might think a fundamental difference lies in both of these types of interactive language use being generally formal and structured. That is, they both have a rigid set of

[27] It is worth noting that although less obviously formalised, these hierarchical structures exist in what we might also regard as day to day conversations. A seemingly informal chat between someone in a position of power over their interlocutor(s) can often have this type of character too, and often the less-powerful party to the conversation will have very little input as to the direction of discussion and will, in effect, be silenced in this regard. In such cases there may be social penalties, or penalties for their employment or career, for example. Although I don't have space to elaborate on this point as much as I would like to, I think considering conversation in terms of synocracy can help us to understand some of the fundamental reasons as to why we should regard these types of interaction with suspicion. For without synocracy, I would argue perhaps they aren't really conversations at all, or at least they are distant facsimiles of conversations. And indeed, a part of the injustices of such interactions is based on the power-broker not affording an interlocutor a genuinely conversational role in the interaction. I pick up this point again in Chapter 2 Section 4 where I begin to refine the nature of the synocracy in conversation.

[28] Although there may be some cases in which this isn't the case. It may well be that the interviewer is desperately keen on employing that particular person who they know has also been courted by other employers.

goals[29] and the direction of the talk is often that one party to the interaction asks questions, and the other party answers them. Or perhaps one party dictates the direction of discussion to such a degree that the dynamic is one of information extractor and information dispenser. All of these types of reason, I think, are related to the asymmetry of control in such interactions. In his 'Retrospective Epilogue' Grice considers such examples[30] to be a "secondary range of cases" from the types of interaction he is considering. Grice suggests that in cross-examination, for example, "the common objectives are spurious, apparent rather than real; the joint enterprise is a simulation, rather than an instance of even the most minimal conversational cooperation".[31] Grice's idea that what is happening in such cases is a 'simulation' touches on what I take to differentiate (1) and the other examples. Although a cross-examination or job-interview might appear on the surface to be of a similar type to examples such as (1), there is something fundamentally different about them. In Section 3 I begin to discuss in more detail what it might be that makes them different types of activities, and in Chapter 2 (§4.2.2) I sharpen further the notion of synocracy.

2.4 *Summary*

In this section I begin to characterise what a cluster notion of a paradigm conversation might look like by highlighting two features I suggest we might expect to be present (to some high degree) in a paradigm type of conversational exchange. These features are what I call turn-taking interactivity and synocracy. In Section 2.1 I present example (1) as being a typical conversational extract. In this example we see two people exchanging short turns that roughly follow on from the previous turn, it is in essence interactive and the roles of the participants switch continuously between speaker and hearer. We can hypothesise that the participants are relatively unrestricted in what they contribute, and there is an approximate balance between the contributors to the direction of the conversation. In (1), then, it seems that A and B each contribute to the activity as approximate equals to the progression of the discussion. This is in contrast, however, to some other types of linguistic activity. In Section 2.2 we looked at two other types of linguistic activity – the reading of a novel and an

[29] For example, in a cross-examination this rigid goal will depend on who is doing the cross-examining; be it, for example, a representative for the defence trying to show cause to regard her client as being not guilty. Or in a job interview, the goal is to determine a candidate's suitability for a particular job.
[30] Although he names only cross-examination here, I take what he says to apply to cases such as job interviews.
[31] (1989e, pp.369-370). I return to discussion of this in Chapter 2 Section 4).

academic lecture. These types of activity are much less obviously interactive. In the case of a novel, on the one hand we have an author and on the other we have a reader. And in the case of an academic lecture we have the lecturer and the audience. In the case of a novel the roles remain entirely static - in most cases the reader has no bearing on the contributions of the author. In the case of a lecture the roles are not as strictly rigid; an audience member could directly contribute in some ways, but generally the lecturer directs almost all of the discussion and makes most of the contributions. From this I suggest that the paradigm type of conversation will be highly turn-taking in nature.

Whether an activity is a highly turn-taking activity or not is not enough to suggest it is a paradigm case of conversation, however. For it is the case that there are turn-taking linguistic activities that also seem to differ in some ways to examples such as (1). In Section 2.3.1 I suggested that in the case of courtroom cross-examination we see a highly interactive exchange that is nonetheless quite different in some important senses to example (1). In the courtroom the cross-examiner is largely in control of the direction of the exchange, she asks the questions and can exert a degree of control over the manner in which they are answered. I argue there is a similar dynamic in examples such as a job interview, though to a lesser extent. In the job interview the interviewer largely directs the interviewee as to the topic to discuss, though the interviewee will more likely be given more freedom to answer in a way she sees fit. Indeed an interviewee might be given the opportunity to ask her own questions, and there may be sections of the interview which are much less rigidly structured than court dialogue ever is. What we can take from this, I suggest, is that in cases such as (1) the conversation is something like a synocratic endeavour. In a paradigm type of conversation the control of direction of the conversation rests to some degree with each of the participants. As such, on a cluster notion of conversation, the paradigm cases will be ones in which the activity is generally a turn-taking interaction in which the control of the conversation is closer towards synocratic on a scale that ranges from synocracy to autocracy. What I want to consider next is how these features of turn-taking synocracy create the conditions for us to experience conversation as being something like a unified joint-activity.

3 Conversation as unified activity

In this section I consider how the turn-taking interactivity and synocratic nature of a paradigm conversation as described in Section 2 contribute to the creation of what Maurice

Merleau-Ponty's describes as "a being-shared-by-two".[32] I take Merleau-Ponty's idea to be similar to Erving Goffman's suggestion that when conversing conversational partners enter into an "*unio mystico*",[33] and I think that what this creates is something like what Sandford Goldberg's describes as a co-authored piece of performance art using the spoken word.[34] I want to emphasise that a paradigm type of conversation is best considered to be something like a unified action and that this unity is based in continuous and necessary perspective sharing and attention to the conversation. In order to do this, in Section 3.2 I consider another type of exchange (that of an exchange of scholarly journal articles) which seems to possess the turn-taking and synocracy of exchanges such as (1), but which we might still have reservations about adding to the class of paradigm conversations. I suggest, then, that such interactions lack the continuous attention of its interlocutors to the task of communication that characterise face-to-face conversations and that this is fundamentally significant for the type of activity we might consider it to be.

This section comes with something of a promissory note, for although much of what I do here is descriptive and hypothetical, in Chapter 2 I offer what I take to be compelling evidence from psycholinguistics on the structural processes of conversation that indicate why the paradigm type of conversation has some of the characteristics it does. The overall aim here, then, is to ready us for the discussions of Chapter 2 where we look at the coordinative requirements of such an activity and consider how this is underpinned by a distinct and necessary form of coordination and cooperation.

3.1 *Interdependence of contribution*

Paradigm types of conversation in which we find a high degree of turn-taking and of synocracy have a character captured by Erving Goffman in his essay 'Alienation from Interaction'.[35] On the topic of conversation he says;

> As a main focus of attention talk is unique… for talk creates for the participant a world and a reality that has other participants in it. Joint spontaneous involvement is a *unio mystico*, a socialized trance… the individual must not only maintain proper involvement himself but also act so as to ensure that others will maintain theirs. [36]

[32] (Merleau-Ponty, 2012, pp. 370–371)
[33] (Goffman, 1967, p.113)
[34] (2020)
[35] (Goffman, 1967)
[36] (Goffman, 1967, p.113)

And later;

> Many social encounters of the conversational type seem to share a fundamental requirement: the spontaneous involvement of the participants in an official focus of attention must be called forth and sustained.[37]

What Goffman is getting at is that fundamentally conversation is an endeavour to which participants must give almost unceasing attention. Though it's true that conversation can be conducted whilst eating or drinking or carrying out other habituated tasks, to participate in conversation is to be directing something like one's attentional capacities[38] to the task of the interaction. That there is, according to Goffman, an 'official focus of attention' seems to fit with Mitchell Green's idea that conversation is an activity that is directed towards an inquiry or deliberation, and so we certainly have no need to exclude that this may be a fundamental aspect of most interactions. However, in many types of interaction there is also a more fundamental further focus required of the participants; that of continuously maintaining the interaction. For in many of these types of interaction, there is a sense in which participants must not only ensure that they themselves are sufficiently focused on the interaction, but also to make sure their conversational partners are similarly sufficiently focused. Disengagement in conversation is often obvious, and also contagious – if I notice my conversational partner's eyes glaze over, my own attention can begin to waver. It is therefore essential for successful communication that participants remain focused on the communicative responsibilities of the task. And as I will argue in more depth in chapter 2, this level of required focus is indicative of how a particularly enjoyable conversation leads us into a 'socialised trance' or *unio mystico.* This also accords closely to something like the following description Maurice Merleau-Ponty ascribes to dialogue;

> Here there is a being-shared-by-two…We are, for each other, collaborators in perfect reciprocity: our perspectives slip into each other, we coexist through a single world.

---

[37] (Goffman, 1967, p.134) It's probably worth noting that I don't take the spontaneity Goffman refers to here as being the spontaneity of entering a conversation, for example one could call someone on the phone with the purpose of having a conversation, or arrange to meet them for that purpose, but the way it proceeds will, in most cases, be quite spontaneous.

[38] I'm aware 'attentional capacities' sounds quite woolly. I don't have a precise definition of what I take this to be, though my hope is it becomes clear what I mean from the discussion around it. Maybe we could take it to be something such as described by William James in Chapter XI of his *Principles of Psychology* wherein attention "is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others" (1890, pp. 403–404).

> I am freed from myself in the present dialogue, even though the other's thoughts are certainly his own, since I do not form them, I nonetheless grasp them as soon as they are born or I even anticipate them. And even the objection raised by my interlocutor draws from me thoughts I did not know I possessed such that if I lend him thoughts, he makes me think in return.[39]

Merleau-Ponty touches upon here the overall unity of the activity we find in certain types of interaction. By 'being-shared-by-two', Merleau-Ponty seems to be referring to something like Goffman's *unio mystic*, but further to this he also seems to be drawing attention to the innate interdependence of between the contributions of each participant. So when he says that "even the objection raised by my interlocutor draws from me thoughts I did not know I possessed such that if I lend him thoughts, he makes me think in return." Merleau-Ponty is alluding to the way in which in many exchanges, the contribution of A will very often provoke novel thoughts in B. B might then in turn respond in a way that has a similar effect on A and so on… So the contributions to this type of interaction is dependent not only on the interlocutors as individuals, but on the effects they have on each other.

I take this idea to also chime with Goldberg's idea of a conversation as being like a co-authored piece of performance art. He says;

> I have often thought that a truly great conversation is akin to a piece of performance art involving the spoken word in which each participant is a co-author. And I've often thought of the joy of conversations as at least in part the joys of co-authorship.[40]

Although Goldberg states that he is exaggerating for the purposes of vividness here, he touches upon something important which is that being in conversation can often be a positive experience. And part of this positivity relates to the fact that it *co-authored* or, we might say, highly collaborative at every stage. As such, for interactions of this type it seems there is an interdependence of participants that makes the conversation something like a unified thing. This feature of interdependence, together with the turn-taking synocracy discussed in Section 2 gives us good reason to think there is something sufficiently different between the types of exchange I consider here to be examples of paradigm conversations and types of linguistic activity that are further away from the paradigm (such as cross-

[39] Merleau-Ponty (2012, pp. 370–371)
[40] Goldberg (2020)

examination, job interviews, novels or academic lectures). For if this reliance on the contribution of others for the formation of one's own contributions is real, then it would entail that a conversation is not dependent in a trivial sense merely on, for example, the interlocutors A and B. That is to say, although it is trivially true that some particular conversation couldn't exist without A and B, what is of interest is that it relies on the relationship of A to B (or perhaps more precisely put, it is the relationship of A's contributions to B's contributions). And this, I suggest, points us towards another of Merleau-Ponty's observations – that in conversation participants continuously share perspectives.

3.2 *Shared perspectives*

When considering what type of exchanges sit closer to the paradigm type of conversation the interdependence of the contributions of the participants that comes from the turn-taking synocratic nature of certain exchanges is important. However, there is something else about the character of these certain types of interaction that I think worth considering. Returning again to the passage from Merleau-Ponty in Section 3.1, another thing to note is the idea that during dialogue 'perspectives slip into each other'. What I take Merleau-Ponty to be referring to is that during interactions of a certain type there will often be a requirement that interlocutors allow that their own thoughts be close to the thoughts of their interactive partner's. In a sense this is what it is to properly and actively listen to someone speak. We might however think that this notion of perspective sharing is common in many sorts of language exchange. Let us consider next another type of exchange - that of a series of journal articles written in response to each other. [41] [42] These types of exchange involve turn-taking, synocracy and taken as a whole might be considered to be something like Goldberg's co-authored piece, though they are, I suggest interestingly different to what I consider to be the paradigm types of conversation.

[41] I choose this example again because Mitchell Green categorises such an interaction as a type of conversation.

[42] The following exchange is one such example worth drawing attention to (chronologically ordered here) (Bach, 2006c; Cappelen & Lepore, 2006b; Bach, 2006b; Cappelen & Lepore, 2006a; Bach 2006a). Although strictly speaking it starts with Bach (2006c) replying in a journal to (particularly a criticism of Bach in) Cappelen & Lepore's book *Insensitive Semantics* (2005), the responses that follow it are all manuscripts. That they are published in journal or not doesn't seem to be a significant difference. Arguably though, there are two different questions being discussed, one by Bach who suggests that Cappelen & Lepore's semantic minimalism relies on a form of what he calls 'propositionalism' and Cappelen & Lepore seem interested in discussing whether Bach's radical semantic minimalism leads down a slippery slope to contextualism.

Suppose we have a series of journal papers S and for simplicity's sake let's say each of the papers are written by one of the same two authors – Autorin and Escritora. These two authors write in response to each other on the topic of some particular question.[43] Each paper in S may well be intelligible on its own, but if the format of S is that Autorin responds to Escritora's points and vice versa then there is a sense in which what Autorin writes will, to some degree at least, have a direct bearing on what Escritora writes. And that which Escritora writes will have consequences for how Autorin responds. As such, in the case of S we appear to have a turn-taking exchange (unlike the example of a novel) with a degree of synocracy (unlike, say, a cross-examination) and also it has the type of interdependence of contributions that I suggest above is characteristic of paradigm types of conversation. But not only that, it appears, too that in order to respond to each other's points, Autorin and Escritora must also at least attempt to share the perspectives of each other in the sense of trying to understand what they have written.[44]

When Autorin is reading a reply from Escritora to a previous paper, she may well allow herself (or be required) to take on Escritora's perspective to comprehend what Escritora might be saying. And the reverse might happen once Autorin's reply is published and read by Escritora. But there is an aspect of perspective taking that does not occur in the case of S but which necessarily must happen in a paradigm type of conversation. For in face-to-face interactions, this perspective sharing is a continuous and synchronous occurrence, and importantly it is done in the presence of an interlocutor as opposed to alone which has consequences for the way in which communication is coordinated (as will be detailed in Chapter 2). In cases like a series of journal articles, more often than not a reader will read an interlocutor's contribution remotely, think about it remotely and, if they write a response, they will write it remotely.[45] The influence that an interlocutor has on the direction of the exchange in such cases therefore ends when their 'turn' ends. Unless they send a further reply, they are no longer actively communicating with their interlocutor. And

[43] For Green S would be a conversation of type *symmetrical inquiry* – one of the two categories (along with *symmetrical deliberation*) into which most, though not all, of the types of exchange I regard as conversation would most likely be taxonomised.

[44] Put this way we might think that most types of language exchange involve *perspective sharing*, and this will be true even in cases such as the reading of a novel. However, in such cases as the novel, this perspective sharing is generally uni-directional. The reader takes on the specific perspective of the author, whereas most often the author will only at best take on the perspective of some hypothetical audience.

[45] By 'remotely' I specifically mean remotely from the other participant(s). So upon publication of one of Autorin's papers, Escritora will read it remotely from Autorin, think about it remotely from Autorin etc…

in cases of face-to-face interaction this isn't the case. Although turns to speak might end, participation in the interaction continues – that is, in a paradigm case of face-to-face conversation the job of communication is continuous whether a participant is speaking or not (and so the perspective sharing it entails is fluid and continuous). I don't expect the importance of this point to be completely clear at this stage but I will argue in Chapter 2 and Chapter 4 that this difference has consequences for the necessary psycholinguistic processes required to sustain face-to-face and digital conversations and that these consequences are therefore significant. To foreshadow the argument in those chapters slightly, there we will see that whereas in face-to-face conversation what I call *the communicative load* (that is, the burden of communication) is continuously shared between the participants, in digital conversation (much like in the case of S above) this communicative load is passed almost in its entirety from one participant to the other. For now though, let's finish the work of this chapter by highlighting one more point from Merleau-Ponty's characterisation of conversation that will play an important role when we look in Chapter 2 at the psycholinguistic conditions required for successful face-to-face communication.

3.3 *The role of prediction*

Merleau-Ponty notes that when we are involved in a face-to-face interaction, it is common that an interlocutor can grasp the thoughts of another 'as soon as they are born' or even 'anticipate them'. In Chapter 2 I highlight some of the medium-specific features of face-to-face communication, for example speech is evanescent, as such this places upon interlocutors a set of restrictions – they must hear and process an utterance in real time. We will also see in Chapter 2 (§§2-3) how some of the consequences of these requirements have a profound effect on the types of cognitive tasks conversational participants must carry out in order to sustain a conversation. To be able to respond in a timely manner in a verbal face-to-face conversation, for example, requires that the listener make predictions as to what type of speech act their interlocutor is performing and also what the specific content of what they are saying might be before they have even said it. And in Ch.2 Section 3 we look at some of the ways conversational partners interactively align during face-to-face conversation, which again mark it as a different type of activity to examples such as S. So what separates, for example, S and a face-to-face interaction is that although in many respects the interactions are the same, the demands placed on the interlocutors are significantly different. At the crux of this is that in cases such as S the interlocutors to the

exchange, Autorin and Escritora, both contribute to the exchange solitarily. They read alone, think alone and respond alone. Face-to-face interactions, on the other hand, require not only a continuous attention be paid to an interlocutor, but also that interlocutors predict what each other might be doing or saying.

4 Summary

In this chapter I have suggested a cluster of features of a type of interaction that I am calling a paradigm class of conversation. Such conversations are fundamentally turn-taking activities. Though turn-taking alone is not sufficient for an exchange to be regarded as a paradigm type of conversation. Exchanges such as a cross-examination or job interview are also turn-taking in nature but seem to have a different character to many day-to-day exchanges we might have. What I suggest is different in these types of cases is related to the control of the exchange. Cross-examinations are particularly autocratic in nature, for example, in that the cross-examiner is the one who directs the exchange. What I suggest, therefore, is a feature of paradigm types of conversation is that they be synocratic in nature – the control of the direction of the exchange is generally more evenly distributed. This turn-taking synocracy helps create some of the conditions for participants to enter into a conversational union of sorts with their interlocutors, (or in Goffman's words, they enter an *unio mystico*). In order for this union to occur, the conversational partners must commit both to the topic under discussion, in the sense of giving it their attention, but they must also commit to the interaction itself. That is to say, parties to a conversation must not only be engaged in paying attention to what is being said by others, but they must also be also be engaged in predicting what others might say, or how they might respond. And they must also make sure others too are engaged with the task of the conversation. In the next chapter, then, we shall look at what unifies these features by looking at Herbert Clark's work on the role of coordination in conversation, Paul Grice's work on cooperation in conversation and then considering how observations from psycholinguistics can help us ground this conception of conversation.

2

Conversational Requirements; Coordination and Cooperation

In Chapter 1 I proposed a type of interaction that we might call a paradigm type of conversation. In this chapter, I suggest we can understand more about the nature of such conversations by considering some of the requirements placed on interlocutors in such interactions. When we look at the detail of the type of tasks participants must necessarily undertake in order for a conversation of this type to function, we start to see how highly coordinated conversation is. Using Herbert Clark's work in his *Using Language* (1996) on coordination in conversation, and in particular his argument that conversational participants must coordinate at both the level of content and process, I suggest that when we focus on the process aspect of this coordination we can see that conversational participants are engaged in a rich set of continuous interactive tasks. In Section 3 I draw on recent psycholinguistic work on response speeds, language production latency and prediction to paint a more vivid picture of what some of these necessary tasks are. Finally in Section 4 I argue that in order for interlocutors to coordinate in such a way requires that for conversation to function, interlocutors must at the very least agree to cooperate with interlocutors at the process level. As such face-to-face conversations such as our paradigm type are necessarily highly cooperative.

1 Basic features of paradigm conversation

Having given an outline of some of what I take to be the nature of the paradigm cases of conversation in Chapter 1, it is worth beginning this exploration of some of the requirements of conversation by thinking about some of the most obvious surface-level features of such exchanges. We might say of our paradigm case of conversation that;

(1)

(i) It has two or more participants;

(ii) It involves turn-taking exchanges;

(iii) It is conducted in a language shared by participants;

Let's briefly consider each of these features in turn.

1.1 *Two or more participants*

It may seem in some sense that it is possible to have a conversation with oneself either silently or out loud, and so we might wonder whether this is genuinely a requirement of conversation. Mitchell Green's account of conversation is much more accommodating of different types of linguistic activity and so soliloquy for Green can be a form of "conversation with ourselves".[46] For my purposes of contrasting face-to-face and digital conversation, however, I specifically want to focus on how different parties to a conversation interact with each other and so the requirement for multiple participants is essential. Recall too the discussion in Chapter 1 (§3) and Goffman's idea that conversational partners enter into a *unio mystico*, as such this would *prima facie* seem to entail that there is more than one participant (there being nothing particularly mystifying about the union of one). So what is primarily of interest to me here is how people interact with each other. And as discussion turns later to the nature of coordination and cooperation between interlocutors in a conversation it is important that there be more than one participant.[47]

---

[46] (Green, 2017, p. 1595). Even harking back to the *infelicitous report test* used in Chapter 1, there perhaps seems nothing unusual about saying "I had a conversation with myself".

[47] It's worth noting that what 'two participants' might mean is not completely clear. We might wonder about cases of the condition Dissociative Identity Disorder (DID). Some argue that such a condition is indicative of two individuals in one body (for example; (Gray Hardcastle & Flanagan, 1999; Tye, 2005)), others argue that symptoms in such cases should be treated as resulting from internal conflicts of a unified self (for example; (Maiese, 2016)). Whatever position we take on this, however, if we were to generally regard such cases as being examples of two (or more) distinct personalities, reports of these dissociated personalities interacting directly with each other are not common. As the nomenclature would suggest, in cases of DID it almost certainly doesn't occur, being as one of the most common aspects to the disorder is that one identity 'takes over.' This being the case we shouldn't expect that two would directly interact. It's also worth noting one recent study into a less extreme condition referred to as 'multiplicity'. (Ribáry et al., 2017) explain this as being on a spectrum with DID. Different identities in cases of multiplicity are aware of each other, and usually have a central controlling host personality (a 'system'). So this may suggest a potential for interaction between different identities. For current purposes, however, I just wish to consider cases in which there are two distinct physical entities.

1.2 *Turn-taking exchanges*

The construction and organisation of turn-taking in the course of conversation is most notably discussed in the field that has become known as 'conversation analysis' following from the work of Harvey Sacks, Emanuel Schegloff & Gail Jefferson *(*1974*)* and Schegloff*,* Jefferson & Sacks (1977). For current purposes we don't need to go into too much detail about some of the general properties Sacks *et al.* and conversation analysts take to be stable features of conversations regardless of context. Though it is worth noting a little on what Sacks, Schegloff and Jefferson say;

> Conversation can accommodate a wide range of situations, interactions in which persons in varieties (or varieties of groups) of identities are operating; it can be sensitive to the various combinations; and it can be capable of dealing with a change of situation within a situation. Hence there must be some formal apparatus which is itself context-free, in such ways that it can, in local instances of its operation, be sensitive to and exhibit its sensitivity to various parameters of social reality in a local context. Some aspects of the organization of conversation must be expected to have this context-free, context-sensitive status; for, of course, conversation is a vehicle for interaction between parties with any potential identities, and with any potential familiarity. We have concluded that the organization of TURN-TAKING for conversation might be such a thing. That is, it appears to have an appropriate sort of general abstractness and local particularization potential.[48]

So turn-taking, for the project of conversation analysis, is taken to be the basic component of conversational language use; it is a feature of conversation regardless of context. And as we shall see later in this chapter, the way in which turns in a conversation operate is insightful as to the type of activity it is. For this reason and for reasons outlined in Chapter 1 (§2.2) I will take it as given that conversation depends essentially on turn-taking for its character.[49]

---

[48] (Sacks *et al*., 1974, pp.699-700)

[49] Or paradigm cases of conversation do at least. One might argue that an occasion in which one person speaks at length with no response from an interlocutor (be that informally in a one-sided conversation, or formally in a lecture etc…) might, in some sense, still appear to be a conversation of sorts. Recall Green's conception of conversation. For any exchange categorised as *asymmetrical didactic* conversations in Green's taxonomy, there is certainly no requirement for a turn-taking exchange to take place, just so long as there are inquiring or deliberative questions salient, there is an occurrence of conversation. To reiterate, I don't here intend to make the point that conversation requires turn-taking and thus anything that doesn't involve it isn't conversation. It can just as easily

Turn-taking (as Sacks *et al.* note)[50] is of course not unique to conversation; it is the basis for games, traffic management, or serving customers in a business, for example. It is even essential to some of the talk-exchanges I suggest are more distant from the ideal of a paradigm type of conversation discussed in Chapter 1 (recall a series of journal articles addressing each other are dependent on turns being exchanged, and cross-examinations and job interviews are formally and rigidly turn-taking enterprises). So we can't hope to define conversation on this alone. However, that so much of what I argue in the remaining chapters will be dependent on how this turn-taking is deeply embedded into conversation it is worth making explicit the importance of turn-taking, and this is something I will also reflect in the terminology I use from here.

1.2.1 *A terminological point*

For the remainder of this thesis the notion of 'turns' will be an important defining feature of conversational participants. For although in the previous chapter I stick to the usual nomenclature of speaker/listener to define the roles of conversational participants, it doesn't sit comfortably with the aim of this project to be a multimodal comparison. From here on I therefore use *in-turner* to denote a conversational participant who is actively speaking, signing, or writing (amongst many other things) and *out-turner* to denote a participant listening, seeing or reading (amongst many other things).[51] Beyond not accounting for other modes of language (such as signed languages, which for my purposes will be structurally equal to any other paradigm types of conversation should they possess the requisite properties), it is also the case that describing a participant as a 'listener' also doesn't do justice to the complexity and variation of tasks being performed by an out-turner in anticipation of becoming an in-turner. Switching to this terminology allows for easier comparison of like-for-like using different modes, but it also pays respect to an important point I want to continuously emphasise; interlocutors in a conversation are generally not performing simple roles of 'speaking' or 'listening', they are engaged in a much more

---

be read as the point that there is an interesting sub-class of language use that has as an essential feature turn-taking – the paradigm cases of conversation.

[50] (1974, p. 696)

[51] As will be made clearer in Chapter 4 – this distinction between *in-turn* and *out-turn* is much less clear in the case of a writer/reader as in most cases there is no simultaneous joint-activity between *in-turners* and *out-turners* (at least, it isn't simultaneous in terms of both participants being required to perform communicative duties). When the *in-turner* is writing, the *out-turner* is waiting, yet once the writing has finished the roles don't swap cleanly. When the writer has finished (and sent their text to the *out-turner*), we might think they become *out-turner*, but at this point the text has most often not yet been read.

complex set of tasks (which we will look at in more detail in Section 3). One last terminological point to note is that the space in between conversational turns – that is, when the current in-turner stops speaking and before the out-turner becomes the in-turner – will be referred to as a *turn interval*.

1.3 *A shared language*

It seems obvious to say we can interact without using a language.[52] Pre-linguistic infants and animals are capable of interacting without the use of what we might ordinarily call 'language'. Even once we acquire language we will routinely interact with others without using it, be it a shared glance with a friend in reaction to something someone else has said, or a colleague making an elaborate 'moving a delicate piece of fine bone china towards the mouth with pinkie finger raised'[53] gesture to ask if someone would like a cup of tea, or stumbling through a shop transaction in a place where the language is unfamiliar to the customer. We might wonder then whether these types of interaction are conversations. It might be thought that they involve two or more participants, turn-taking, a degree of synocracy and a type of perspective sharing (if only fleeting). My intuition on examples such as these, however, is that they fall more closely under the category of a type of

[52] I use 'language' quite broadly here. It might be thought that such a statement assumes that there is some *one* language that, for example, all the speakers of English speak. On closer consideration, however, this is possibly a little too simplistic. For it seems to be the case that there isn't one single version of English that all English speakers speak, there are regional dialects, pidgins and creoles. They may all be similar, but there is perhaps no singular language that we could point to and call Ideal English, for example. This point is touched upon by Paul Pietroski in his discussion of his notion of *Slang* in the precis of his *Conjoining Meaning*. (2018b) Pietroski uses this notion for several reasons but perhaps of interest here is how it accommodates the fact that there can be significant regional variation in syntax and semantics even within one broad 'language'. Pietroski cites the example of 'robin' in British English which refers to a bird of the chat type, whereas 'robin' in American English denotes a type of thrush, or how 'biscuits in the lift' in British English would be expressed as 'cookies in the elevator' in American English. So what it is to be a speaker of English is perhaps best expressed as being someone who has acquired one of the numerous Slangs of the English Slang family. If this is correct, then reference to an intuitive sense of 'a shared language', as suggested in (1iii), doesn't quite capture what is required of conversational participants. Their relationship to a shared language is much looser. Indeed, if we consider the speed of development of pidgin and creole languages, we can see how the barrier of not having a 'shared language' is overcome relatively easily even at the community-level. For discussion on pidgin and creoles, see for example (Bickerton, 1984) who discusses the growth from Hawaiian pidgin to Hawaiian creole over the course of a single generation, or (Lefebvre, 2004) for an overview of the features of pidgins and creoles (esp. pp.9-12). That such complex systems of communication develop within a generation (as opposed to the centuries it often takes other 'languages' to develop), seems to suggest that a problem such as a lack of shared 'language' is not one that happens for long, and where two groups may not be users of Slangs that are suitably similar, they will coordinate to resolve this problem by creating and developing a Slang based on the similarities they find.

[53] I was really hoping to find a name for this gesture that summed it up, but beyond 'tea-sipping' gesture, I couldn't find one so this will have to do.

exchange Grice discusses early in Logic and Conversation – cases of "purposive… rational, behaviour"[54] analogous to what he calls "talk exchanges", but of which conversation is simply a special case.[55] The claim that a conversation requires a shared language does meet with at least one very clear counter-example, however; the phenomenon of cross-signing.

Cross-signing occurs when Deaf sign-language users, generally from different countries and with no written or signed-language in common, are able to conduct conversations despite the absence of a common language. Although anecdotally this phenomenon has been known about for some time, it only became known as 'cross-signing' in 2013[56] and is only now starting to receive attention from linguists.[57] Cross-signed conversations are certainly much more typically what we might call 'conversational' in content than, for example, attempting to buy some toothpaste on holiday when one doesn't know the local language. Some of the examples cited in the literature are discussions of arrival times, differences in home countries and Deaf culture in different areas. So in terms of the content of these interactions, they can go far beyond the simple transactional interactions in a shop. On the surface, then, this seems to cast doubt on the notion that conversation requires a shared language.[58] If cross-signers are able to converse and don't have a shared language, then the stipulation that a shared language is necessary for conversation seems too stringent. So we should note that although conversations will most often be conducted using a shared language, it is not essential. Or we might say, if there is no shared language between participants, then what is required is that interlocutors have the requisite abilities to interact.[59] However we wish to classify this feature, though, is not of major consequence for the view developed here. What I take such examples to be illustrative of, however, is the difficulties we can meet when trying to define conversational features. For even features such as having a shared language, which we might expect to be a common feature across conversations isn't necessary.

In summary then, at a most basic level our paradigm type of conversation must have at least two participants who take turns and most often they will use a shared language

[54] (1989b, p.28)
[55] Some of the famous examples Grice gives are of a person helping another to fix a car or mix the ingredients of a cake.
[56] In Bradford *et al*. (2013).
[57] (Zeshan, 2015, 2019; Byun, de Vos, Bradford, Zeshan, & Levinson, 2018)
[58] Or, in Pietroski's terms (see fn.52) we might say here 'knowledge of similar Slangs'.
[59] See (Levinson, 2006, pp.40-42) for more evidence that human interactive abilities are not completely dependent on language.

(though this is not essential). In the next section we will look in more detail at some other features of the conversational environment and how these features require that conversational participants must engage in a set of highly coordinated tasks.

2 Coordination in face-to-face conversation

> [L]anguage use requires continuous coordination. The participants have to coordinate not only on *what* they do but on *when* they do what they do. They accomplish that…by coordinating on the entry times, content and exit times of each phase of their actions on the assumption that the addressees' processing of the current phase is expected to be complete roughly by the initiation of the next phase.
>
> (Clark, 1996, p.91)

Although he specifies only 'language use' in the above passage, Herbert Clark is discussing specifically language use in conversation.[60] Clark's insights into the role of coordination and some of the features of conversation form the basis of the type of view of conversation I develop here. The first task then is to start to unpack the notion of 'conversational coordination'.

2.1 *A simple structure of paradigm face-to-face conversation*

Suppose we have the following extract of a face-to-face conversation between A and B.

(2)

01: A: Do you know the way to San Jose?

02: B: No, I've been away too long.

A rough chronological depiction of what A and B are up to at each turn (01, 02) is given in Table 1.

(3) Table 1[61]

[60] So for example, there would appear to be no such coordination required in language use such as letter writing.

[61] Note: during in-turn we may well also include a task group such as 'respond to out-turner feedback' in parentheses, however, though this is likely, it is not strictly necessary.

	Turn marker	
	01	02
Person	Necessary conversational task	
A	Making utterance	Comprehending B's utterance Formulating response——————>
B	Comprehending A's utterance Formulating response——————	———>
	Time ———>	

Table 1: A simple structure of face-to-face conversation and its necessary tasks.

In-turn
Out-turn

It might seem obvious that when A speaks B is listening and formulating a response, and vice-versa, but it is worth being explicit about the concurrence of these tasks as it will mark an important distinction between face-to-face and digital conversations that will be discussed in Chapters 3 and 4. The type of structure outlined in (3) is necessary for a successful conversational exchange be it face-to-face, on the telephone, or using video-call technology. Essential to their efficacy is that when the *in-turner* speaks or signs, the *out-turner* attends to what they utter or sign. At any one time, then, all participants in a conversation will be fulfilling some specific coordinative role, their focus will be, in part, on the output of the *in-turner* and their role will dictate which set of other tasks they will need to be carrying out. As such, there is a strict symmetry of attention that all participants must direct towards what is being uttered/signed in the conversation. Without this coordination face-to-face communication would fail, this is due, in part, to some of the features of face-to-face conversation noted by Clark.

2.2 *Clark's features of conversation*

Clark[62] adapts from his work with Susan Brennan[63] 10 features of face-to-face conversation. These can be separated into three groups – immediacy, medium and control. I will discuss each group in turn.

2.2.1 *Immediacy features*

[62] (1996, pp. 9–11)
[63] (Clark & Brennan, 1991)

Clark notes that a face-to-face conversation generally has the following two features;

Copresence: the participants share the same physical environment.

Instantaneity: the participants perceive each other's actions with no perceptible delay.

It will also have either one of or both of the following;

Visibility: the participants can see each other.

Audibility: the participants can hear each other.

What Clark is discussing here are features of what he calls *basic* conversation.[64] The idea of conversation being *basic* is similar to the idea suggested in Chapter 1 (§1.2) that the paradigm type of conversation is the conversation universal to human societies. To be able to participate in such conversation requires no special technology and any skills required are most often-developed without formal training (unlike writing skills or cross-examining skills, for example) or other forms of training (such as might be required for effectively giving speeches). The training most of us receive in how to have a conversation starts around the time we are born and we're just left to learn on the job.[65] So with this type of conversation in mind, the above features seem self-explanatory.

For our purposes here we need not assume all the types of exchange we might regard as paradigm types of conversation have all of these features. Indeed, although they often will have all of them, it is certainly not necessary that they do. In terms of the types of activities I suggested in Chapter 1 might be considered paradigm cases of conversation, there is no obvious reason we should exclude a telephone or video call. As such we might think that copresence isn't necessary for a paradigm type of conversation, nor visibility in the case of a telephone call, one or the other of these two properties will suffice in these respective cases. And anyone familiar with video-calling technology's occasional defects might be familiar with the occasional perceptible delays using such media, and so *instantaneity* as described above might not be quite accurate. In some cases of spoken

[64] Following Fillmore's description of conversation as being 'basic' (Chapter1, §1).
[65] It's worth noting that development of awareness of turn-taking has been observed at 3 months (Hilbrink, Gattis, & Levinson, 2015), as well as sensitivity to interpersonal timings (Striano, Henning, & Stahl, 2006), and sensitivity to facial expressions (Striano & Liszkowski, 2005). (Moore & Meltzoff, 1977) suggest that 12-21 day olds can mimic facial expressions. So even though pre-linguistic humans are not fully conversational, they are developing the abilities to conversationally interact from a very young age.

conversation one or more participants might not be able to see others (perhaps due to problems with eyesight, or the lighting, or not be facing them) and so visibility is not essential. In many signed conversations, too, it will likely be that interlocutors can't hear each other and so audibility need not be a feature.

2.2.2 *Medium-specific features*

Clarke suggests in the second group that face-to-face conversation has the following three features;

Evanescence: the medium fades quickly.

Recordlessness: there is (usually) no record or artefact of actions.

Simultaneity: participants can produce and receive at once and simultaneously.

The implications of this medium-specific group of features will be important when we consider the requirements placed on interlocutors in a paradigm type of conversation. For when we factor in these aspects of conversation, we start to understand some of the necessary requirements that need to be accounted for in order for conversation to function against this backdrop.

Whether it be a spoken or signed utterance, the visual and auditory properties of the language produced will disappear quickly.[66] Once something is said, for example, its

[66] Though it is worth considering the role of working memory when considering quite how evanescent language is, for in some sense at least, it may be that some version of the language produced doesn't fade quite as quickly as suggested here. Although the sounds or signs themselves might fade quickly, on one prominent theory of working memory, there is potentially at least a delay of 1.5-2 seconds before sounds fade from the phonological loop component of working memory, this is potentially even longer if we accept the existence of subvocal rehearsal. Alan Baddeley and Graham Hitch (1974) propose a multi-component model of working memory that consists of a controlling central executive mechanism and two "slave" systems – the phonological loop and the visuospatial sketchpad. These two slave systems are each concerned with the processing and storage of material from the relevant domains (as picked out by their respective names). The Baddeley-Hitch model has more recently been adapted to include a fourth component - the episodic buffer. See for example (Baddeley, 2000). This fourth component is taken to be multi-modal in that it is capable of integrating information from various sources such as the long-term memory, the phonological loop and the visuospatial sketchpad, yet it remains under the control of the central executive, which can draw upon this temporary information store as required. Baddeley suggests the phonological loop component of working memory consists of two sub-components: a phonological store and articulatory rehearsal/control process. (Baddeley, 1986, 1990, pp. 71–87, 2003) The phonological store holds speech-based information (If the linguistic input is, for example, written, then the articulatory control process converts the written string into a phonological code and place it into the phonological store. It helps to try this yourself. Look at the following sequence of numbers, close

existence as a piece of sound is almost instantaneously over. It is also the case that should an out-turner not hear or see a particular part of some utterance, then unless she is able to piece together a prediction from other parts of the in-turners utterance, then she will most often have no record of it to return to for clarification. In such a circumstance, then, to be able to recover this missing piece, the out-turner would need to initiate a process of repair whereby she requests the in-turner repeat or clarify the missed utterance.[67] Such resources are generally only to be used sparingly, however. It seems obvious to suggest that processes of repair could only really be used sparingly as overuse would lead to circuitous discussion.[68]

So taken all together, what the medium-specific features above make clear is that in order for spoken/signed conversation to function, an out-turner is required to attend to the

---

your eyes and try to remember the sequence. 8 4 9 2 0 7 2. Chances are you "vocalise" the numbers internally) and the assumption is that memories in here fade and become irretrievable at around 1.5-2 seconds. This degradation, however, can be delayed by an articulatory rehearsal process in which the phonological information is rehearsed subvocally. The idea being that this rehearsal acts as something like a refresh of the information. (Criticism of the notion of subvocal rehearsal, and in turn the phonological loop, can be found in Nairne (2002). One of the arguments Baddeley puts forward for the existence of the phonological loop can be found in (Gathercole & Baddeley, 1993). They suggest that how we can return to a correct interpretation of a *garden path* sentence is indicative of its existence. The example they use is:

(X1) i. I saw that gasoline can explode.
ii. And a brand new can it was too.

Let's say that Susan utters (X1) to Alan. The assumption is that when Alan hears (X1i), he will initially process 'can' as a verb and 'that' as a conjunction. Let's call this Interpretation 1. Once he hears (X1ii), though, it will likely become clear that Interpretation 1 is incorrect. It should become clear that rather than 'that' being used as a conjunction in (X1i), it is actually being used as a pronoun. As such, 'can' will need reclassifying as a noun, and this reparsing would then make (X1i) and (X1ii) consistent (Interpretation 2). (For more on this, Baars (1988, pp. 93–95) provides an amusing and enlightening illustration of how such language processing occurs using a global workspace model.) One of the things this suggests, according to Gathercole and Baddeley, is that although most language processing occurs 'online' in real-time, the ability to deal with garden path sentences requires something more. In a spoken exchange in order to reprocess (X1ii), the listener would need access to a verbatim representation of the sentence (X1i). However, she is hindered in that the physical manifestation of (X1i) lasts not much longer than the time it takes the utterer to say it. That she can do it, the idea goes, suggests that this verbatim representation is stored in her working memory. (I should note, although I do agree in principle with this type of example, this specific example only really works on paper – although we represent the two 'that's which create the supposed ambiguity in (X1) the same way in written English, in most cases the pronunciation is different. The conjunction is pronounced /ðət/ whereas the pronoun /ðat/ or /ðæt/. As such, I think it unlikely such ambiguity would arise in the case of (X1) were it a spoken example, and the force of this example is reduced if we treat it as a written example, for there is no requirement that we store anything in our working memory – we can simply reread the sentence.)

[67] Specifically, then, an other-initiated repair. (Schegloff, Jefferson, & Sacks, 1977).

[68] And, indeed, although repair is common, Enfield (2017, Ch.7) notes that on average such repair initiating expressions as 'huh?', 'what?' or 'sorry…?' occur approximately every 84 seconds in a conversation. As noted later, conversational turns average around 2secs, so this would be an average of approximately one such repair every 40 turns or so.

linguistic output of the in-turner just as the in-turner makes her utterances, but added to this as the in-turner makes those utterances she too must attend to the attentive state of the out-turner(s). Clark describes this as follows;

> If Roger is to succeed in telling Nina something, he must make sure she is trying to attend to his sounds *at the very instant* he is articulating them. Executing behaviours to be attended to and attending to those behaviors, then are participatory acts: Roger cannot do his part without Nina doing hers, and vice versa.[69]

Recall Chapter 1(§2); there it was suggested that interactivity and continuity of attention are some of the properties of the paradigm type of conversation being characterised, and now we see start to see some of the detail as to why this is so. The interactivity of such exchanges is not merely at the level of the most obvious structure of the exchange, by which I mean it is not interactive merely because each turn follows a prior one; this interactivity is continuous due to the medium-specific features of evanescence, and recordlessness (and simultaneity to a lesser extent). Therefore, as an utterance is being made the out-turner must be synchronised with the in-turner.

Consider how participants in a cooperative game of throw and catch are required to synchronise their behaviours in order to successfully throw and catch a ball. To be optimally successful we might expect that the catcher makes clear to the thrower her readiness to catch, which the thrower then reads. The thrower then tosses the ball in the rough direction of the catcher giving cues in her bodily movements prior to the throw as to the approximate direction she intends the ball to travel. The catcher in order to be successful in catching the ball will need to read both the body language of the thrower and the flight of the ball and position herself and her hands accordingly. If successful, the roles then switch and the same thing plays out in reverse. All of this behaviour, I think, is continuous, synchronised and interactive. The various moves of each player are interdependent. And I think this is close in nature to what occurs in the paradigm type of conversation. An out-turner in a verbal conversation, for example, must be performing tasks relevant to the actual listening (for example, comprehending and contextualising, many more such tasks are discussed in Section 3), but she also needs to provide feedback to the in-turner; she must communicate that she is paying attention to the in-turner's utterances

[69] (1996, p.275)

in harmony with them occurring. And to ensure the communication is successful, the in-turner must also pay heed to this feedback.

The most economical of the responses an out-turner can use to communicate their continued attention is an eye gaze,[70] but there are other methods we naturally might use be it a smile, a nod, or a thoughtful furrowed brow.[71] [72] Face-to-face conversation without this simultaneity of response can be disconcerting. For example, as most philosophers are acutely aware, sometimes the minutiae of some matter are not as interesting to anyone else as it is to the person elucidating it and talking about it to someone not remotely interested can induce a glassy-eyed vacancy. In such a discussion, the out-turner is likely (intentionally or not) sending feedback to the in-turner that her commitment to the tasks required of conversation has ceased – she is no longer listening. So often the cues we take as in-turners from our out-turn interlocutors guide us in how we proceed both with our turn, and how we approach our commitment to the interaction. And as Clark says, Roger can't do his part without Nina doing hers, and vice versa.

These medium-specific features, then, are salient to how we are required to interactively coordinate ourselves when participating in a conversation. The sounds or signs we receive will vanish almost instantly and without record. As out-turners, in many cases, we also need to communicate to the in-turner that we are engaged with what is being uttered. In Section 3 we look in much greater detail at some of the data on the type of comprehension tasks out-turners necessarily must perform due to these medium-specific features, but next let's consider Clark's final group of features – the control group.

2.2.3 *Control features*

The final group of features Clark discusses are the *control* group (specifically about who controls what).

> **Extemporaneity:** The participants formulate and execute their actions extemporaneously, in realtime.

---

[70] Clark cites (Argyle & Cook, 1976) for evidence, we also see evidence of the role of mutual gaze in pre-schooler conversation in (Krantz, George, & Hursh, 1983), suggesting this is developed early.
[71] A note of caution however, what I always assumed to be a 'thoughtful' furrowed brow appears is also very often taken to express confusion. See (Domaneschi, Passarelli, & Chiorri, 2017, p. 294, particularly fig.2.)
[72] As we shall see in §3.1.5, when looking at interactive alignment, this process of interaction goes even further than these most obvious types of example.

Self-determination: The participants determine for themselves what actions to take when.

Self-expression: The participants take actions as themselves.

Clark says of these:

> In face-to-face conversation, the participants are in full control. They speak for themselves, jointly determine who says what when, and formulate their utterances as they go. In other settings, the participants are restricted in what they can say when. The church, for example, determines the wording of many prayers and responses. In fictional settings, speakers and writers only make as if they are taking certain actions — Gielgud is only play-acting his role as Hamlet — and that alters what they do and how they are understood. And in mediated settings, there are really two communications.[73]

I will discuss the features of self-determination and self-expression in more detail when looking again at synocracy in conversation in Section 4.2.2, but for now I want to focus most particularly on extemporaneity. Because it is this feature, together with the features of the medium-specific group (evanescence, recordlessness and simultaneity), that determines some of what is particularly unique about the paradigm cases of conversation. Later in the *Using Language* Clark states that (emphasis mine);

> Conversations… are *purposive* but *unplanned*. People achieve most of what they do by means of joint projects, both large and small, in which they establish and carry out joint purposes they are willing and able to commit to. To complete these, they have to work at the level of minimal joint projects,[74] for it is with these that they

---

[73] (Clark, 1996, pp.10-11)

[74] One of Clark's overall arguments is that fundamentally a conversation (and by extension language use more generally) is a series of projects interlocutors propose and complete collaboratively. These, then are joint projects. And for Clark these joint projects will generally take the form of adjacency pairs (See (Schegloff & Sacks, 1973) for initial discussion of adjacency pairs.) in their most basic form and stretch up to a whole conversation at its most complex. For example, a paradigmatic example of a basic type of joint project might be a question-answer pair such as (2) above, or (x2):

(x2)

01: A: How much?
02: B: No more than a fiver.

What happens in (x2) is not simply the case of A says something and B understands it. At (x2-01), A proposes that B provides information to A, and at (x2-02) B takes up this proposal and completes the joint project by providing the information. So although it is essential for the successful completion of the joint project that B understands the meaning of A's utterance, the joint project's completion is dependent on B understanding and providing the type of response required. In Clark's terms, (x2)

> negotiate broader purposes and complete extended joint projects. What emerges are sections and, ultimately, the entire conversation itself. *Conversations look planned and goal-oriented only in retrospect. In reality, they are created opportunistically piece by piece as the participants negotiate joint purposes and then fulfill them*. Let me call this the *opportunistic view* of conversation.[75]

Later;

> [C]onversations are controlled jointly. Each turn is shaped by all the participants as they engineer the selection of the current speaker, and influence the course and length of each turn. [76]

And later;

> [A conversation will often look] orderly even though each step of the way was achieved locally and opportunistically… Much of the structure of conversations is really an emergent orderliness. Although the participants appear to follow rules in turn taking, they are merely trying to succeed in contributing to the conversation.[77]

What Clark says above seems to be right, for although conversations might appear in hindsight to look as if they were planned and ordered, the way that they develop is opportunistic (and this too echoes Goffman's description of conversation as *spontaneous* discussed in Chapter 1 §3.2). It is, in a sense, dependent on continuous reactions to the actions of others.[78] And not only this, it is placed under the joint control of the participants. Recall in Chapter 1 the notion of perspective sharing and interdependence of contributions

---

would be a type of minimal joint project, and in conversation interlocutors will embed, chain and sequence these minimal joint projects into extended joint projects (with the entirety of the conversation being a joint project composed of all these minimal and extended joint projects).

[75] (1996, p.319)

[76] (1996, p.331)

[77] (1996, p.351). This last point, too, chimes with some of what Searle says in his criticism of the project of conversation analysis (1992a). Searle argues that talk of 'turn-taking rules' is misguided because these so-called 'rules' are not constitutive rules. The argument being that what conversation analysts observe as 'rules' are not rules that conversational participants can or can't follow, they must behave in the way that these supposed 'rules' would instruct them. And if interlocutors were adhering to rules, it doesn't have any bearing on their conversational behaviour anyway. Emanuel Schegloff (1992) responds to Searle that it would be possible to replace 'rule' with 'practice'. Searle's response (1992b) is that the force of the problem remains – description of patterns is not the same as proving explanatory force.

[78] Recall too from chapter 1 Goldberg's description of a great conversation as being akin to a piece of co-authored performance art. (Goldberg, 2020) For interlocutors to 'author' the conversation as they do collaboratively requires that they must work opportunistically.

was also invoked to discuss how the contribution of A to a conversation might provoke novel and unexpected thoughts in B. So although both may enter the conversation with some sense of individual purpose - maybe to ask a particular question or to arrange some event – it is highly implausible that they will have considered and planned many of their utterances in advance (and almost imperceptible that they could have planned all of them). It is in the very nature of conversation that if Kim is responding to an utterance made by Thurston and Thurston to Kim, that unless they have some foresighted access to the precise future thoughts of each other, their responses will be extemporaneously developed. But not only will the responses be extemporaneous, they will most often also be reflective of the preceding talk of an interlocutor. Further to this, an interlocutor can only respond to these once they have been produced (or to be more precise, as they are being produced – for as we shall see, out-turners do usually have to predict at least some of what an in-turner is going to say in order to respond in a timely manner). Contrast this with the example of cross-examination first discussed in Chapter 1 (§2.3.1).

(4)

01 Q. Did he tell you he had seen a pickup truck?

02 A. Yes.

03 Q. All right. Did he describe the pickup truck to you?

04 A. Yes. He said it was a dark-colored –

05 Q. Hold it a second. Did he describe it? Yes or no.[79]

On some level, an exchange such as (4) requires some degree of extemporaneity, for example at (4-05) Q's contribution to the interaction is an interruption that responds directly to the way A at (4-04) responds to the question at (4-03). But we should expect that for the cross-examiner to be competently fulfilling their role as a cross-examiner, much of what they say will be planned to a much greater extent than in a paradigm case of conversation. They will have a line of questioning prepared in order to extract the precise information they wish to from the cross-examinee. Not only this but, as discussed in Chapter 1, in scenarios such as the courtroom (or the job interview), by the very conventions of such institutions one party to the exchange is afforded autocratic control

[79] (Hansen 2008, p.1400).

over the direction of discussion. As we see in the interruption at (4-05), in the above exchange A is not afforded much self-determination nor self-expression in the exchange.[80] Now of course in a trivial sense we might say A *is* afforded this, she could refuse to answer the questions, or she could even respond in a nonsensical manner, though to do so could result in penalty such as contempt of court (or in the case of a job interview – failure to be considered for the position). So although trivially speaking she is free to determine for herself what actions to take and free to express herself as she chooses, relative to most conversational types of interaction, these freedoms are severely restricted (and we shall see the importance of this in Section 4.2.2). As such, it is this set of features which start to help us see important differences between interactions such as a cross-examination and our paradigm cases of conversation. Generally conversations will be extemporaneous and the participants will determine their own actions and how to express themselves.

2.2.4 *Summary of Clark's features of paradigm conversation*

The picture of face-to-face conversation we get then when we consider Clark's features is one of a fast moving and interchanging series of concurrent tasks. Instantaneity and simultaneity make it so that such interaction is a continuous process of feedback whether a person is in the role of in-turner or out-turner. In the role of out-turner a participant must convey their commitment to the conversation and the in-turner must monitor this feedback. This is necessary due to the evanescence and recordlessness of the utterances produced. As such, conversation requires that as an in-turner is making an utterance an out-turner must attend to it in real time before it disappears (and so to not hear or see it as it occurs will mean that it fails to function as a tool of communication). Finally, the nature of conversation as being driven by its participants' self-determined and self-expressed contributions means that responding contributions will need to reflect the direction set by the content of previous utterances. As the participants in a paradigm type of conversation are in joint control of the proceedings contributions will most often be formulated and executed extemporaneously. This extemporaneousness means, therefore, that conversation is opportunistic in nature. Now of course, there is something obvious that these features contribute to; that is the idea that a paradigm case of conversation is highly personalised to the conversational participants. More often than not, a conversation is conducted with a

[80] And indeed, one might argue that even the cross-examiner isn't afforded this – the court itself has supreme authority over all of the interactions in a case and can halt the cross-examiner should a judge see fit to.

single audience - all members of which are the participants of the conversation. When this is the case, each turn of a conversation is generally directed towards, and for the benefit of, the other conversational participants, and how an in-turner presents an utterance will be based on the utterances and feedback of those other participants.[81]

This all points towards the fact that face-to-face conversation is quick and highly improvised. As we see in more detail in Section 3, to sustain a conversation there's a lot to do and very little time in which to do it. So the observations of Merleau-Ponty of conversation as a *being-shared-by-two* and of Goffman's *unio mystico*, are, I suggest, in part explained by the requirements conversations place on interlocutors. If I were engaged in a conversation with you, then it is imperative that as you speak I must at the very minimum listen, let you know that I'm listening, and prepare to respond to what you say, and when I speak you do the same. We are both continuously attending to each other and using each other's contributions as a means of helping us develop our own contributions to the conversation. And this explains, in part, how sometimes conversation can be so engrossing – to participate in one is cognitively demanding. The features of evanescence, recordlessness, instantaneity and extemporaneousness mean we must stay *tuned-in* to a conversation almost continuously.[82]

2.3 *Coordination: Process and content*

On what has been said so far, we might have some idea why Clark suggests that the features of face-to-face conversation require participants to engage with the overall task of the conversation at both the level of content and the level of processes. The content being the topics of the conversation, the processes being the tasks required in order to deal with the particular features of any face-to-face interaction. Although we can consider these levels of coordination in isolation, they are nonetheless necessarily dependent on each other. Without a topic (content) there is nothing to talk about in a sensible manner and so no way

[81] Contrast this again with some other types of interaction and often the audience dynamics are very different. In a cross-examination the primary audience is the court and not the cross-examinee. Think also perhaps of a TV chat show wherein the set-up is of two people in conversation with each other. Anyone who has watched these will know, however, that these types of conversation are played for the purposes of a watching audience. So although many of the features Clark notes of face-to-face conversation might still hold, the conversation is being performed for the benefit of people not participating in the exchange.

[82] This also recalls something Merleau-Ponty suggests later; "I am freed from myself in the present dialogue…Only *après coup* – when I have withdrawn from the dialogue and I am remembering it – can I reintegrate it into my life, turn it into an episode of my private history, and only then does the other person return to his absence." (2012, p.371)

of coordinating processes. Without coordination of process, there is no way to comprehend the topic of conversation (so at the most obvious level, if I don't listen to your utterance, I of course won't be able to respond to it).

On Clark's view, in order to do coordinate in such a way, interactants will enter into a series of joint projects which link to form extended joint projects, which in turn will compose the overall joint project we might call 'a conversation'. To complete these joint projects we must coordinate at the level of content (that is to say, the participants must in some way have their focus of talk on the topic relating to some joint project), but in order to coordinate at the content level we must also coordinate at the level of the processes. In the next section, then, we shall look at some of the processes involved in face-to-face conversation, many of which are made necessary by the medium-specific features and control features of face-to-face interaction. The direction this is going in, then, will be to consider in Section 4 how this systematic coordination is reliant on an agent-led cooperation. That is to say, in order to achieve this level of coordination, speakers in a conversation must actively intend to cooperate with an interlocutor, even if just minimally at the level of process. And since coordination of process and content are co-dependent, the picture is one in which even if interlocutors are in disagreement at the level of content (in, for example, an argument), they are nevertheless engaged in a form of highly cooperative joint action.

3 Coordinating processes

For Clark, salient to understanding how language use works is understanding how interlocutors coordinate their behaviour in terms of content and processes.[83] Clark takes 'content' to refer to what participants intend to do and *processes* relates to the physical and mental systems recruited to carry out those intentions.[84] Although Clark doesn't state it explicitly, we might expect from this that in our paradigm cases of face-to-face conversation many of the types of processes undertaken will remain relatively stable both within a single conversation and even when moving from conversation to conversation. On the other hand, the specific content will vary greatly within a conversation and from conversation to conversation.

[83] (Clark, 1996, see Ch.3 and Ch.7)
[84] It is worth noting that coordination of content and process need not relate solely to language use, dance partners will coordinate in content and process, so too a group of people trying to move a piano, or an orchestra playing a concerto.

In the role of out-turner or in-turner, much of what is required at the process level will remain the same whoever the participants are, and whatever they happen to be discussing. Put simply; in a paradigm case of verbal conversation an in-turner is required to perform all the *process* tasks related broadly to speaking (vocalising, conveying meaning, giving cues as to turn length, checking for feedback etc…), and the out-turner is required to carry-out all of the tasks relating broadly to listening (comprehending, predicting, giving feedback, preparing to speak etc…). On the other hand, the *content* of a conversation will be highly specific to particular conversations and, indeed, will even be transient within a conversation as the joint projects and shared goals of a conversation continuously progress. The content, then, is dependent on the specific context of a particular conversation; the intentions of an individual, the shared intentions of the interlocutors and the common ground.[85] So the content will vary significantly even within a single conversation and when moving from conversation to conversation. Now of course, these two levels of coordination are, as mentioned above, co-dependent - it isn't possible to coordinate on one but not the other - but it *is* possible to consider them in isolation. My focus for the remainder of this dissertation, then, will primarily be on the coordination of processes – the things participants in our paradigm type of conversation are required to do in any such type of conversation. The co-dependence of these aspects of conversation will naturally mean that this isn't to say that content doesn't feature, it does, but rather the point is that in comparing the paradigm cases to other cases, the stability of processes across conversations make this a particularly interesting point of comparison.

[85] Of Clark's two levels of coordination, the common ground is essential to understanding coordination of content. As my focus in this project is weighted towards coordination of process and the effects this has on content, I have chosen to avoid detailed discussion of the common ground. In a fuller account of paradigm conversational exchanges that placed equal emphasis on content and process, the common ground would be a key component. And indeed, the common ground would be worthy of its own dedicated chapter. It is worth recapitulating the basic idea for the sake of background consideration. Simply speaking, we might say that the common ground is the basic context that conversational exchanges take place within. So the conversational common ground is the relevant information that conversational partners draw upon when conversing; that is, it is the information they take to be shared. It has a rich history in recent philosophy through David Lewis's talk of common knowledge and conversational scorekeeping ([1969] 2002; 1979), Schiffer's on mutual knowledge (1972), Aumann's common knowledge (1976), perhaps most famously in Stalnaker work on common ground (for example, 1978[1999]), and more recently in Bach's broad context (2012), and the Stalnakerian common ground of von Fintel (2008), Green (2009; 2017) and Stokke (2018). It is also resonates through much of the work of psychology and psycholinguistics I draw upon in this and the following chapters, for example Clark and Brennan (1991), Clark (1996, Ch.4, Ch8), Pickering & Garrod (2004), Horton & Gerrig (2005, 2016), Enfield (2006), Richardson, Dale, & Kirkham (2007), Brown-Schmidt & Duff (2016). The common ground, then, is a major omission, though for the sake of clarity and brevity, one I think necessary for now. I will make passing references in footnotes to some areas in which the common ground is salient.

In this section, then, I consider in more detail some of these stable process level tasks by looking at some recent work in psycholinguistics. We shall see that some of the consequences of Clark's features of face-to-face conversation are that in paradigm cases of conversation participants are required to commit to a continuous and fast-moving set of necessary process tasks in order to sustain the conversation. In the following section I will hypothesise that the required commitment to such tasks gives us good grounds to consider the fundamental role of cooperation in these types of linguistic activity.

3.1 *Response speeds*[86]

3.1.1 *The one second window*

The first interesting observation is what Gail Jefferson (1989) refers to as the "standard maximum silence" and which Nick Enfield (2017) refers to as *the one second window*.[87] Roughly speaking, this one second window is the approximate amount of silence afforded by the most recent speaker in-between conversational turns before they either prompt again and/or begin to attach meaning to the silence. Take the following example, for instance.

(5)

01:	R:	What about coming here on the way?
02:	(.)	
03:	R:	Or doesn't that give you enough time?
04:	C:	Well no I'm supervising here.[88]

Here (.) represents one second of silence. What we can see from this example is how R takes C's silence at (5-02) to signify a dispreferred response to the question at (5-01), and so hypothesises why this might be the case at (5-03).

[86] I don't critically appraise the ecological validity or methodology of the studies mentioned here. And although I present some of the headline figures and results, I don't discuss the experiments in much depth. The specific timings themselves are not what I really take to be of interest, though they are astonishing, rather it is the overall picture of the type of commitments participants in our paradigm case of conversation must make in order for conversation to proceed in the manner it does.

[87] See Enfield (2017 Ch.1 & 4) and Levinson (2016, p. 9) for more on this.

[88] (Levinson, 1983, p. 335)

The type of occurrence such as seen in (5) leads some to hypothesise that responses taking longer than one second can signal some trouble in the interaction. Jefferson suggests one second as the average such time after studying 1000 examples in English. Clark[89] suggests this may be variable between languages, but a similar average time was observed in a study by Stivers *et al.* (2009) of 10 different languages. And more recently studies of the one second window in English suggest that this window is potentially even shorter, for example;

> In the absence of other vocal or visual cues, there is clearly a critical range for the effect of the inter-turn silence after requests: perceived willingness is consistently higher before 500 ms, begins to drop after 600 ms, and then clearly and significantly steps down from 700 to 800 ms. After 900 ms, in the absence of other cues, there appears to be a simple floor effect. (Roberts & Francis, 2013)

This accords with results in Kendrick & Torreira (2015) which found that a gap of over 700ms in telephone conversation indicated that at a dispreferred response (i.e. a rejection as opposed to acceptance) was more likely. Whatever the precise time may be before silence becomes a sign of unwillingness, it isn't long. Whether this one second window is an innate feature of interactive language use, or rather an example of a wide-ranging conventional aspect to language is not certain, thought the former, we might expect, is most likely;[90] what we can take from this, however, is that cross-linguistically it appears that there is an expected limit to response speeds, and this limit itself is important. The one second window imposes on interlocutors a time-limit in which to respond, and with that time limit comes added pressure to focus on the conversation because language production itself is subject to time-limitation.

3.1.2 *Language production latency*

Although up to a second might be afforded for to out-turners to respond, this usually this isn't all taken. In a wide study of three corpora in different languages (Dutch, Swedish, Scottish English), Heldner & Edlund (2010) observed that 55–59% of all turn intervals were either not noticeable gaps, or were overlaps.[91] Stivers et al. (2009) also found that mode

[89] (1996, pp.268-269). Clark refers to it as the *one second limit*.
[90] Cf; "Clearly, something generalized about human perceptual processing generates an observable-reportable phenomenon of silence as indicative of trouble in conversation." (Roberts, Margutti, & Takano, 2011, p. 350)
[91] (Heldner & Edlund, 2010, p. 563). By 'overlaps' they refer to a response starting before the in-turner has finished speaking. By 'noticeable gaps', at the time Heldner & Edlund took this to be

average response speeds of around 0-200 ms occur across all 10 languages (with varied distributions for different languages).[92] If the *average* response speed is around 200 ms, then on average interactants respond to a previous turn at close to the limit of human reaction speed.[93] Although this 200 ms is fast, it is even more notable when contrasted with observations of the latency involved with speech planning. For example, in meta-analysis of studies that look at timed picture naming, Indefrey & Levelt (2004) and Indefrey (2011) suggest planning a single word takes approximately 600 ms when participants are primed.[94] Bates *et al*. (2003), which looked at timed picture naming across seven languages, suggests this increases to 1000 ms without priming. Also using picture naming as a basis Griffin & Bock, (2000) and Schnur, Costa, & Caramazza, (2006) suggest the planning of a simple clause sentence (such as 'the girl jumps' or 'the man sneezes') takes approximately 1500 ms. And one last headline figure to mention, Levinson (2016) notes that average turn duration is approximately 2000 ms. What we shall see next, then, is that in order for interlocutors in a paradigm case of conversation to respond in the way they typically do, the processes required for comprehension and production must overlap.

3.1.3 *Prediction and overlap*

For ease of reference, I collate some of the data from Section 3.1.2 into Table 2.

approximately 200ms as per (Walker & Trimboli, 1982), since then, however, following Heldner's own (2011) research, he has suggested approximately 120ms is about the length of acoustic silence required before a gap is noticed.

[92] It should be noted that Stivers *et al*. used video to compile the data and it was allowed that a gesture such as a nod or the inhalation of breath in preparation to speak was considered the start of the next turn.

[93] For context on quite how fast this is, IAAF guidelines for a 100m race don't deem a false-start to be one in which the sprinter leaves their block before the starter pistol fires, but rather if they leave within 100ms of the starter pistol firing. Enfield (2017) notes that 200ms is about the time it takes to blink an eye, Levinson (2016) that this is the about the time it takes to utter a syllable.

[94] In Indefrey & Levelt (2004, see Table 3), they estimate the breakdown of this 600 ms as follows: Conceptual preparation 175 ms; Lemma retrieval 75 ms; Phonological code retrieval 80 ms; Syllabification 125 ms; Phonetic encoding (until initiation of articulation) 145 ms.

(6) Table 2

1	2	3	4	5
Avg turn	Avg response time	Planning (single word primed)	Available speech before planning	Overlap
2000 ms	200 ms	600 ms	1600 ms	400 ms
		Planning (simple clause sentence)	Available speech before planning	Overlap
		1500 ms	700 ms	1300 ms

Table showing average turn duration (1) and average response duration (2) in comparison to planning (3) with resultant duration of speech available to out-turner before commencement of planning begins (4), and amount of time out-turner planning overlaps with current in-turner's speech (5)

What we see from Table 2, then, is that based on an average turn duration and average response time, even the out-turner's planning of a single word response will need to begin 400 ms before the in-turner completes her turn, and this rises to 1300 ms in order to prepare a simple single clause sentence. This leads some to propose the *early-planning hypothesis* (for example, Barthel & Levinson, 2020; Levinson & Torreira, 2015) – speakers need to start planning their response as early as possible. One such example of this view can be found in Pickering & Garrod (2013). There it is proposed that language production and understanding in face-to-face conversations are tightly linked; "[interlocutors] do not simply transmit messages to each other in turn but rather negotiate the form and meaning of expressions they use by inter-weaving their contributions."[95] Pickering and Garrod illustrate this interweaving using the following example;[96]

(7)

01: A: I'm afraid I burnt the kitchen ceiling.

02: B: But have you…

03: A: burned myself? Fortunately not.

[95] (2013, p.330). It is worth mentioning that Pickering and Garrod take this is interweaving approach to also work at the level of the individual. They base this on Susan Hurley's Shared Circuits Model (2008) and her rejection of the separation of what she calls "the classical sandwich" (2001), in which perception and action are the bread and cognition the filling. Hurley argues that perception and action are mutually and symmetrically interdependent.

[96] From (Gregoromichelaki et al., 2011).

So B begins to ask a question (7-02), but A interrupts (7-03) to both complete and answer the question. In this example, then, B alone doesn't individually encode in sound the question 'But have you burned [yourself]?', it is, rather, jointly encoded by B and A over the course of (7-02/03). This is an example of what Clark (1996, p.238) classifies as a *midutterance action* in which we see a *collaborative completion*. One of Clark's own examples of this is (8), wherein Herb is wondering where another umbrella might be.

(8)

01: Herb: Where's the other –

02: Eve: On the back shelf.

03: Herb: Good.

This example differs slightly in that Eve at (8-03) *doesn't* encode in sound the remainder of Herb's question, though what she does do is answer it in a way acceptable to Herb at (8-03), and so it is taken by both that they took Herb to be asking; 'Where's the other umbrella?'. Cases such as these, then, illustrate what we might expect if we accept the early planning hypothesis – a part of the role of an out-turner is to predict what the in-turner is going to utter.

3.1.4 *Prediction and perspective aligning*

When we consider together the overlaps in exchanges such as (7) and (8) and the difference in response speeds in relation to language production latency, then we see evidence of how prediction is essential for the functioning of face-to-face conversation. The time constraints of the one second window, how long it takes to produce an utterance, and how quickly interlocutors respond on average make it seem likely that out-turners must try to *predict* what an interlocutor might say, or the speech act they might be performing, as very often they won't have a complete utterance to respond to. Not only must an out-turner predict *what* the in-turner will say, they must also predict when they will finish saying it.[97] What this seems to suggest, then, is that one of the key groups of cognitive tasks an out-turner is required to complete is centred around not simply listening to what an in-turner *has* uttered (though this is essential), but also to using what has been uttered to forecast what an interlocutor is *going* to utter, and when they will finish uttering it. This, then, perhaps

[97] See (Garrod & Pickering, 2015) for more on the prediction of turn duration.

seems to overlap somewhat with what Merleau-Ponty was referring to (see Ch. 1, §3) when talking about *perspective sharing*.[98]

There seems to be an aspect to paradigm conversations in which both the out-turner and in-turner must all be arriving at the same (or similar) thoughts at the same time. These thoughts may be initiated by the in-turner in the way they begin their utterance, but we might expect that for the out-turner to respond in a timely manner, she too must (try to, at least) take the perspective of the in-turner. And this needn't simply be a case of prediction on a turn-by-turn basis – that is as A starts an utterance, B starts the processes of prediction during that particular turn – it may be that this prediction overlaps into other turns. So A might start an utterance, which also involves prediction of how she will complete her own turn,[99] and based on what she is uttering she might also be making predictions as to how B might respond to her before B even begins to respond. This is the type of idea Pickering and Garrod (2013) propose - that conversational partners use forward models as a means of predicting beyond a mere turn-by-turn basis;

> [W]e propose that speakers use forward models to predict aspects of their upcoming utterances and listeners covertly imitate speakers and then use forward models based on their own potential utterances to predict what the speakers are likely to say. The account helps explain the rapidity of production and comprehension and the remarkable fluency of dialogue"[100]

Although whether or not interlocutors begin such predictions as early as Pickering and Garrod hypothesise is correct, it still appears from the other data that at the very least out-turner prediction often must begin before the in-turner has completed a turn. So we might think out-turners must attempt to at least match their perspective with that of the in-turner. Put in this way, it might sound more like an out-turner is the conversational participant with the burden of making perspectives align, but the task of prediction is not an enterprise undertaken by the out-turner alone. The in-turner has a part to play too, and they must also make sure to align with an out-turner just as the out-turner aligns with them, and one of the ways of doing this is suggested by some of the research into interactive alignment.

---

[98] And also recall; "I am freed from myself in the present dialogue, even though the other's thoughts are certainly his own, since I do not form them, I nonetheless grasp them as soon as they are born or I even anticipate them." (2012, pp.370-371).

[99] Because, of course, we very often have not planned a whole utterance in advance of making it, but rather we prepare just how we will start it.

[100] (2013, p.346)

3.1.5 *Interactive alignment*

The notion of interactive alignment stems from studies showing that interacting groups will align in a number of different ways, both behaviourally and cognitively. For example, as conversation progresses eye-movements, hand positions, facial expressions and word choices of interlocutors will often become aligned. As Marlou Rasenberg, Asli Özyürek, & Mark Dingemanse (forthcoming) note;

> In social interactions, speakers coordinate their actions in an effort to incrementally and interactively reach their communicative goals. A pervasive component of such joint actions is cross-participant repetition of communicative behavior. Work across a wide range of fields shows that when people are engaged in interaction, their behaviors may grow to be in tune with each other at several levels: from body postures and eye gaze, to words and gestures. (Rasenberg, Özyürek, & Dingemanse, forthcoming)

And when we think about it, this is not unexpected, as Dijksterhuis & Bargh note;

> [W]e have an innate tendency to imitate. We whisper to someone who is whispering, we start to speak much louder when others do so. We scratch our head upon seeing someone else scratch their head. We walk slower in the presence of the elderly, we cycle faster after we have seen a cycling race on TV"[101]

We can postulate that this is a powerful tool when trying to establish commonality during interaction. Simon Garrod and Martin Pickering (2004) address why it is that conversation is so "easy" and argue that studies into interactive alignment show that it's easy precisely because participants in a conversation interactively align. So during an interaction in which participants align in various ways, the eventual outcome of this is an alignment of their linguistic representations. They also present evidence that interactive alignment processes occur automatically and suggest that "the interactive nature of a dialogue supports interactive alignment of linguistic representations." Further to this, they also note that:

> [P]eople 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

[101] (Dijksterhuis & Bargh, 2004)

> other levels… Conversations succeed, not because of complex reasoning, but rather because of alignment at seemingly disparate linguistic levels.
>
> (Garrod & Pickering, 2004, p. 9)

The key point to note here is that alignment at one level leads to more alignment at other levels.

On interactive alignment Deborah Tollefsen and colleagues (2012)[102] state that "it has now become evident that when individuals engage in a joint activity such as conversation or joint problem solving they become aligned at a variety of different levels."[103] They also refer to alignment as "the dynamic 'matching' between the behavioral or cognitive states of two or more people. Gestures, eye gaze, word choice, and various other behavioral features may become coordinated in human interaction." So although Garrod and Pickering focus primarily on the different linguistic levels, we might suppose that as conversational partners align at the behavioural level and the linguistic level, this might lead to alignments of other cognitive processes. Indeed, they suggest later in the same paper that "the alignment of representations has the effect of distributing the processing load between the interlocutors because each reuses information computed by the other."[104] If all this is true, then, it would seem to suggest that in a paradigm type of conversation not only is an out-turner required to predict what the in-turner will say, the in-turner and out-turner will often align in other ways to make this predictive job easier. But not only does this make the predictive job easier, there is evidence, too, that synchronising with an interlocutor also gives participants a sense that the interaction is smooth and harmonious.

For example, Kerry Marsh and colleagues (2009) review studies that look at the role of synchrony between interactants and how subjects reported this synchrony made them

[102] See also Tollefsen *et al.* (2013)
[103] (2012, p. 386)
[104] It is this distributing of the processing load that occurs through alignment that then raises the question of what type of system is at work here – so if this is the case, there is perhaps a question of how do we go from simple alignment to sharing cognitive loads.

feel.[105] In one of the studies[106] participants working in pairs were assigned either an easy or difficult task in which they were asked to bodily synchronise with each other. They report that participants responded in a way that suggested; "[e]asy coordination condition interactions were viewed as significantly friendlier, smoother, and more harmonious and these effects were not reduced when mood effects were partialed out."[107] That is, when participants found it easier to synchronise their bodily movements, they reported that this made them feel as if they more closely interacted with their experimental partner. And it isn't merely explained by the fact that that they succeeded in performing the experimental task. Other studies they looked at used distraction tasks not related to bodily synchrony whilst still testing for the same phenomenon. They note;

> One striking feature of the bodily synchrony studies is that…the tasks typically did not involve an interpersonal goal. Yet even in situations where an unintentional interpersonal pull toward synchrony might be at odds with (distract from) the purported experimental goal (e.g., memorizing words), individuals spontaneously coordinated their incidental movements with another individual. The ease of doing so was associated with greater feelings of connectedness—a feeling of readily being a team with the other person.[108]

Based on such studies, then it seems that not only do processes of alignment and synchronisation between interlocutors make the necessary predictive tasks of conversation easier, they also make participants feel more positive about the interaction. This becomes particularly pertinent for the discussion in Chapters 3 and 4 on the contrast between face-to-face and digital conversation, but for now let us return to consider how interlocutors help each other with the tasks of prediction.

3.1.6 *Prediction and perspective sharing*

[105] I use only one example here for ease of reference, but the correlation between synchrony and prosocial attitudes is well observed. See for example meta-analysis of 42 different studies found in (Mogan, Fischer, & Bulbulia, 2017) and analysis of 60 studies found in (Rennung & Göritz, 2016). Evidence exists too that in cases of disputation, synchrony also decreases, see for example (Paxton & Dale, 2013). There is evidence too that even observing synchrony of a dyad from the outside can make participation more desirable for the observer (Marques-Quinteiro, Mata, Simão, Gaspar, & Farias, 2019). And (Reddish, Tong, Jong, & Whitehouse, 2019) finds that synchrony increased reports of joint-agency.
[106] (Marsh et al., 2007)
[107] (Marsh et al., 2009, p.330)
[108] (Marsh et al., 2009, p.332)

In-turners too, it seems, play a role in providing an out-turner with cues and clues as to the things they are going to utter, such as making clear the type of speech act apparent to the out-turner early into a turn. For example, it was observed by Sicoli *et al*. (2015) in a study across 10 different languages that even speakers of languages that do not include speech act indicating particles at the beginning of an utterance, speakers will, for example, boost the pitch at the beginning of a question. And based on an electroencephalogram (EEG) study, (Gisladottir, Chwilla, & Levinson, 2015) suggest that "speech act recognition begins very early in the incoming turn, starting already from 200 milliseconds after first word onset when the utterance has only been partially processed."[109] [110] We might hypothesise then, that by making clear to an out-turner early on that, for example, a question is forthcoming; this helps the out-turner narrow down the scope of what will be uttered and potential responses. The out-turner might also assist by providing particularly salient content early into a turn. For example, Bögels, Magyari, & Levinson (2015) show the efficacy of doing this. They used an EEG to look, in part, at the timing of production planning in response to questions such as the following;

(9)

Early: Which character, also called **007**, appears in the famous movies?

Late: Which character from the famous movies, is also called **007**?

So '007' here represents the information the testers deemed necessary for participants to answer the question. Based on observations from the EEG, they found that in preparing to respond to questions such as those in (9);

> [E]ffects started already around 500 ms after the onset of the information that enables participants to retrieve the answer. This indicates that interlocutors started production planning within half a second of hearing the critical information necessary to start answer preparation.[111]

[109] For context, see again fn.51, there it was noted that uttering a syllable takes approximately 200ms.

[110] There is evidence, too, that the use of gestures accompanying questions also improves response speeds, for example Holler, Kendrick, & Levinson (2017). Indeed, the consideration of how such multi-modal language components as gesture, facial expression and linguistic output align to generally produce faster responses is interesting. See for example, (Holler & Levinson, 2019).

[111] (Bögels, Magyari, & Levinson, 2015)

It is in the interest of the proceeding of a conversation, then, that an in-turner provides the out-turner with as much information as possible early in a turn to help the out-turner(s) predict the type and content of the utterance they are making. This inter-weaving, or perspective sharing, then, appears to be essential for conversation to function in the way that it does. And as Stephen Levinson notes;

> As far as we know, the overall system employed in conversation is strongly universal, with only slight variations in timing, and it contrasts with other more specialized speech exchange systems such as those employed in classrooms, courtrooms, presidential press briefings, etc., which tend to be culture-specific. (Levinson, 2016)

3.2 *Levinson's preconditions*

In light of the above data, we have a clearer idea of the processes involved in conversation and the timescales these process tasks must be performed within. It seems obvious, to me at least, that for paradigm types of conversation to function as they do, that whether in the role of out-turner or in-turner conversational participants must remain continually active – participants must devote their attention to the conversation continuously. So even if *prima facie* it might appear that there is a neat division of the roles in a conversation – for example, a speaker and a listener in a verbal conversation wherein the speaker is active and the listener passively listens - this is far from the case. Very often the role of listener in a verbal conversation is much richer than merely receiving and processing information, they also have a number of further requirements to fulfil in order to be able to respond in a timely manner. In summary of these tasks, then, Levinson suggests the following preconditions would need to exist in a conversation between A and B and where A is the in-turner, to enable B to respond within the average 200ms timescale:

(10)

> (i) B must attempt to predict the speech act (detect whether A's utterance is a question, offer, request, etc.) as early as possible, because this is what B will respond to;
>
> (ii) B should at once begin to formulate a response, going through all the stages of conceptualization, word retrieval, syntactic construction, phonological encoding, articulation;

(iii) meanwhile, B should use the unfolding syntax and semantics of A's turn to estimate its likely duration, listening for prosodic cues to closure;

(iv) as soon as those cues are detected B should launch the response. [112]

We might expect, then, that these preconditions (or at least something like them) are general for all of our cases of paradigm conversation.

Contrast this with some of the other cases discussed in Chapter 1. In cases such as the courtroom or job interview, we have no reason to expect there is anything resembling the one second window (though this would be an empirical matter), if there is no pressure to answer within such a short time-frame, then there is no requirement for prediction on the part of the interviewee, for example. In the case of an academic lecture, there is no requirement at all on the listener to respond. And as such Levinson's preconditions need not hold. That is not to say they wouldn't to some degree. It may well be that language processing requires something like prediction of speech act to help understand an utterance (and so would meet (10i)). Though if one member of the audience fails to do this, there will be no consequence for the lecture itself. Whether an attendee understands a lecture is of minor significance for the lecture as a whole. This isn't to say that listening to a lecture might be less cognitively demanding, it may well be more difficult. An attendee might, for example, need to do such as draw upon previous knowledge, consider where the teachings of the lecture fit in with a wider picture, take notes etc… The notable contrast between the lecture and our paradigm type of conversation, though, is there is no requirement that a lecture attendee do any of these things, whereas in conversation, it is very often minimally required for the conversation to proceed that an out-turner be engaged in a set of required tasks.

3.3 *Summary of processes*

In this section, then, we have looked at some of the types of process tasks interlocutors in our paradigm type of conversation must complete. The evanescence and recordlessness of the medium, and the time constraints imposed on interlocutors mean that in order to respond in a timely manner, out-turners must start planning their turn before the in-turner finishes their turn. So it is the case that the out-turner must predict the type of speech act being performed and predict how and when it will finish. To assist in this the in-turner must

[112] (Levinson, 2016, p. 7)

also give cues and clues to the out-turner. One hypothesis as to how this occurs is that interlocutors interactively align in various ways across different communication channels. So in a paradigm type of conversation, interlocutors necessarily must coordinate at the level of process in an interdependent way. I want to argue next that underlying this coordination is a type of cooperation. That is to say, in order for conversation to succeed and proceed it requires that conversational participants mutually (and tacitly) agree to cooperate with each other to fulfil the roles expected of them. So even if conversational participants are in disagreement at the level of content, they must still cooperate at the level of process, and as such paradigm cases of conversation will be essentially cooperative.

4 Cooperation and coordination of process

Paul Grice famously makes cooperation central to his account of conversation. Though using the distinction between process and content discussed so far, most would probably interpret Grice as referring to cooperation at the content level. In this section, however, I want to suggest that the interdependent nature of content and process levels in conversation can help us understand a little better the cooperative principle. More precisely I will argue that we can understand better some of the tasks a conversational participant must undertake to enable them to adhere to the cooperative principle. So we shall see the influence of cooperation of process on the coordination of content. Which leads me to suggest that at the very least a paradigm case of conversation is essentially cooperative at the level of process, and that this cooperation at the level of process is influential, too, at the level of content. And so even in a case of disagreement, a paradigm case of conversation is is in some sense intrinsically cooperative. My aim here, then, is to bridge the notion of coordination suggested above, and the type of cooperation that underpins the Gricean account. Space precludes introducing a fully-fledged theory of cooperation at this stage, and so the notion of cooperation I work with here will remain intuitive.

As a starting point I am considering that the coordination discussed above is a systematic set of tasks necessary for the functioning of conversation. Cooperation on the other hand, I take to be the tacit agreement that interlocutors will commit to the coordinative tasks required of them in order to sustain the conversation - that is, interlocutors cooperate to meet a general shared goal of sustaining the conversation. Importantly, this need not be cooperation in the sense of being in agreement at the level of content. So although it might be seem unusual to describe two people vehemently disagreeing about some matter as being two people cooperating, in some sense they are

cooperating so long as they are at least carrying out something like Levinson's preconditions (11).[113] Instead what I suggest is that they cooperate primarily at the level of the process – they cooperate in their agreement to coordinate the processes required of them in order to sustain the interaction.[114] That is to say, in a paradigm case of conversation participants share a goal of maintaining the conversation and work together to achieve that goal for as long as they wish the interaction to continue, and in doing so they are cooperating in a deep and interesting way.[115]

4.1 *Coordination without cooperation*

Coordination can occur without cooperation. Two machines sending and receiving data are coordinated to an extent, though machines don't appear to be the type of thing that have goals. I can coordinate my body by hopping from one foot to another. Or to put to a different use an example Searle[116] gives; in the UK cars drive on the left-hand side. Assuming everyone in the UK adheres to the rule to drive on the left-hand side, then all drivers in the UK are coordinated. They systematically align to drive on the left-hand side, to drive around roundabouts in a clockwise direction, and to overtake on the right. There are also other ways they coordinate. They agree to stop for traffic at a junction, observe traffic lights, use indicators etc… But although they coordinate such a way, it doesn't follow that they require a shared goal and are acting together to achieve it. At a certain level of abstraction they might have the same goals with regards to following the rules of the road. Driver A and Driver B might both adhere to the rules so as to avoid trouble with the police, for example. But Driver A and Driver B aren't cooperating to meet this goal, they are simply two

---

[113] Of course, there will be some disagreements in which interlocutors are no longer listening to what the other has to say, or perhaps predicting what they might say but getting it wrong. These cases no longer seem to appear to be a paradigm case, perspectives are no longer being shared and such interactions can take the form of two people airing grievances regardless of the previous interactant's contributions.

[114] And the close relationship of content and process will entail that in these cases most often the cooperation at the level of process will also translate into cooperation at the level of content, indeed it would appear that cooperation at the level of content would require cooperation at the level of process.

[115] This definition of cooperation is close to Raimo Tuomela's *Basic Thesis of Cooperation*: "Two or more actors cooperate in the full sense if and only if they share a collective (or joint) goal and act together to achieve the goal." Of this he says: "The present thesis must be understood broadly enough to be compatible with the claim that not all cooperation needs be acting towards a collective end-state. This is because there can be full-blown cooperation which only involves shared activity, a collective action-goal, but does not purport to lead to a shared collective end or purpose at all." (Tuomela, 2000, p.12) This seems true of conversation, it may well be that there is no shared collective end, indeed often there won't be one beyond 'have a nice conversation' or something similar.

[116] (1992, pp.16-17)

individuals with the same individual goals. But there are perhaps cases where it might not look quite as clear.

Suppose two drivers Dot and Dash both have the same goal 'don't crash'. For it to be safe for Dot to drive on the left-hand side also requires that if Dash is coming in the other direction that Dash too is adhering to the left-hand side rule, if not then their goals will be compromised. So in a sense at the moment their journeys intersect, it is in the interests of Dot and Dash that they both have the goal 'don't crash'. So suppose Dot and Dash both share the goal 'don't crash', and both pass each on the road by driving on the left-hand side. In a sense it might seem that they have acted together to achieve a shared goal, and Dot and Dash don't crash. So perhaps it would appear Dot and Dash have cooperated to avoid crashing into each other. But even if we were to call this a case of cooperation, it differs in some respects from the type of cooperation we see in a paradigm case of conversation. For although Dot and Dash do cooperate, there is no requirement that they do. They could have very different goals and still not crash into each other.[117] [118] What I want to suggest of conversational cooperation, however, is that for conversation to function the participants must necessarily cooperate and that at the most basic level, conversational participants must share the same simple goal. That is, for a conversation to function, it requires that each participant must at least have the shared goal of wishing for the conversation to proceed. If any participant does not share this goal, then the conversation will cease to function as a conversation.[119]

---

[117] If we were to go further into aligning the intuitive notion of cooperation I work with here to a developed theoretical notion, it might be possible to say a little more on the distinction Tuomela draws between 'I-mode' and 'we-mode' cooperation, or 'full cooperation' and 'cooperation as coordination'. (See Tuomela, 2005; Tuomela 2011, pp.66-69 for a summary). If we wished to denote the example of Dot and Dash as cooperative, it might be deemed an 'I-mode' cooperative act. Both Dot and Dash have goals that are met by driving on the left-hand side, and by doing this both played their part in avoiding crashing and so meeting each other's goal. But this, really, is incidental. That their goals align and are acted upon does not give us the stronger notion of 'we-cooperation', it is not that Dot and Dash worked towards the goal together, their goals are merely (and fleetingly) contingent.

[118] As Rosanna Keefe points out to me, with Lewisian co-ordination problems solved by conventions, arguably there are always shared goals.

[119] Although I don't have space to develop this beyond the level of a suggestion, it strikes me that we could also relate this notion of cooperation at the level of process to something Searle (1992) says on the role of shared intentionality in his criticism of conversation analysis. "I believe that a recognition of shared intentionality and its implications is one of the basic concepts we need in order to understand how conversations work". Elsewhere Searle (1990) says of shared intentions; "Ask yourself what you must take for granted in order that you can ever have or act on collective intentions. What you must suppose is that the others are agents like yourself, that they have similar

4.2.1 *Symmetry of processes*

Recall earlier it was noted that paradigm cases of conversation are in some senses symmetrical. We are now in a position to understand a little more about the notion of symmetry in process coordination. Let's consider again the following example discussed in Chapter 1 (§2.1).

(11)

01	A:	Just think of how many people would miss you. You would know who cared.
02	B:	Sure. I have a *lot* of friends who would come

awarenesses of you as an agent like themselves, and that these awarenesses coalesce in a sense of *us* as possible or actual collective agents. And these conditions hold even for total strangers…my stance toward others with whom I am engaged in collective behavior is that of their being conscious agents in a cooperative activity, without my needing or having a special belief to that effect… The biologically primitive sense of the other person as a candidate for shared intentionality is a necessary condition of all collective behaviour and hence of all conversation." (pp.414-415) It is perhaps worth considering the notion of a shared goal to cooperate on sustaining a conversation to be a form of shared intentionality. This too could be adapted to work on two levels mirroring the Clarkian notions of content and process and can help understand the link between these two coordinative levels. So the shared intentionality is directed at a higher-level in terms an adherence to something like carrying out the process required for conversation to function. We could say then that it is thus directed at maintaining the interaction. On a secondary level is the type of intentionality Searle possibly has in mind (and something like Green's Stalknakarian expansion project has at its core) – the general question (or questions) that the discussion is directed towards. These two levels of intentionality are essential for the continuation of the discussion and should one participant withdraw from either, they will withdraw from both. Suppose Verity and Mia are in conversation and Mia realises she needs to go for a bus. Mia might start to disengage from the higher level of process-intentionality, she no longer will wish to continue the conversation and the shared intentional activity, which might in turn dictate how she responds at the content level, she might start to look at her watch, or mention she has a bus to catch. Conversely, something at the content level might make Verity no longer wish to participate in the shared intentional joint activity of the conversation, suppose a third-party Ros joins the conversation and Mia and Ros start to discuss a topic of little interest to Verity. Verity's lack of interest towards the topic will disengage her from the content of the conversation, but it will also serve to disengage her from the processes that make interaction possible. That she loses her commitment at the content-level therefore will start to bear upon her ability (or intention) to maintain the higher-level shared intentionality. She might stop providing cues that she is engaged, she might stop all but the most basic form of listening (i.e. the words are *heard* in the sense that they are stored in working memory as phonological replica), and ultimately she might try and excuse herself from the conversation.

to the funeral and say what an intelligent,

bright, witty, interesting person I was.

03 A: They would*n't* say that you were *humble*

04 B: No. Humble, I'm not.

One notable point of symmetry in (11) relates to previous discussions on turn-taking and the different roles of in-turner and out-turner; we can assume that when A is engaged in the tasks associated with being an in-turner, B is engaged in the tasks relevant to being an out-turner. Considering that each new line as transcribed represents approximately one second of time, the whole exchange takes less than seven seconds. A and B are both contributing to the discussion, and as the time it takes to complete is short, then it also seems reasonable to assume that A and B are attentively engaged with their tasks. As noted in Section 2, the evanescence and recordlessness of the medium require that if the conversation is to proceed as it does, that in (11) both A and B are attentive to the process tasks required of them. So there is symmetry in their joint-commitment to the overall task of sustaining the conversation. This symmetry is not merely at the level of process either, for it also bears relation to the idea of conversation as a synocracy first discussed in Chapter 1 (§2.2).

4.2.2 *Symmetry of influence*

Suppose A and B are involved in a verbal conversation. If B is listening to what A utters then what A utters is, to some degree at least, influencing the direction of thought B will have, which in turn will dictate, to some degree, the potential content of B's response. Similarly, when A is listening to what B utters, then what B says is, to some degree also, influencing the direction of thoughts A will have, which in turn will provide the scope for A's response.[120] This might seem to come into some conflict with the features of self-direction and self-expression discussed in (§2.2.3), a person about to take on the role of in-turner will be restricted to some degree on what they can say in response if they are to make their contribution sensible. However, the synocratic nature of a paradigm type of conversation would suggest that the interlocutors are still free to, for example, attempt to change the topic of conversation. So there is a symmetry of influence reflected in a synocratic exchange too. This symmetry relies on the conversational participants tacitly accepting their changing roles as in-turner and out-turner, the commitments each role entails and also accepting

[120] Again, this goes back to the idea of perspective sharing first discussed in Chapter 1 Section 3.2.

their interlocutors'. This acceptance, I suggest, is rooted in a willingness of participants to cooperate in the processes required of a conversation. However, we need to be cautious here because, as alluded to in Chapter 1 (§2.2), it is not always the case in face-to-face interaction that each participant will have the freedom to determine the direction of the conversation. That is, in some exchanges one of the participants might not have the same opportunity for self-determination or self-expression (in the non-trivial sense) as their interlocutor(s). Naturally we see this in formal and institutional examples such as cross-examination or a job interview, but this can also be the case due to other factors relating to the social standing or role of interlocutors.

So although the notion of symmetry I discuss above is primarily the symmetry of processes interlocutors tacitly agree to cooperate on, part of the task here is to consider a paradigm type of conversation. In such cases the synocratic nature of the exchange is such that participants should be able to self-determine and self-express, and this too requires a type of symmetry – a symmetry of influence. And so in an ideal conversation we should also expect that all the participants have the freedom to influence the direction or topic of the exchange should they wish to (and reciprocally, their conversational partners are free to accept or reject this). Now clearly there are some cases such as a job interview where this freedom is restricted to an extent; the interviewee will conventionally do most of the talking but will most often be doing so in response to the interviewer's questions. So the interviewer holds a degree of autonomy on the general direction of the exchange, and therefore the dynamic is more asymmetrical than the paradigm case. In the case of cross-examination, this freedom is even more restricted. Recall the following extract from example (4).

(12)

04 A. Yes. He said it was a dark-colored –

05 Q. Hold it a second. Did he describe it? Yes or no

Here A is restricted in self-expression by the cross-examiner, the balance of control of the exchange is heavily weighted towards the cross-examiner. The power and authority of the court makes it so the cross-examinee must answer questions in the way the cross-examiner determines. Though they will still be free, to some extent, to determine what their answer will be. This said, the restrictions on self-expression in such cases would take it even further

away from our paradigm cases. There will also be many examples where this balance is skewed even further.

Rachel McKinney (2016) uses the case of the Central Park Five to discuss what she terms *extracted speech* – speech that an agent is made to produce. In that racially charged 1990 case, five black and latino teenagers were convicted of the rape and attempted murder of a 28-year-old white woman. The main evidence presented to the court was police video recordings of each of the five confessing to the crime in the interrogation room. All five recanted their confessions claiming they were false and wrongly obtained. After the emergence of new evidence the convictions were vacated in 2002. What McKinney argues is that the Central Park Five case shows how in some instances asymmetrical power dynamics in an exchange can, to use the terminology from Clark's control features, remove entirely an agent's self-determination and self-expression. Power dynamics, then, are important for considering conversation, and these relate to their effects on conversational participants' self-determination and self-expression. The greater the asymmetry of influence the more autocratic the exchange becomes and recalling the argument of Chapter 1 (§2.2), therefore the further away we get from the paradigm type of conversation. So a job interview may seem closer to the paradigm case than, say, a cross-examination. In those cases, the job interviewee is typically free to say what they want in response to the interviewer's questions, and so are only restricted in a limited sense. There appears to be a difference between the symmetry of cross-examination examples and extracted speech cases too. In a cross-examination, the control of the direction of talk is clearly with the cross-examiner, but the cross-examinee is not subordinated to the extent as we see in extracted speech cases wherein the person that the speech is extracted from has little control over the content of their contributions.

So we might think that on a continuum that starts with the paradigm case of conversation, the more asymmetrical the control of the discussion becomes, then the further away from the paradigm we get. So next to the paradigm case might be a job interview, at a few stages of further removal we might have a cross-examination, and then much further away again we have cases of extracted speech. And at this distant remove, too, we would probably also want to consider other types of case where an injustice or power imbalance restricts the freedom of an interlocutor to self-express or self-determine. So for

example cases where there is illocutionary silencing,[121] or *discursive injustice*,[122] would sit on the continuum closer to extracted speech than the paradigm case. A part of the injustice in such cases, we might say, are *conversational injustices* – a participant in an exchange is not afforded the status of a full conversational participant with the opportunity to self-express, but rather they are given a subordinate status in which they are not free to influence the direction of talk to the same extent as other participants. We should be careful to distinguish these types of exchange with other exchanges in which there may be a power imbalance of sorts; but where there is a willing imbalance of power and which does not greatly impede on a speaker's self-determination or self-expression.

Consider interactions between a willing student and a teacher. Most often the teacher will be in a position of power relative to the student, and the student might defer to the teacher in a number of different ways. She might allow that what the teacher says on a matter of fact is to be taken to be true, or allow the teacher to speak more on some matter.[123] But this need not entail that a student cedes all influence of the exchange to, for example, the teacher, she may still retain the right to ask questions of what the teacher says, for example. All I wish to highlight here, then, is that power-structures can impose restrictions on conversational participants that can suggest an exchange is distant in some respects to the paradigm case, however it need not necessarily be the case.

4.3 *Symmetry, synocracy and cooperation*

What I want to suggest now is that for the symmetry of the in-turner/out-turner relationship to develop, we might expect it requires that an in-turner adheres to something like Grice's Cooperative Principle;

(CP)

[121] For example (Langton, 1993)
[122] Kukla (2014)
[123] This being a type of *asymmetrical socratic* exchange noted in Green's taxonomy (2017) of types of conversation. In such cases there is an agreement (tacit or explicit) that one participant leads the discussion and the other agrees to follow. Though picking up on this being 'socratic' it's difficult to know how willing some of Plato's Socrates' interlocutors were, usually it appears it was simply a case of wrong place, wrong time.

> Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.[124]

Now of course Grice proposes this principle, along with the maxims that accompany it, primarily as a means of explaining the role of implicature in conversation, so it is most commonly viewed as cooperation at the level of *content*. But I don't think it is too far of a stretch to use this principle as a means of understanding the relationship between the process-level tasks discussed above and the content-level cooperation of Grice's CP. That is, I also think that with some modifications we can regard CP as being applicable to the processes involved in conversation.

I want to start by picking up on Grice's defence of CP as put forward in Strand Six of his 'Retrospective Epilogue',[125] there Grice says the following;

> While the conversational maxims have on the whole been quite well received, the same cannot, I think, be said about my invocation of a supreme principle of conversational cooperation. One source of trouble has perhaps been that it has been felt that even in the talk exchanges of civilized people browbeating disputation and conversational sharp practice are far too common to be offenses against the fundamental dictates of conversational practice. Another source of discomfort has perhaps been the thought that, whether its tone is agreeable or disagreeable, much of our talk-exchange is too haphazard to be directed toward any end cooperative or otherwise. Chitchat goes nowhere, unless making the time pass is a journey.[126]

He then goes on to offer three points of refinement of the notion of CP. It is the third of these, stated below, that particularly interests me here.

> [S]ince we are concerned as theorists only with concerted talking, we should recognize that within the dimension of voluntary exchanges (which are all that concern us) collaboration in achieving exchange of information or the institution of decisions may coexist with a high degree of reserve, hostility, and chicanery and

[124] (1989c, p. 26). Naturally if we are expecting the CP to hold, we might also expect something like the Grice's maxims to be important too. To recap these suggest that a contribution to be just as informative as needed (quantity), not knowingly false or under supported (quality), relevant and not ambiguous, obscure or disorderly (manner).
[125] (1989e)
[126] (1989e, pp. 368–369)

> with a high degree of diversity in the motivations underlying quite meagre common objectives.[127]

So the idea here is that even in exchanges that seem to be confrontational, there is still some basic level of cooperation between the interlocutors even if their ultimate aims for the talk exchange differ vividly. Part of what Grice is maybe getting at here, even if he stops short of saying it explicitly, is that the structural nature of a conversational exchange is a necessary collaboration between the participants. They simply *have* to cooperate, otherwise the talk exchange can't proceed as a communicative endeavour. So cooperation need not be cooperation in the sense of aiding each other in aims specific to a particular conversation, so, for example, my trying to convince you to do something and you complying. That would be cooperation of a sort, a cooperation specific to a particular conversation. But for current purposes the cooperation that I will apply CP to is specifically a process-level cooperation that applies across conversations. On this reading CP is simply a statement of background processes required to be completed in order for the conversation to proceed. Recall again that the content of a specific conversation is not what we're interested in here, but rather the general processes involved. If that is our consideration, then CP shouldn't seem controversial at all. The idea that two (or more) people involved in a language exchange (that genuinely functions as a conversation) need to cooperate to some degree, even if just to make the flow of the exchange sensible, seems almost trivially true. Though when we consider what this entails, it isn't of mere trivial interest.

I proceed now with the assumption that CP has use as a basis for understanding what makes something a paradigm conversational exchange in the first place, particularly in helping us grasp the inherent coordinative and cooperative aspects of a conversational exchange. For current purposes though, although CP is a starting point, it is not complete, or at least not explicitly. The primary focus of CP is on one specific part of a conversational exchange - the contribution of the in-turner at the time of their turn. To help with the current task though, we need also to think of the role of other participant(s) in a conversation; the out-turner(s).[128]

---

[127] (1989e, p. 369) It is worth pointing out Grice's use of "we are concerned as theorists only with concerted talking" in the Retrospective Epilogue and that he notes that 'voluntary exchanges' are what he is interested in. As per the discussion in Chapter 1, (§2.3.3) about simulations, it is in passages like this that make me think Grice has in mind a similarly narrow conception of conversation as the paradigm case being used here.

[128] This follows to some extent from a point made by Clark about the maxims being "exhortations to speakers, not addressees" (1996, p.146). I think this is correct in the sense that the way Grice

4.4 *Elements of the Cooperative Principle*

As it stands, CP acts as a principle we take an in-turner to be adhering to *qua* in-turner. The element of a conversational exchange it doesn't address explicitly is what an out-turner(s) need to do to allow them to fulfil CP once they become the in-turner. Grice does allude to the role of the out-turner to some degree, but only in so far as to say that the out-turner assumes the in-turner to be adhering to CP. But I think we can go further with this and develop a complementary principle that applies to the out-turner of a conversation. To start to do this let us now divide CP into three informational elements.[129] So using Grice's own wording, let's say the three closely related informational elements of CP are:

(13)

a) what is required of a contribution;
b) what stage the conversation is at;
c) what the accepted purpose or direction of the talk exchange is.

The first thing to note here is that each element contains distinct, though likely highly interlinked, information. To know (13a), one must probably first know (13b) and (13c). The requirements of a contribution will be dependent on the stage and direction or purpose of the exchange. If this is the case, then, in terms of processing, we should expect that (13b) and (13c) to be established before arriving at an understanding of what qualifies for (13a). So (13b) and (13c) would seem to be prior to (13a). Depending on our reading of (13c), it could be that the case of the relationship of (13b) and (13c) is less clear however.

We could read (13c) in a quite rigid way. So a conversation about the weather simply has the purpose of broadly discussing the weather. If this were the case then it would seem (13c) is prior to (13b), to understand the stage of the conversation seems to require knowing the accepted purpose of the exchange. But this is quite an unappealing

---

phrases it ,the focus is on what the in-turner must do, though on my reading it seems not to be an exhortation to a speaker, but rather what an addressee will assume the speaker is adhering to.

[129] I'm almost certain this goes beyond what Grice was thinking with the Cooperative Principle, I don't think he viewed it as containing three elements, and if we do wish to break it down into smaller elements I'm open to the idea that there may well be better ways of doing it. I also don't intend this to be taken as reflective of the psychological reality, that is, I don't expect that any participant in a conversation considers each of these points individually before making a conversational contribution. This is only intended to be illustrative of the types of immediately salient information changing during a single turn. So what I really want to show with this breakdown is that meeting CP requires being aware of the different dynamic facts of a conversation. I'll get to this point in §4.5.

reading of (13c), for it doesn't account for the rapidly changing scope of content in a conversation. Think again about (11).

(11)

01	A:	Just think of how many people would miss you. You would know who cared.
02	B:	Sure. I have a *lot* of friends who would come to the funeral and say what an intelligent, bright, witty, interesting person I was.
03	A:	They would*n't* say that you were *humble*
04	B:	No. Humble, I'm not.

It seems difficult to pinpoint a single purpose or direction this talk exchange has. The first two turns seem to be directed roughly towards death or funerals, with some change of subject occurring in 02, whereas the second two turns address traits of one of the interlocutors. So, based on (11) at least, it seems that this purpose or direction is subject to revision as the conversation progresses. If this is the case, then reading (13c) as prior to (13b) in the way suggested previously is unsatisfactory.

A rigid reading of (13c) doesn't represent what occurs in many conversations. So we need to allow that (13c) be dynamic in the sense that it is frequently changing throughout the course of a conversation. This being the case, we might then wonder how (13b) and (13c) relate. To know what stage a conversation is at would seem to require at least some grasp of the direction and purpose of an exchange, and knowing the direction and purpose of an exchange would also presumably require an understanding of what stage the conversation is at. Why keep (13b) and (13c) apart then we might wonder. One reason we might want to do so is that although (13b) and (13c) are co-dependent, they could also potentially each change at different points in a conversation. If we imagine that (13c), very roughly, picks out something like the current topic or theme of the conversation, and (13b) approximately picks out something like the progression of talk on that particular theme, then (13c) is generally likely to progress more slowly than (13b). Looking again at (11). At 01 A is talking generally on the theme or topic of death or funerals. At 02 B is also talking on

this theme, though the progression on the topic they make within 02 means that by 03 the topic has changed to being specifically about B. On its own I don't take this to be quite enough alone to suggest that (13b) and (13c) are genuinely distinct, however if we consider the type of thing these three elements pick out, the difference becomes clearer.

4.5 *Dynamic facts*

What (13a), (13b) and (13c) ultimately are is a set of dynamic facts about a particular exchange.[130] In many cases these particular facts would only hold at the point an in-turner begins her turn. Each new turn would, most likely; modify what is required of the following contribution; alter what stage the conversation is at; and, even if only subtly, change the purpose or direction of the talk exchange. So although an in-turner's turn may be adhering to CP, the facts that CP relies on will alter during her turn. This alteration, though, becomes relevant not to the in-turner, but rather to the next person to take on the role of in-turner. As such, an out-turner, to enable herself to adhere to CP on her transition to in-turner, must be alert to any changes made to the facts underpinning CP that are enacted by the in-turner during her turn. So being able to fulfil CP when becoming in-turner requires that the conversational participant, when taking on the role of out-turner, is required to carry out the types of tasks discussed in Section 3. Notable about all of these elements and the dynamic facts relating to them, then, is they all relate to the content-level of a conversation. They are facts about some particular conversation. Recall in Section 4.3 I suggested that we can use CP to understand a little more about the interaction at the process-level and content-level, let us now do this by considering the types of process an out-turner must engage with in order to meet CP.

4.6 *Meeting the cooperative principle*

If we consider the requirements created by the evanescence and recordlessness of face-to-face conversation, then an out-turner in a conversation such as (11) would need to listen to the in-turner just as the in-turner is making her utterances. She is required to process what the in-turner is uttering in relation to the conversation itself in order to establish any alterations to dynamic facts such as (13c) - what the accepted purpose or direction of the talk exchange is. From this, or in-line with this, the out-turner must also determine what the in-turner's utterance does to dynamic fact (13b) – so at what stage does some utterance put the conversation. Further to this, in order to carry out (13a) when becoming in-turner, she

[130] Which, of course, is perhaps best considered part of the common ground.

must then also determine what scope remains within which to offer a reply that contributes to the conversation based on the updated facts of (13b) and (13c). Once this scope is established, she will also need to determine the direction she then wishes to take the conversation in.[131]

But as we know by now, and taking a verbal exchange as an example, the listener in must also do a lot more than listen, the listener must also start the process of preparing to speak which also involves its own set of sub-processes such as conceptualization, word retrieval, syntactic construction, phonological encoding, before reaching the stage of articulation. Typically, all of these cognitive tasks are being carried out in *real time*. That is, they are carried out concurrently as someone else is speaking or signing, *not* when they have finished their turn. As such, the complication is that being able to adhere to CP requires the out-turner not only to process what the in-turner is uttering as she utters it, but in order to respond in a timely manner, the out-turner need also *predict* what the in-turner is *going* to be saying in the rest of her turn too. So even if conversational participants are disagreeing in terms of the content of what they are saying, they are still coordinating in the sense of keeping the conversation going, which requires they both keep track of the dynamic facts of the conversation as they alter. In light of this, and with Levinson's preconditions (10) in mind, we might then suggest that in order for CP to be adhered to, a conversational participant must also engage in something like the following:

Precondition to the Cooperative Principle (PCP):

> In the role of out-turner, concurrently with the in-turner producing an utterance; a conversational participant must predict alterations to the purpose or direction of the language exchange, the stage the exchange will be at following the present turn, and begin the set of sub-processes in preparation to produce an utterance.[132]

Once PCP has been met, then when the out-turner transitions to in-turner, the new in-turner would be in a position to adhere to CP. We should note that the dynamic facts

[131] Now of course, as in our paradigm case of conversation all participants are free to self-determine and self-express, they are also free to not adhere to CP in the sense discussed here. They may choose to start the conversation on a completely new path unrelated to previous discussion. Naturally this relies on the other interlocutors accepting this change of direction, and it is not something that can be readily deployed. In the great majority of turns this won't happen.

[132] Implicit in here too is that in face-to-face conversation, when in the role of out-turner there is generally an either speaking or signing as PCP is being met.

predicted by the out-turner are still subject to change until the end of the in-turner's turn, and so may require some revision if the speaker says something that further alters one of the dynamic facts required for adhering to CP.

What I hope PCP makes clearer is the role the process-level tasks have in helping to determine and formulate the content required to meet CP. In the role of out-turner there are prediction and production tasks required and these are dependent on the dynamic nature of the informational elements of CP. For it is those *contents* that the *process* tasks will track. As discussed in Section 3, there is a much richer set of process tasks required for conversation to proceed than merely those presented in PCP, and I argued that the high degree of coordination at that process level is indicative of a highly cooperative activity (at least in terms of process). So all I want to suggest here is that PCP should be viewed as a postulate of how these rich process tasks interface with the content of some conversation. Interlocutors coordinate and cooperate at the process level to sustain conversation, and if cooperation is contagious, then we might expect the necessary cooperation at the process level also makes cooperation at the content level more likely (if not guaranteed). So the cooperation at the process-level can in turn promote at least some cooperation at the content level.

4.7 *Cooperation and content; a conjecture*

There is good reason, then, to think is that paradigm conversations have cooperation built into them. That's not to say they will be agreeable exchanges in all instances, that's clearly not the case. However, we might wonder if this in-built necessary cooperation is significant. This is only conjecture really at this stage, but I wish to plant the seeds for some of the discussion in Chapter 4. We might wonder whether the intrinsic cooperative nature of conversation is actually a key defining aspect to conversation and its cooperative nature explains why conversation has been such a powerful and uniting part of human sociality. For most of us, I imagine, understand how effective conversation can be for resolving differences. Sitting down in someone's presence is very often the best method of sorting out disputes, apologising for mistakes, or trying to rebuild a relationship. It doesn't always work, of course, sometimes it simply isn't enough. But what we might wonder is whether the deep level of cooperation at the process level of conversation might make us more receptive to cooperate in other ways. If two people in conflict are able to cooperate to converse, it perhaps would give them a sense that cooperation is possible.

5 Conclusion

I have argued that taking the paradigm case of conversations sketched in Chapter 1 we can start to see some general features of such interactions, and that these features suggest the paradigm cases of conversation are intrinsically highly-cooperative activities. For conversation to function in the ecology it does wherein the medium is evanescent and goes unrecorded, and contributions are made extemporaneously and opportunistically in response to the contributions of others requires a high level of coordination. This coordination has two components as pointed out by Clark – process and content. As content will vary to a large degree from conversation to conversation, I focused the rest of the chapter primarily on process coordination (though these levels of coordination are interdependent). After discussing some of the recent work from psychology, psycholinguistics, linguistics and sociology on conversational response speeds, language production and interactive alignment, I argue that coordinating at the level of process is highly cognitively demanding and requires a significant commitment from conversational participants. This being the case, I suggest that essential to cases of paradigm conversation is an agreement by interlocutors to cooperate in sustaining the interaction. That is, paradigm conversations are highly and continuously cooperative at the level of process and thus are essentially highly cooperative activities. In the next chapter, then, I begin the task of comparing this paradigm case of conversation to the type of conversations we have online.

3

The Written-Spoken Distinction

In Chapter 1 we looked at what we might call a paradigm type of conversation, and in Chapter 2 we looked at some of the processes that participants in such conversations are required to perform in order to sustain those conversations. We saw how some of the features of face-to-face conversation, such as the evanescence, recordlessness and simultaneity of the medium, and the instantaneous and extemporaneous nature of contributions, interact to create a particular set of demands on interlocutors. These features coupled with the rapid response speeds we typically see in face-to-face conversation give us a picture of a paradigm case of face-to-face conversation as being a fast, synchronous and intricately coordinated joint action. I then argued that this coordination is indicative of a deep and sustained cooperation on the part of interlocutors. In order to sustain a conversation, conversational participants must continuously share the goal of sustaining the interaction, and work together to achieve it. In this chapter I begin the task (to be continued in Chapter 4) of considering how much of this model applies to cases of digital conversation - the types of which we typically find mediated by technology such as mobile phones and computers. The point of the comparison is based on the following intuition; despite the spread and availability of digital conversation in many of our lives, the quality of these conversations and the satisfaction we derive from them is, very often, inferior when contrasted with face-to-face conversations. To put it another way, considering again Erving Goffman's description of conversation noted in Chapter 1, it seems that digital conversation somehow seems less likely to give us a sense of entering a *unio mystico* with an interlocutor.

And so it seems that although very similar in some ways, face-to-face and digital conversation are notably different.[133]

What I consider in this chapter is whether there is something particular about the modalities themselves that might help us to understand why the above intuitive point might hold (and I do this by focusing specifically on the differences between speaking and writing). The overall question being asked here is whether or not differences in digital and face-to-face interactions might have some explanation in what I will call the 'surface-level' differences between the modalities used. Section 1 is a broad discussion on the emergence of writing as a conversational mode. I suggest that these digital conversations can have many of the features of a paradigm type of conversation (such as discussed in Chapter 1). I also make the point that we should pay attention to this form of conversation because despite its relative youth as a widespread conversational mode, we have good reason to expect it is no passing phase – written digital conversation appears to be here to stay. Following this I briefly discuss some of the phenomena that we might associate with digital conversation before considering this again in relation to the intuition that these conversations are oftentimes somehow less satisfying. In Section 2 I look at what I call the *surface level differences.* There I contrast writing and speaking in terms of their acquisitional differences, the differences in their historical roles, and the differences in what I call the *communicative bandwidth*. I argue that it may well be the case that all of these differences play some part in explaining the current differences between face-to-face and digital conversation, though it is not inconceivable that some of these differences can be wholly or in-part reconciled. That digital conversation is so new in our communicative lives makes it conceivable that in time such differences may become less-pronounced, and so I argue they are not explanatorily sufficient alone to explain the intuitive idea that face-to-face and digital conversation are importantly different.

1 Written digital conversation

The internet and mobile phones have transformed the possibilities we have to communicate with each other. In the fourth quarter of 2019, the internet had 4.1 billion users,[134] social

[133] Which is certainly not to say we can't have good quality and deeply satisfying digital conversations, we can, but rather the thought is that we often don't when contrasted with face-to-face conversation.
[134] https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

media platform Facebook had 1.6 billion active daily users,[135] and as of June 2017 there were 5 billion unique mobile phone subscribers in the world.[136] According to Internet Live Stats, in 1 second on average 8,855 tweets are sent, 1,657 Tumblr posts are made, and 2.8 million emails are sent.[137] One thing that is particularly interesting about this is the mode of language much of this new type of communication uses - writing. Although writing has been used for communication since at least the Bronze Age, literacy itself has only become widespread in the past hundred years.[138] And up until recent technological developments, writing was, in comparison to speaking, not a common modality used for everyday social interaction.

1.1 *Paradigm digital conversations*

What I refer to by *digital conversation* relates to the types of exchanges that take place using online services and platforms in which people can exchange (usually quite short) written messages with each other. So for example SMS[139] and its derivatives such as Twitter,[140] Weibo[141] and WhatsApp,[142] and other social media sites such as Facebook, Instagram, WeChat and QQ. Other examples could include message boards in which users post responses to each other, real-time chat rooms wherein users meet in a virtual 'room' and exchange messages quasi-synchronously and instant messenger services such as Facebook messenger. These types of digital conversations will mostly be conducted online, or be mediated by an electronic device, and as a result are rarely conducted face-to-face. The

[135] https://s21.q4cdn.com/399680738/files/doc_financials/2019/q4/Q4-2019-Earnings-Presentation-_final.pdf
[136] https://www.gsmaintelligence.com/research/2017/06/number-of-unique-mobile-subscribers-worldwide-hits-five-billion/624/
[137] https://www.internetlivestats.com/one-second/#traffic-band
[138] For example (Mira d'Ercole, van Zanden, Baten, Rijpma, & Timmer, 2014) estimate that in 1820 less than 20% of the global population were literate, and most of those who were literate were concentrated in Western Europe. According to (Roser & Ortiz-Ospina, 2020), global literacy didn't rise above 50% until 1950.
[139] 'SMS' stands for short messaging service and was the standardised service used by most mobile networks during the growth of text messaging, a single message is capped at 160 characters and originally this restriction was apparent to the writer of such messages. As technology developed, though, mobile networks have allowed concatenation of what are still individual packets of 160-character messages and phone manufacturers have integrated internet-reliant messaging clients such as iMessage which allow for longer messages.
[140] A service in which users post character limited messages (called 'Tweets', originally 140 characters, changed to 280 in 2017) either directly (though usually publicly) to another user, or to their 'followers'. The 140-character limit originally was used to allow users to post Tweets using SMS.
[141] A similar service to Twitter popular in China due to a ban on Twitter.
[142] An encrypted messaging service similar to SMS text messaging but with no character limits and wider applications for group chat.

widespread use of written language as a mode of conversation is new, it has only been happening on anything we might call a significant scale since the mid to late 1990s. Arguably this timescale is even shorter and probably only truly emerges as significant in the mid-2000s with the arrival of Web 2.0. For many of us social media is an ubiquitous feature of life and so we are engaged in or observing digital conversations routinely as part of our communicative lives.

It is certainly the case that we can have digital conversations. And these digital conversations appear to have some of the features of face-to-face conversations; there are usually two or more participants, a shared language and turns are taken. And just like in our paradigm type of conversation there is often a synocratic symmetry of influence in many such exchanges; I am free to self-express and self-determine, so is my interlocutor, and we can both contribute or control the direction of the conversation. In digital conversation we can enter the same kinds of joint-projects and perform many of the same types of speech act; I can make a request, which you can accept or decline. The topics of a digital conversation can be as varied and fleeting as in a paradigm case of face-to-face conversation and we can enter into them with close friends or with complete strangers. Indeed there are many features of such conversations that we might think make writing ideal as a conversational mode. We might expect that the ability to easily find interlocutors with similar interests would be conducive to good conversations. That very many of us are continuously connected allows that if we feel in need of searching out conversation we can. And there is a degree of greater freedom available - we can choose much more easily who and when to interact with, and the choice of interlocutors isn't impeded by geographical location. And so there are some close similarities to our paradigm type of conversation, and some ways in which we might think digital conversation provides even better conditions for entering into such conversation.

1.2 *Conversation written down*

Speaking and writing have generally operated within very different spheres of social life. As recently as 1989, linguist MAK Halliday observed that "[w]ritten language, never was, and never has been, conversation written down."[143] Although Halliday's point wasn't that written language *couldn't* be used for conversation,[144] it is of course true for almost all of

[143] (1989, p.41).
[144] His comment is more directed to the differences in how language is presented in writing vs conversationally.

our linguistic history it has only rarely been used as such. Writing has many uses, but rarely has it been used for conversation. There are exceptions to this, we might think of notes being passed back and forth a classroom when the teacher's back is turned, for example, or a series of letters in between correspondents, but generally it is true that writing has never been conversation written down. So it is interesting that in recent years we have adopted written language to perform a new role as a mode of conversation. This doesn't seem to be a mere passing moment either. If we presume hyperconnectivity – the instant availability of interlocutors for communication anytime and anywhere[145] - is here to stay, then we should expect written digital conversation is too. For although the use of text for digital conversation was most likely necessary when the technology was rudimentary, the bandwidth and technology available would very easily allow us to revert to other modalities should we so wish. As technologies have developed the environments in which we have digital conversations have become more multimedia by incorporating pictures, emoji, gifs and videos, yet we don't appear to be abandoning the use of the written mode, so we might think that written digital conversation is here to stay.

1.3 *Digital conversational phenomena*

Emerging out of these developments have been a number of interesting phenomena many familiar with the online world may have observed. For example, the permanence of writing and the often-public nature of this new form of conversation has resulted in the involvement of unintended or unexpected audiences in discussions and we see frequent failures of communication (for example, without explicit signposting jokes, irony and sarcasm on the internet can often fail). And so there are interesting consequences for how we consider the boundary between what is said and what is implicated and how we hold people liable for what they have uttered.[146] The emergence of this type of phenomena is amplified by how easy it is to remove an extract of text from one context and present it in a new context – a joke made to a friend might not translate well to a larger, disparate audience. We also see arbitrary limitations placed on utterances – so this could be actual word or character limits such as Twitter's 280 characters, or even limitations set by an utterer aware that often their utterance will be placed in some environment where there is a stream of other comments vying for attention.

[145] (Quan-haase & Wellman, 2005)
[146] See for example discussion of the case of the 'Twitter joke trial' in Borg & Connolly (forthcoming).

Another thing we might have noticed is the proliferation of dishonest behaviour (not that this is new, of course, but rather its scale and success are what is novel). So the spread of digital conversation has coincided with, for example, the rise of trolling and trolling behaviour (more on this in Chapter 5) and the spread of 'fake news'.[147] It is also now the case that non-human conversational participants such as so called 'bots' and AIs becoming part of our conversational lives. And our social spaces have become open to the spread of targeted attacks and advertising by governments and their agencies, corporate interests and activists. Mixed in with all of this we also see emergence of new types of speech act, for example retweeting,[148] hashtags,[149] and the topic of Chapter 5 - the act of trolling. So despite the similarities with face-to-face conversation at the level of an individual conversation, there are some larger-scale phenomenological differences that arise from it. In this chapter and the next my focus is on how individual digital conversations contrast with face-to-face conversations, and so I save the task of discussing these particular socio-linguistic phenomena for Chapter 5. Let's now finish this section by returning to the intuition that there is something importantly distinct between spoken and digital conversation.

1.4 *Two types of conversation?*

Of most interest to me in this chapter and the following chapter is the difference in how face-to-face and digital conversation makes us feel – how we experience them. Intuitively it appears to me that face-to-face conversation is quite different to its written counterpart. As Goldberg (2020) observes, most people when asked to think of the most satisfying conversation they have had recently will probably recall a spoken conversation. This isn't to say digital conversations can't be satisfying and engrossing in much the way spoken conversations can often be, they certainly can, but rather it seems that they very often aren't. This despite the fact that we can easily find communities and groups with similar interests, we have a much wider range of potential conversational partners and with hyperconnectivity we often have available to us a potential interlocutor at precisely the point when we might feel the desire to have a conversation with someone.

[147] For more on fake news see for example (Pepp, Michaelson, & Sterken, 2019a) for an explanation of how the particularities of digital communication give us good reason to consider the notion of 'fake news' to be a genuinely new phenomenon. Also see Habgood-Coote (2019) for an argument as to why we should stop talking about it and Pepp, Michaelson, & Sterken's response to Habgood-Coote (2019b).
[148] The reposting of an utterance made by another. See (Marsili, forthcoming)
[149] The use of '#' preceding a word or phrase, see (Scott, 2015; Wikström, 2014)

2 Surface level differences

In this section I consider three of the most obvious differences between spoken and written modes of language. These are differences in acquisition (§2.1), traditional communicative roles (§2.2) and, finally, the differing availability of different communicative channels (§2.3). What I conclude when looking at each of these differences is that considered alone they don't give us much insight into any interesting differences there might be between face-to-face and digital conversations. In fact, I will argue, the first two of these differences need not be particularly consequential for there seems to be no particular reason why the consequences of these differences can't be in some way reconciled. The difference in communicative channels is not quite as straightforward and indeed it is in small part contributory to the more significant differences discussed in Chapter 4. However, taken alone we also have reason to think that this difference too is theoretically (to some extent at least) reconcilable as digital conversation matures.

2.1 *Acquisitional differences*

We acquire spoken language through conversation. But learning to write is not itself learned in conversation it is learned much more painstakingly in a very different environment. When learning to write we learn to associate graphemes with sounds and then learn how to combine those graphemes with others to make words. We must then learn of the idiosyncrasies of these constructions and the formal rules. In learning to speak or sign, though, we ultimately learn how to have a conversation,[150] learning a first spoken language is almost always done through immersion. Sure we have to learn how to mimic sounds, and how to construct words and sentences out of these sounds, but not in the same granular way in which we learn to write. And so perhaps the thought might be that there is something here that can help us explain why spoken and digital conversations differ, perhaps related to the fact that spoken language is acquired in conversation and written language isn't. Indeed, spoken language is a direct product of conversation whereas written language isn't. So we might think, perhaps, this might go some way to explaining the difference between spoken and digital conversation. A linguistic-evolutionary argument for this might be something like the following; spoken and signed language is acquired in

[150] Though on the way to this we might start simply by naming objects, expressing desires, or asking for assistance, the point being that this all leads directly to us being conversational.

conversation to allow people to converse, and so we should expect that it has evolved to perform this role better than other modes.

Another conjecture this might lead us to is related to competency and fluency discrepancies between the modes. Although, of course, levels of articulacy with spoken language varies from person to person, the method of acquisition means that almost everyone has been practicing spoken language for a number of years longer and, in a great number of cases, we might expect that proficiency with the spoken mode of language is higher than with the written mode. Most of *homo sapiens* were illiterate, yet most would have had proficiency with a spoken language. And we might expect that it will only be a small percentage of highly educated people who are equally proficient at writing as they are at speaking. And so the acquisitional differences might also seem to have potential consequences for our proficiency. An argument arising from this might be that because we acquire the modes in such different ways, which we might expect results in greater proficiency in one mode over the other, then we might expect that conversation will be better in the mode with which most are more proficient.

We might think, then, these acquisitional differences and the subsequent potential proficiency discrepancies they cause, may form part of the reason why digital conversation is often experienced differently when compared to face-to-face conversation. It certainly could be the case that being as we acquire one mode through conversation (and to enable us to converse), then that is the mode we perform best in. And so when we converse using that particular mode, we are simply better at conversing than in the mode not acquired through and for the purposes of conversation. However, we should also consider that this could be merely contingent. It may be that these differences amount to nothing. That we learn spoken or signed languages first could simply be explained as being a physiological necessity with no bearing on the suitability of other modes for conversation. A 1 year old simply doesn't have the fine motor skills required to hold a pen, but she might be able to gesture or mimic sounds. And so it could be argued that it is potentially incidental that we learn to speak or sign first. It may be the case that having learned to name objects, express desires, and request help one of the first types of linguistic activity we need to master is something like conversation. If this is the case, then perhaps we just need to do it in any way possible, and as writing simply isn't possible in infancy, we simply use a mode we *can* use. And in terms of proficiency, although it seems difficult to imagine that the way we acquire these modes will alter much, it is not inconceivable to imagine the difference of proficiency

between the modes might balance further over time. As digital conversation becomes more common, deeply established and understood better by the people using it,[151] then any proficiency discrepancy could potentially decrease enough so as to render this point redundant.

This all said, it's difficult to say with any great certainty whether acquisitional differences result in the fact that very often digital conversation doesn't seem to be as immersive as face-to-face conversation. These differences may be contributory - it may be that because spoken and signed languages are developed in conversation in order to be able to have conversations, then they are somehow naturally more suited. But on the other hand, this may simply be explained as being contingent. We might simply acquire the different modes in the way that we do out of necessity, and this has no direct bearing on their suitability for conversation. So this alone certainly doesn't seem to be explanatory enough to tell us why face-to-face conversation is so often more immersive. So let's next consider another point of difference – the different historical roles of the modes – and consider whether this might help us to explain the difference.

2.2 *Traditional roles of written language and spoken language*

In broad terms, historically written language was probably most commonly found in formal institutions - be they governments, courts or churches. At least this is the way in which written examples of language have best survived. As such, literacy tended to be most prevalent in educated elites; be they rulers, clergy, lawmakers or scholars. And as mentioned earlier, literacy itself was historically very low until very recently.[152] Writing was used to lay down the laws of gods and monarchs, or the fiction or treatises of scholarly people for the consumption by other similarly well-educated people. In the Europe of the

[151] In a 2016 survey, for example, the Pew Research Center found the following changes in percentages of adults using social media between 2005 and 2016 (2005 figure appears first). 7% - 86% of 18-29-year olds, 7% - 80% of 30-49-year olds and 4% - 64% of 50-64 years olds use at least one social media site. Even accounting for the fact that some of these may not necessarily be text based services and so may be centred primarily on media such as photographs, and also that not all people will use the services frequently, this still seems to suggest over a period of 11 years highly significant growth. Now it may be that this is the peak of its growth, and social media is only one means of out of range conversation, but even if it does level out the point still stands that as this type of conversation becomes firmly established, we might expect people become more proficient.

[152] Although Claus Wilcke's (2000) study of archaeological findings on the spread literacy levels with one of the earliest known cuneiform script in southern Mesopotamia suggests although kings, priests, conjurors, doctors and soothsayers wrote most frequently and claimed it as a special knowledge, literacy was possibly widespread throughout Mesopotamian society. Though this, of course, does not suppose global literacy was high.

middle ages written language was even quite literally a different language, being as the convention was to write in Latin rather than in the local spoken tongue. Since the invention of the printing press the scope of writing's role increased to include more books and newspapers, for example. But even though there was more writing, it still remained up until very recently the mode of language of authority, be it as laws, scriptures, books, exams, public notices, newspapers, treatises, tests, licences, advertisements, signs and, most authoritatively, philosophical treatises.[153] It has been, in general, a formal and standardised mode of language not available to everyone and often it's rules of construction are much more strictly monitored and policed. The theories of grammar from the late 1800s and early 1900s that Saussure was reacting to were theories of *written* language that gave writing an exalted status. We might then wonder whether a mode with such a history is suitable for conversational purposes. None of these traditional functions or characteristics of formality, elitism and authority make it sound like the kind of medium one would choose as being ideal for a shared joint collaboration such as conversation. Spoken language, on the other hand, has a much more egalitarian background.

Now although spoken language has also certainly played some of the roles above, and many laws and judgements written down for posterity will have been transcriptions of spoken language. However, one obvious traditional role for spoken language not shared by writing is of being the mode of the masses. It is the most common, most used, furthest spread and easiest to acquire. You don't have to have been born at a particular point in history, to wealthy enough parents in a wealthy enough society to learn to speak. And unless spoken under oath or in some other formal capacity, most often spoken language is informal, and unless it is happens to be recorded, its evanescence means that a spoken utterance exists only for as long as it is being uttered.

Clearly these roles are no longer quite so distinct, whereas writing retains its formal roles it also has expanded into being a mode of conversation. We might then wonder whether a mode that has had such a formal and institutional traditional role could be suitable for the much less formal activity of conversation. So the thought might perhaps be that written language has developed in the way it has to fulfil the roles required of it, and if it is now being employed to perform a very different role, perhaps it is less suitable as a mode for conversation in some way. So we might say that writing developed from spoken

[153] To reiterate, I am of course speaking very generally. There will have been examples of writing that weren't any of these things, personal letters, for example.

language to perform a different functional role (so for example, its persistence is useful for laws etc… as opposed to the evanescence of spoken language), though it is inadequate as a substitute for spoken language in conversation. As with the discussion of acquisitional differences above, this could plausibly be the case, but there's certainly nothing that makes it essentially so.

If these historical differences are simply functional, then we should keep in mind that functions can and do change. And in some sense the function *has* already changed – writing *is* being used conversationally. Written language still retains its job description of old as being the mode of formality, but it has added to it a modern function as a tool for conversation. This, in turn, has also lessened some of the previously strict formal rules governing writing. It is commonplace to use abbreviations and grammatically incomplete sentences in digital conversation, for example. So as the functional role of writing is expanding and because this expansion has only started happening very recently, then there seems no strong reason to expect that as the role of writing changes that any differences that might be a result of historical roles couldn't also change.

2.3 *Communicative bandwidth*

One final point of difference we might consider, then, relates to the rich channels of communication available in a face-to-face conversation. Consider, for example, the prosodic and paralinguistic features, which operate above the level of individual phonemes in spoken language. Prosodic features of spoken language include intonation, rhythm, pausing and pitch. Generally, such prosodic features will operate over whole utterances, or sequences of words grouped into prosodic phrases within utterances and refine or be used to alter meaning. These features certainly seem to be important and very useful features of spoken language. Or at least, they are helpful tools of expression. Similarly, paralinguistic features can also have comparable repercussions for what is meant by an utterance. These are features such as tempo, volume and timbre or they could be gestures or facial expressions. These types of feature can give contextual clues about speakers such as their health or emotional state, and even provide sociolinguistic cues such as accent and pronunciation.[154]

[154] This has some positive and negative connotations. It could be advantageous in establishing a common ground - noting an interlocutor's accent might help one to establish commonality of knowledge. Hearing that someone is from the same country, for example, might allow for presuppositions about cultural or political knowledge to be considered common ground. In some ways though, maybe written language has some egalitarian advantages over spoken language in this regard. So perhaps we might consider the role of accents and pronunciation play in implicit biases. In the UK, for example, accents and pronunciation can be a guide to social class, educational

Some of these prosodic and paralinguistic features have useful functional roles beyond meaning, but we might wonder what the loss of such a rich method of modifying meaning entails. So perhaps we might think that the loss of these meaning-modification features in digital conversation may go some way to explain some phenomenological issues frequently notable in online conversation, such as failure to detect irony or jokes, or incorrect ambiguity resolution. This may again lead to questions similar to those raised previously about how well digital conversation can cope with being used conversationally when contrasted with spoken or signed language.

There are a number of examples we might consider that illustrate quite how powerful such features are, but swearing is a good case to consider. Note how many written words it takes Dostoevsky to describe what is effectively a dialogue between six people in his *A Writer's Diary* (1873) in which only one word is spoken six different times in six different ways.

> One Sunday, quite late in the evening, I happened to be walking some fifteen paces away from a group of six drunken tradesmen; suddenly I realized that it was possible to express all thoughts, sensations, and even entire, profound propositions using only this one noun which, besides, has very few syllables. One of the lads first pronounces this noun sharply and forcefully to express his scornful dismissal of something they had been discussing earlier. Another replies by repeating this same noun, but now in quite a different tone and sense-specifically, in the sense that he thoroughly doubts the expediency of the first lad's denial. A third one becomes indignant at what the first has said; sharply and excitedly, he gets into the discussion, shouting out this same noun, but now in the sense of disparagement and abuse. The second fellow again interrupts, angry at the third, who's offended him, and stops him as if to say: "Why do you have to stick your oar in, chum? We've been having quite a discussion here; what d'you mean by getting on to our Filka!" And this whole notion he expressed by using this same forbidden word, this same

---

background (private vs. state funded), and region of origin. All things being equal, one might expect such biases to be less likely to occur where pronunciation styles and accents are removed. Of course there may be a correlate issue if someone's written proficiency, or their adherence to the strict rules governing writing at least, are not regarded to be "proper", in such cases then those biases may still occur. Theoretically though, if a person privately educated and a person educated at a state funded school had similar writing abilities, then there would be no way, unless it is made explicit, to differentiate between their educational background. In Chapter 4 Section 3 we shall look at another way in which the removal of some of the communicative bandwidth could also create better conditions for conversational interaction in cases such as people with ASD.

> monosyllabic name of a certain object, and raised his hand to take the third fellow by the shoulder. But then, suddenly, the fourth lad, the youngest of the group, who had kept silent to this point but who probably had found the solution to the original problem that had caused the dispute, raised his arm and shouted. . . . "Eureka!" you might think. "I've got it! I've got it!" No, it wasn't eureka, and he hadn't got it. He only went on repeating this same noun, not found in the dictionary; just one word, only a single word, but with delight, with a scream of rapture, and, it seems, a little too exuberantly, because the sixth, a morose fellow and the eldest of them, didn't like the sound of it and at once put a stop to the youngster's delight by turning to him and repeating in a gloomy, didactic bass . . . that same noun which isn't mentioned in the presence of ladies and which clearly and accurately signified: "What're you bawling about?" And so, without having said anything else at all, they repeated this same little word of theirs six times in succession and understood one another completely." ([1876]1993, Ch. 13 'Little Pictures')

Presumably upon hearing the dialogue described above a listener might for themselves glean much of the shades of meaning described at length by Dostoevsky because they have access to the way in which this word has been said. And although Dostoevsky is a vivid writer who wants his reader to imagine the scene in the way that he saw it occur, it still seems to be the case that without access to the prosodic and paralinguistic features available in using the spoken word, using the written word to do justice to the shaded meanings this noun takes on in the discourse is a difficult task.[155] As such, it seems to be the case that currently we have a rich set of communicative tools in face-to-face conversations that aren't as obviously available in digital conversation.

Digital conversation is a recent development though; as such, we can't expect that it would have ways of replicating prosody or paralinguistic features when these were not needed previously.[156] So where prosody relates to conventions, the convention may simply

[155] Another example in this same spirit can be found in the TV series *The Wire*. In Season 1 Episode 4, detectives Bunk and McNulty carry out a three-and-a-half-minute crime scene investigation uttering only very differently expressed variations of the word 'fuck'.

[156] Though there is some evidence of prosody in reading comprehension. For example, Janet Fodor proposes the Implicit Prosody Hypothesis: "In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction." (Fodor, 2002); See also (Fodor, 1998; Kentner, 2012; Kentner & Vasishth, 2016), for examples. This does certainly seem to be the case. Pay attention to how you read, for example; 'You *have* just placed a prosodic emphasis on 'have', haven't you?'

not have been replicated in written language yet. Again though, that's not to say it couldn't be, and we can already see examples of this happening, some of which are not even particularly recent. Punctuation such as question marks and exclamation marks, for example, replace prosodic inflections from spoken language. If I want to *really emphasise* an expression to mirror how I would verbally emphasise it I can do so. More recently we see the increased use of emoticons and emoji[157] that can be used to display some of the things previously expressed prosodically or paralinguistically which can make clear when a writer is joking, for example. It's even now possible TO SHOUT in a sense should you so wish.

Now there may be some features of spoken language that we might find it difficult to imagine can be translated to written language. For example, we might wonder how it could ever be possible to translate the musicality of a human voice into writing.[158] So this surface-level difference perhaps cuts a little deeper than the previous two. And recall in Chapter 2 (§3.1.5) the discussion of interactive alignment. There it was suggested that interlocutors make use of different channels of communication in order to align communicatively, and some of these alignments will occur prosodically or paralinguistically. So there could well be some loss of what we might call communicative bandwidth. Though this itself need not be necessarily problematic. It is very possible to have a paradigm type of conversation on a telephone where some paralinguistic features are removed. So too there seems to be no problem for users of a sign language who don't have the specific prosodic features noted above to call upon. [159] So again, although communicative bandwidth is going to form an important part of the overall picture of the difference between face-to-face and digital conversations, this surface-level feature alone doesn't explain why it is that face-to-face conversation is so often much more immersive and enjoyable.

3 Summary

In this chapter we looked at a few possible ways we might explain the difference in phenomena we find when contrasting a face-to-face and digital conversation. As there is

[157] Emoticons use standard typographic characters to make vaguely facial-looking displays, there are many such examples; :) for a smiley face, :(for a sad face, to make it clearer one is joking a winking face such as ;) might be used, if you ever need to digitally blow a raspberry at someone there is :P, and if a shrug is all you can muster at this, then ¯_(ツ)_/¯ could be one way of expressing it. Emoji work in similar ways though are graphical in form and can be used to express such diverse ideas as 'having a haircut' to 'wearing a suit *and levitating' (genuinely!)* to simply giving an a-ok gesture. (See https://emojipedia.org/ for other examples.

[158] Thanks to James Lewis for introducing me to thinking about musicality in conversation.

[159] See (Brentari *et al*., 2018) for some examples of the types of prosody we do find in signed language.

one very notable obvious difference – they both use different modalities, it was suggested that it perhaps follows that it might be helpful to consider the difference between speaking and writing. We then looked at how this distinction between modes can be understood in terms of what I call surface-level differences; acquisitional differences, traditional function differences and the different communicative bandwidths available. These differences do all potentially help us to understand why face-to-face conversation is often more satisfying, though as I hope to convince you in the following chapter, there are some much deeper fundamental differences that tell us a lot more. And indeed considering the relative youth of digital conversation, we can foresee how such differences as outlined in this chapter might be overcome or negated in some ways. In Chapter 4, I discuss a difference that although related to these surface-level differences, is not one that will resolve in time. I will argue we can best understand the difference by considering that the most fundamental contrast between the modes as arising from the different ways in which we are required to cooperate with an interlocutor in order to coordinate at the process level.

4

Two Types of Conversation: Face-to-Face and Digital

In the previous chapter it was argued that there may be a number of contributory surface-level differences that explain why so often written digital conversation (from here shortened to *digital conversation*) is not as satisfying or engrossing as spoken conversations often are. However, it was also argued that each of these surface-level differences are theoretically surmountable. The direction of investigation in this chapter remains the same as it was in Chapter 3 – that is, I am occupied by wondering why it might be that face-to-face conversation seems to be better suited to entering something like, using Goffman's term, an *unio mystico* with a conversational partner. In this chapter, however, we return to themes first discussed in Chapter 2 relating to the necessary tasks that participants in a face-to-face conversation must attend to in order to sustain a conversation. These differences, I argue here, are much more fundamental than the differences discussed in Chapter 3 and as such don't seem to be as potentially reconcilable in the same way.

The core of the argument here, then, is that when we compare how interlocutors in face-to-face and digital conversations coordinate at the process level, we see a distinct contrast in the tasks required of interlocutors. Importantly, these differences have interesting consequences. In face-to-face conversation, the requirements of the media entail that interlocutors must work concurrently to ensure the success of communication. In doing this, face-to-face interactants continuously share the communicative load between them. In digital conversation, however, participants are operating in isolation from each other both geographically and temporally. The result of this is that rather than a continuous sharing of the communicative load, the burden of process coordination is passed from one participant to the next. I argue that these differences are important for three main reasons.

First, I argue that the difference in process coordination entails that despite the absence of rigid response speeds and the requirement of continuous attention in digital conversation, the burdens of process coordination make the tasks of comprehension and language production more difficult, generally, than they are in face-to-face conversation. Second, I argue that we might expect that the greater synchrony of face-to-face conversation when compared to digital conversation has the consequence of making it appear to interlocutors to be more harmonious. This argument recalls the discussions of interactive alignment in Chapter 2 (§3.1.5) and specifically experimental results showing greater levels of perceived harmony and friendliness are reported in synchronous activity. Third, I argue that what this all suggests is that the way in which we cooperate in face-to-face conversation is interestingly different than it is in digital conversation. The basis for this recalls the argument made in Chapter 2 that the intricate and continuous nature of the process tasks in conversation was indicative of a deep cooperation between conversational participants, and as it seems in digital conversation this is not required, then we have good reason to think the nature of the cooperation in the two types of conversation is different. So it is for these three reasons that face-to-face and digital conversation are different in nature and why we might think face-to-face conversation is more conducive to the type of interaction that is a *being-shared-by-one* such as described by Merleau-Ponty (see Ch.1 §3.1). I conclude the chapter on a different note by considering some of the benefits of these contrasting aspects of digital conversation.

1 Coordination in digital conversation

Using the Clark's discussion of features of face-to-face conversation from Chapter 2 (§2.2) as a basis, in this section the features of face-to-face and digital conversations are contrasted. By doing this what becomes clear is that the restrictions placed on interlocutors due to the nature of the media used for face-to-face interaction - features such as the evanescence and recordlessness of speech - and the intricate rapid coordination at the process level this entails, doesn't exist in digital conversation. In fact, coordinating at the process level is very different in digital conversation, and as we see in Section 2, the burden of this coordination rests almost completely with only one participant at any one time. In digital conversation, there is no requirement to attend in real time to an interlocutor as they are making an utterance, there is no requirement to predict how an in-turner[160] will

[160] Recall from chapter 2, it was suggested that 'in-turner' be used to denote the conversational role usually termed 'speaker' and 'out-turner' be used to refer to the role traditionally defined as 'listener'. This allows easier cross-modal comparison and, I suggested there, is respectful of the fact

complete her utterance and there is no pressure to respond within a certain time (or even if there is, it is almost certainly longer than the one second window discussed in Ch.2 §3.1.1 allows). So there is no requirement to do anything concurrently with our conversational partner in digital conversation, and as is discussed in Section 3, this has some interesting consequences. For now, though, let's begin to look at the different features of the two modes by considering a simple structure of digital conversation in contrast to face-to-face conversation.

1.1 *A simple chronological structure of digital conversation*

Recall in Chapter 2 (§2.1), Table 1 was offered to represent the concurrence of the required tasks placed on in-turners and out-turners.

(1)

	Turn marker	
	01	02
Person	Necessary conversational task	
A	Making utterance	Comprehending B's utterance Formulating response———————>
B	Comprehending A's utterance Formulating response———————	——> Making utterance
	Time ——>	

Table 1: A simple structure of face-to-face conversation and its necessary tasks.

In-turn
Out-turn

In Table 3 we see a representation of the chronology of the required tasks in digital conversation (I divide turn 02 into sub-turns to show more clearly how the same tasks shown in Table 1 become spread out).

(2)

that both of these roles are much richer than that of merely speaking or listening (or signing and seeing or writing and reading).

	Turn marker			
	01	02-a	02-b	03
Person	Necessary conversational task			
A	Making utterance	n/a	n/a	Reading B's utterance
B	n/a	Reading A's utterance	Making utterance Formulating response	n/a
	Time ——>			

Table 3: A simple structure of written conversation and its necessary tasks.

In-turn
Out-turn

As we can see from Table 3, once the in-turner completes a turn she no longer has any necessary tasks to perform (02-a & 02-b). Of course there may be some tasks such as checking for a response, but beyond that there are few requirements placed on the out-turner. Now of course, this isn't to say she can't attend in some way to the conversation. At 02, A could be considering what B might potentially utter in response, she might reread some other parts of the conversation, or she may be considering what she herself might want to say later. Importantly, though, there is no *requirement* to do so. The synchrony and symmetry of attention to the conversation is no longer necessarily present as it is in face-to-face conversation. Whereas in face-to-face interaction at any one time during a conversation the participants must be fulfilling some specific coordinative role essential for the conversation to function, in digital conversation the necessary tasks are left to the in-turner to perform alone. And note too that the absence of synchrony also moves the comprehension of utterance from being part of the role of the out-turner to being part of the role of the in-turner – it is only once the previous in-turner has finished her turn that the new in-turner receives the most recent contribution to the conversation. So the in-turner has all the work to do and the out-turner no longer has any responsibilities towards the maintenance of the conversation.[161] To help us think about this further, let's revisit again Clark's groups of features of face-to-face conversation first discussed in Chapter 2 (§2.2).

[161] Beyond, perhaps, checking for a reply, though in many instances this may be automated for them by way of, for example, notifications. Notifications work by sending a message to an interlocutor informing them a response has been made.

1.2 *Clark's features of conversation*

I present these features with only short commentary on the immediacy and medium-specific groups of features as it is here we see the greatest contrast between face-to-face and digital conversation. But it is the consequences of the difference in these features between face-to-face and digital conversation that are of most interest for the present chapter, and the discussion of this will be taken up in Section 2. I say more on the control features group as here the differences are less pronounced and indeed it is the parallels in this group that allow us to see the similarities between digital conversation and the type of paradigm type of face-to-face conversation discussed in Chapters 1 and 2.

1.2.1 *Immediacy features*

Recall the following features of face-to-face conversation taken from Clark (as discussed in Ch.2 §2.2);

Copresence:	the participants share the same physical environment.
Instantaneity:	the participants perceive each other's actions with no perceptible delay.
Visibility:	the participants can see each other.
Audibility:	the participants can hear each other.[162]

It was argued in Chapter 2 that not all of these features need necessarily be present in a face-to-face conversation, though very often most will be, and it will always be the case that at least some of them are present. Yet in most cases of digital conversation, none of the immediacy group features apply. Naturally, in a digital conversation most interlocutors will be remote from each other and so it follows they won't be copresent. As interlocutors in a digital conversation don't share the same physical environment, this has limiting effects on the things they can use to communicate with each other; that is, they can't use the environment and shared knowledge of the environment to aid communication.[163] Note too, that the absence of instantaneity in digital conversation entails that interlocutors also can't use the immediate reactions of an interlocutor to gauge communicative success or uptake of speech acts in real time in the way that is possible in face-to-face conversation. And as is

[162] (Clark, 1996, p.9)
[163] Thinking in terms of the common ground, then, we might say that the available common ground is reduced.

almost always the case, in digital conversation interlocutors usually won't see or hear each other, and so, as discussed in Chapter 3 (§2.3), the communicative bandwidth is comparatively reduced with the absence of many of the gestural, prosodic and paralinguistic features available to face-to-face interactants. Without these features of the immediacy group, then, in a digital conversation between Eliza and Parry what is available to the interlocutors is a relatively impoverished set of communicative tools. So, for example, Eliza might only directly have access to Parry's username, Parry's avatar[164] and Parry's previous written utterances (and vice versa).[165] We shall return to consider some of the consequences of the absence of such features later, but next let's think again about the medium-specific features of digital conversation in contrast to face-to-face conversation.

1.2.2 *Medium-specific features*

Recall the features of this group when applied to face-to-face conversation;

Evanescence: the medium fades quickly.

Recordlessness: there is (usually) no record of actions.

Simultaneity: participants can produce and receive at once and simultaneously.[166]

In almost all examples of face-to-face conversation this set of features will all be present; however all of the features of this group will be generally absent in digital conversation. Almost all written utterances persist and will be recorded (be it on data servers or on the devices of the interlocutors). Once a digital utterance is written and sent, it can be read and reread at any time and as many times as desired by the recipient(s).[167] And although, theoretically, participants can produce and receive at once and simultaneously, the lack of instantaneity and the narrowing of communicative bandwidth noted in Chapter 3 (§2.3), also entails that there will be at least some delay in these responses being received. So in

[164] Usually a pictorial representation of a person online.
[165] Of course Eliza and Parry might be familiar acquaintances and have a rich common ground of shared knowledge they can draw upon too.
[166] (Clark, 1996, p.9)
[167] There are some exceptions, apps such as Snapchat found initial appeal in that messages sent on the platform could be made to 'self-destruct' 10 second after being sent, and the app includes features that made it difficult for the receiver to create duplicates of messages. Though Snapchat is primarily used to send photos and videos, it is also possible to send written messages. Other social media platforms, such as Twitter, also allow users to send messages that expire after a certain time period.

some sense it appears that in digital conversation interlocutors are freed from the constraints imposed by the media used for face-to-face conversations. It is notable, however, that this freedom of response-speed constraints is also a freedom from having to necessarily commit continuously to a digital conversation. Due to the *permanence* and *recordedness* of written language, in a digital conversation there is no requirement to attend to an utterance just as it is being produced (and indeed this is generally not possible). A response could take hours and so often it will be the case that we are not required to dedicate attention to the task of the conversation in the way we are required to in face-to-face conversation, or if we do, then it need only be merely intermittent attention.

Recall the following passage from Clark (first presented in Ch. 2 §2.2) that typifies some of the necessary requirements of face-to-face conversation;

> If Roger is to succeed in telling Nina something, he must make sure she is trying to attend to his sounds *at the very instant* he is articulating them. Executing behaviours to be attended to and attending to those behaviors, then, are participatory acts: Roger cannot do his part without Nina doing hers, and vice versa.[168]

An important feature of face-to-face conversations is that interlocutors are as (roughly, at least) committed to the conversation as each other, and at the same time as each other. Building on the description of a paradigm type of conversation as an interactive and synchronous activity suggested first in Chapter 1 (§2), in Chapter 2(§2.3) I argued that these features of face-to-face conversation give us reason to see that the continuous interactivity we find in face-to-face conversation is not merely a result of each turn following another. Rather, the interactivity of face-to-face conversation is continuous due to the medium-specific features of evanescence and recordlessness, and so interlocutors must enter into an interactive synchrony, with each conversational participant performing the necessary tasks required to fulfil their respective roles. It is required for the continuation and success (in terms of communication, if not interpersonal harmony) of face-to-face conversation that Roger and Nina are both fulfilling their roles concurrently. Yet the absence of these medium-specific features from digital conversation make it so that the interactivity of digital conversation is different in character. It is interactive in that one turn follows another, but there are no requirements of synchrony or continued attention. And so what we might think this entails is that the nature of the necessary commitments to the

[168] (1996, p.275)

conversation interlocutors must make in digital conversation is different. This need not always the case of course. Although strict synchrony might not be possible, participants in a digital conversation could theoretically remain equally focused on a digital conversation as face-to-face interlocutors. They might wait patiently for a reply, attempt to predict what an interlocutor might say, read a response as it is received and then immediately and rapidly respond. The important difference, however, is that there is no *requirement* to do so in a digital conversation whereas in face-to-face conversation such coordination and commitment is essential.

The upshot of this is that the nature of process coordination in digital conversation is quite different to that which we find in cases of face-to-face conversation. There is no need for an out-turner to attend to an in-turner as the in-turner makes her utterance, and there is no pressure or requirement to respond immediately. This isn't to say there is no coordination of processes in digital conversation, there is, but as we shall see in Section 2.3, the burden of this coordination need rest only with one of the participants at any one time and we might suppose this has interesting consequences for the type of activity digital conversation is. Next, though, let's recall Clark's final group of features – the control features.

1.2.3 *Control features*

Recall this group relates to the control participants have in a face-to-face conversation.

Extemporaneity: The participants formulate and execute their actions extemporaneously, in real time.

Self-determination: The participants determine for themselves what actions to take and when.

Self-expression: The participants take actions as themselves.[169]

Unlike the features of the previous two groups, the set of control features is quite similar in both face-to-face and digital conversations.[170] Before looking at each feature in turn, it is worth noting first that each of these control features are perhaps best considered not as binary notions, but as spectral notions. So in the case of a feature such as copresence (from

[169] (Clark, 1996, p.10)

[170] We should expect this to be the case too, recalling the discussion of Chapter 2 §4.2.2 the types of interaction we are interested in here are generally more *synocratic* in nature than autocratic.

the immediacy features group), for example, we could state some simple principle that allows us to give a binary answer to a question of whether interlocutors are copresent or not. Naturally, there will be a few examples where interlocutors are at the borderline of copresence, but generally speaking it seems right to say that most interlocutors in a face-to-face conversation will be copresent and in digital conversation this most often won't be the case. However, in the control category of features the differences are perhaps better considered a matter of degree and with the exception of extemporaneousness (although both types of conversation still have this to some degree), the specifics of which type of conversation has more or less of the features will be relative to the specific conversations being contrasted (as opposed to the more general *type* of conversation being contrasted).

Let us for now consider extemporaneousness to be a matter of degree. This being the case, in many cases of digital conversation there will be some degree of extemporaneousness; for example, we might respond to our interlocutor's utterances as soon as we receive them, and we will use our interlocutor's utterances as a guide for our response to them. And even if we don't respond instantly, this wouldn't preclude the response itself being any less extemporaneous. Though in another sense we might argue that digital conversation is less extemporaneous than its face-to-face correlate. The permanence and recordedness of written utterances and the leniency of expected response times give a writer much greater scope to plan how she might respond to an interlocutor's utterance. As will be discussed in more detail in Section 2.3, when composing a contribution an in-turner might write, reread, rewrite and edit an utterance before she eventually sends it to an out-turner. Recall that face-to-face time-constraints and coordination requires that utterances are produced almost as soon as they are initially planned (see again Levinson's preconditions, Chapter 2, or present chapter §1.4).[171] As such, we might think that face-to-face conversation is often *more* extemporaneous than its written counterpart, however this need not necessarily be the case, and many digital conversations will involve at least some degree of extemporaneousness.[172]

[171] Though of course it should be noted too that in a trivial sense all utterances are at least to some degree planned. Even in face-to-face conversation as described in Chapter 2 there are planning stages in preparation for a turn. Think again about the Precondition to the Cooperative Principle (PCP) discussed in Ch.2. It could be argued that fulfilling such a precondition is a form of planning. Hence why it is best to consider extemporaneousness to be a matter of degree.

[172] It is worth noting that it is possible to contribute something we might consider to be a planned utterance to a face-to-face conversation. Think of the conversational participant offering up a well-rehearsed anecdote, or the response of a researcher when asked what they are researching.

The differences in the role of self-determination and self-expression in the two types of conversation is murkier still. In terms of self-determination, the absence of the one second window in digital conversation and the various pressures of response it entails allows that participants in a digital conversation are able to consider more carefully what they want to do and when they want to do it. In digital conversation an in-turner is often afforded the opportunity to draft, redraft and edit her contribution to a conversation. As such, we might therefore think that writers are given opportunity to consider more carefully what they wish to do or say. So perhaps it is the case that, in some senses at least, digital conversation offers an even greater degree of self-determination. Though in another sense we might think there are limitations placed on this by the particularities of the social media platform being used. So for example, this might be in the form of character limits, such as found on Twitter,[173] which place restrictions on the length of an utterance. There might also be limitations placed on the content of an utterance. For example, most social media platforms have community standards guidelines and to be in breach of them can result in a post being deleted or even an account being suspended or deleted. There might also be technology-imposed restrictions, for example not having access to a mobile signal, the internet or an electronic device. So it won't always be the case that participants in a digital conversation are able to self-determine their actions. What I think this suggests, then, is that although there are some granular differences, there is no striking difference between the two types of conversation in terms of self-determination. This is similarly the case for the final feature in this group – self-expression.

There are some examples where we might consider there is a greater degree of self-expression in digital conversation; perhaps we might consider anonymous media wherein there may be few social penalties for saying things one might not say in a face-to-face conversation.[174] It could be that in such a situation a conversational participant might believe herself to be afforded greater freedom to take actions as herself. Even if a user is not completely anonymous in the sense of using a pseudonym that is difficult to link to the contributor's life beyond the electronic media they are engaging with, there can still also be

[173] There are, of course ways around such limits, one could make clear a tweet is part of a series, for example by including '(1/6)' or similar to indicate it will be in six parts.
[174] Indeed, this may not even need to be anonymous, although anonymity (or relative anonymity at least) may be a disinhibiting factor, there are potentially other factors such as described by John Suler (2004) who suggests the existence of an 'Online Disinhibition Effect' created out of a mix of anonymity, invisibility, the minimisation of authority, "solipsistic introjection", "dissociative imagination" and asynchronicity.

a relative level of anonymity in some sense.[175] For example, a user might use her own name and picture as an avatar, but be a member of an online community with no connections to her life beyond the website. Though self-expression might be restricted by other features of the media being used. The recordedness and permanence of written language might have a stifling effect on the types of actions conversational participants might take. It may make some conversational participants more cautious in what they utter. And that a digital conversation will very often be more public than a face-face conversation (and so may have greater the potential for unintended audiences to see an utterance) could also impose similar restrictions. Though again, as with self-determination, these are speculative differences, and ultimately none are convincing enough that we should think there is a profound difference in relation to self-expression between the two types of conversation.

Clearly, then, there are some differences between the two types of conversation relating to the control group. However, unlike in the case of the immediacy or media-specific groups, the differences aren't ones of absence of a feature (such as copresence's absence in digital conversation) or that one has a feature contradictory to the other (such as evanescence in face-to-face vs. permanence in digital conversation), the differences are linear. That is to say, the control features *are* present in both types of conversation, though how they are instantiated may differ in certain of the specifics between the modalities. These, then, don't seem to be essential differences in the way many of the previously discussed features were. And indeed, we should expect that how these features apply in any particular conversation, be it face-to-face or written, might vary in comparison to another particular conversation. For example, context might place impositions on the amount of self-determination participants have in a particular conversation. We might think of a case in which disgruntled colleagues are in an office discussing their workload within earshot of a tyrannical boss, those conversational partners might not have the same level of self-determination as similarly frustrated co-workers meeting after work away from the office. And as discussed in Chapter 2 Section4.3, power dynamics will vary from one conversation to the next, and so too we would expect the levels of self-determination and self-expression will fluctuate accordingly.

What I think the comparison of control features highlights, then, is that despite its differences digital conversation *does* share important similarities with face-to-face

[175] See Goldberg (2013) for discussion on how assertions made anonymously affects the epistemological aspect of their production by in-turners, and their reception by out-turners.

conversation. Excepting some finer-grained contrasts, the features of the control group will often apply similarly in cases of digital conversation and face-to-face conversation. For current purposes this is a good thing, for it is the similarities in this group that allow us to say that both face-to-face conversation and digital conversation play similar communicative and social roles – both of these types of interaction seem to be conversation in the restricted sense I use in Chapter 1. In paradigm cases of either we might expect that the participants of the conversation have some degree of control on the direction of the conversation and it is this synocratic structure that differentiates these types of interaction from the more autocratic types of linguistic activity such as the academic lectures, job interviews, and novels discussed in Chapter 1. For the purposes of this chapter, then, we need not say much more about the control features, for the overall aim here is to characterise the significant differences between the two types of conversation.

1.3 *Summary of contrast between Clark's features in face-to-face and digital conversation*

It was argued in Chapter 2 that the requirements placed on participants in a face-to-face conversation are a direct result of a combination of some of the features of face-to-face conversation Clark highlights. And so from the discussion above, it seems there are some important differences between face-to-face and digital conversation emerging out of these differences. Interlocutors in a digital conversation are not copresent and perceiving each other's actions with no perceptible delay. Digital conversational partners most often won't be able to see or hear each other, with the implications for the communicative bandwidth that entails (see Chapter 3 Section 2.3). As digital conversational utterances are generally persistent and recorded, then it is also the case that there are no requirements placed on out-turners to attend to the utterance of the in-turner just at the moment she is making it.

As was also argued in Chapter 2, the requirement to attend to each other concurrently in a face-to-face conversation entails that conversational partners must engage in an intricate set of process coordination tasks. This intricate series of tasks are also most often carried out under the time-constraints looked at in Chapter 2 (§3). Most pertinently, participants in a face-to-face conversation are constrained by the requirement to answer within a time limit lest the meaning of their response be adjusted by an unconventional delay (sometimes conceived as 'the one second window'). Added to this, average response speeds in face-to-face conversation are estimated to be approximately 200 ms (Ch. 2 §3.1.1). So when we consider too the latency we find with language production (Ch.2 §3.1.2), it seems that to operate at such speeds in face-to-face

conversation some of the primary tasks out-turners must engage with relate to prediction. So, as discussed in Chapter 2 (§3.1.4, §3.1.6) they must predict speech act type, remaining content and turn duration.

One major consequence that emerges from the different features of face-to-face and digital conversation is that in digital conversation correlative time constraints no longer necessarily apply. As such, there is also no requirement to predict what an interlocutor will utter. Now recall in Chapter 2 (§3.1.6), there it was argued that prediction, and the requirement to predict, are indicative of the continued perspective sharing required for face-to-face conversation to proceed. We might expect, therefore, that this difference has consequences for the type of perspective sharing we find in digital conversation. For if the time constraints in digital conversation are not as rigid, so too the requirement for prediction is diminished. One consequence of this diminishment of the role of prediction, we might expect, is that the nature of perspective sharing in digital conversation is different. So it appears the nature of digital conversation both at the process level and in how interlocutors share perspectives, is notably different to that of face-to-face conversation. Many of the necessary process tasks discussed in Chapter 2 (§3) are simply not required in digital conversation. And we can see how this differs if we consider again the preconditions Stephen Levinson suggests hold for each turn of a face-to-face conversation.

1.4 *Levinson's preconditions*

Recall Chapter 2 (§3.2), there I presented Levinson's suggestion for the preconditions that exist in a conversation between A and B to enable out-turner B to respond to in-turner A within the average 200 ms timescale:

> (3)
>
> (i) B must attempt to predict the speech act (detect whether A's utterance is a question, offer, request, etc.) as early as possible, because this is what B will respond to;
>
> (ii) B should at once begin to formulate a response, going through all the stages of conceptualization, word retrieval, syntactic construction, phonological encoding, articulation;
>
> (iii) meanwhile, B should use the unfolding syntax and semantics of A's turn to estimate its likely duration, listening for prosodic cues to closure;

(iv) as soon as those cues are detected B should launch the response. [176]

Notable from this, then, is that very little of what we might think must hold for face-to-face conversation to function need hold for digital conversation. And even where it does exist, it is distinct in nature due to the different response-speed time constraints.

Consider (3i), in almost all digital conversations an utterance will be received by an out-turner in its completed form.[177] Assuming it is written clearly enough, then there should be no need to *predict* the speech act in the way that Levinson intends here, it should be derivable from the complete utterance. Though it is worth noting here a point of ambiguity with the expression 'predict', for it can be taken to be synonymous with a few notions, in particular it is synonymous with 'forecast' but also with 'calculate'. This is worth mentioning as it will often be the case that a speech act will not accord precisely to *what is said*[178] by an utterance. Consider cases of indirect speech acts such as the following famous example from Searle (1975);

(4) Can you reach the salt?

What is said by (4) is an interrogative, however, in most contexts we would interpret (4) as a request (to pass the salt). Perhaps then, there may be a question about what it is that occurs when (4) is taken to be a request and not an interrogative – specifically we might wonder if it is a form of prediction. If we take 'predict' to be synonymous with 'calculate' then this is perhaps a case that could be made. It seems right to say that we can *calculate* from (4) that despite its form it is meant not as a question but rather as a request, and in this sense we might say it is a 'prediction'. However, it seems clear to me that Levinson is using 'predict' as being synonymous with something like 'forecast'. That is, 'prediction' used here is meant in the sense of being an estimate of some future occurrence. As such, even though in both face-to-face and digital conversation interlocutors might be required to calculate a speech act (or an implicature) based on what *has* been said, it is only in face-to-face conversation that participants will also often be required to forecast a speech act based on what *will* be said. And I argue next that this seems to be an important difference.

[176] (Levinson, 2016, p. 7)
[177] Of course, an in-turner might break her utterance up into smaller chunks, perhaps ending one chunk with an ellipsis. Though in such cases there is still no particular *requirement* that an out-turner predict what might come next (though naturally they could do this).
[178] 'What is said' here is used in the Gricean sense as being something like the conventional meaning of a sentence, see for example (Grice, 1989f, pp.87-88).

We might argue that deriving indirect speech acts or implicatures in both types of conversation is a similarly perspective-sharing practice – that is, we might think that in order to infer what is *meant* by some utterance beyond what is said by it we might need to consider, to some degree, the perspective of an interlocutor. And this may well be the case.[179] However, when we consider that in face-to-face conversation an out-turner is often required to perform this task during the in-turner's production of an utterance, the perspective sharing becomes concurrent – an in-turner must attempt to share the perspective of her interlocutor, just at the moment the in-turner has this perspective. I take this to be significant, but as it stands there is a worry that this might seem arbitrary – we might not think a temporal gap has any bearing on the nature of perspective sharing. When we look at the burdens of process coordination in Section 2.3 I hope to persuade you that this is significant, particularly when considered in relation to trying to understand the difference in the way we experience the two types of conversation. But now let's return to Levinson's preconditions.

We should expect that (3ii) does still hold to some degree, though it will be different in nature. The absence of conventional time-constraints allows that there is no pressure to begin 'at once' the processes of response formulation. A recipient of a written utterance is afforded time to read and reread it, and typically it is only when she chooses to respond that she would need to begin the tasks associated with (3ii). And again, as we see when looking at the burdens of language production in Section 2.3.3, these tasks will differ in nature due to the relatively solitary nature of production found in digital conversation. Let's now consider briefly the final two preconditions.

It will most often be the case that the prediction tasks of (3iii) are simply not applicable to digital conversation. Generally, a turn ends when an interlocutor sends her utterance to the other participant(s), and so it is unlikely participants need to predict turn-endings to ensure they can launch their own response at an appropriate time. And finally, as there is generally no necessary pressure to respond in a timely manner, (3iv) need not take the form suggested by Levinson.

The purpose of considering these preconditions again is to draw the contrast further as to the different types of process task we might expect are required of a digital conversation when compared to a face-to-face conversation. Now we may find examples of

[179] Though I think when we consider the burdens of comprehension in §2.3 this gives us reason to think this is notably different in nature too.

digital conversation where these preconditions might hold in a form closer to Levinson's, but recall from Chapter 2 (§§3.2-3.3) these preconditions we should expect to hold for *most*, if not all, face-to-face conversations and such examples in digital conversation are outliers. Remember, too, that based on an average turn duration of 2000 ms and an average turn interval of 200 ms, then if something like Levinson's preconditions are required for a response in face-to-face conversation, they will often be carried out within 2200 ms. Clearly, in digital conversation, it will be rare that responses will begin so rapidly.

We might think the upshot of these differences is that it frees up cognitive space for participants in a digital conversation to attend more closely to what an interlocutor is trying to communicate. The idea being something like; more time to consider an interlocutor's utterance provides greater opportunity for understanding it. If participants in a digital conversation don't have to carry out concurrent comprehension and production tasks, it would perhaps follow that they can devote more energy and time to these tasks individually. And as participants in digital conversation aren't constrained by the one second window, we might also think that they are able to better formulate responses. The thought being that because we have more time available to consider a written response, this might, in turn, give us opportunity to better construct our utterances. I will argue in Section 2.3, however, that this needn't necessarily be the case. Because even if it were true that digital conversation does indeed afford us more opportunity to better comprehend the utterance of an interlocutor and to formulate our responses, it is also true that this is done at a greater remove from an interlocutor and this too has its own consequences. The lack of synchrony between interlocutors in digital conversation, their lack of copresence, and the absence of rigid response speeds results in participants in a digital conversation doing much of this work alone as opposed to synchronously. And this might lead us to wonder what this means for the type of cooperation we find in digital conversation when compared to the type we find in face-to-face conversation (as discussed in Ch.2 §4).

2 Coordination and cooperation

We arrive now at one of my central claims; that what makes face-to-face and digital conversation interestingly different is the contrast we find between the required process tasks in these two types of conversation. It was argued in Chapter 2 (§4), that the type of coordination we find in face-to-face conversation is indicative of the type of cooperation required to sustain it. That is, due to the features and time constraints of face-to-face conversation interlocutors must almost continuously share a goal of sustaining the

conversation and must work continuously to achieve this goal. The result being a deep and continued cooperation between interlocutors. It was also argued in Chapter 2 (§4.7), that we might expect that the close link between coordination at the process level and coordination at the content level therefore means that this process cooperation creates conditions in which interlocutors might be more likely to cooperate at the level of content.

What I argue here, then, is that the different nature of process coordination in digital conversation is such that the level of required cooperation is also different. The basis of this argument is that it appears that process coordination in digital conversation is necessarily more solitary in nature than it is in face-to-face conversation. This is not significant simply because interlocutors are geographically and temporally isolated from each other which we might reasonably expect will result in them feeling somehow more remote from each other. What I take to be most interesting about this separation, however, is the consequences it has on the tasks of comprehension and production; it entails that the burden of process coordination will rest most heavily at any one time with just one conversational participant. And what this means is that the type of cooperation is different – interlocutors in a digital conversation will still often be cooperating, however this cooperation need not be continuous but rather merely fleeting. So it is when we consider these differences in required cooperation between face-to-face and digital conversation that we start to see the fundamental difference between these two types of conversation. If, as I intuitively take to be the case, we generally think of cooperation with another person as being a harmonious experience, then we might expect feelings of harmony generally increase relative to an increased sense of cooperation.[180] When we add to this the experimental findings (first discussed in Chapter 2, §3.1.5) that alignment and synchrony with another person also increases perceptions of harmony and friendliness in an interaction, we understand why it is we so often seem to enter more freely into an *unio mystico* with an interlocutor in face-to-face conversation than we do in digital conversation. And unlike the surface-level

[180] As an empirical means of testing the hypothesis that the specific requirements of process coordination in face-to-face conversation make it appear more cooperative it might be interesting to contrast cases of disputation. For on the above account it would be expected that the deeper level of cooperation we find in face-to-face conversation should make it more likely that disputation would be less likely to escalate than in that case of digital conversation. I have no specific data on this, though many observers of internet discourse have wondered about the apparent polarising effects it has. I make no claim that this supports my thesis, but rather note it as a potentially interesting line of empirical research that might follow.

differences discussed in Chapter 3, (§2), there seems like no obvious way in which this particular difference might be reconciled.

2.1 *Similar but different: perspective sharing and prediction*

What I want to stress here is that although face-to-face and digital conversations have many similarities and some obvious differences, it is the consequences of the differences that hint towards the real distinction between the two types of conversation. What I want to suggest is that when we contrast face-to-face and digital interactions we have two types of conversation and each has its own set of process tasks required to sustain them. Relevant to a paradigm case of face-to-face conversation are the concurrent, predictive tasks such as those suggested in Chapter 2 Section 3, and relevant to a digital conversation are the more solitary tasks of comprehension and formulation of replies (more of which in §2.3).

As discussed in Chapter 3 (§1.1) and here in Section 1.2.3, face-to-face and digital conversations are, in many respects, similar – they can involve the same number of participants, address the same topics and progress in similar synocratic ways wherein each contributor is afforded a symmetrical influence on the direction of the conversation. And as discussed in section 1.2.3, the difference in control features between the two conversations is only really by a matter of degree. However, that digital conversation isn't synchronous, and that the predictive requirements are different between face-to-face and digital conversation has some important consequences. Recall that it was argued in Chapter 2, (§3.1.6), that there is an aspect to paradigm cases of face-to-face conversation in which both the out-turner and in-turner must all be arriving at the same (or similar) thoughts at the same time. And that this can be best considered to be a form of perspective alignment. What is clear, however, is that in a digital conversation between A and B, A doesn't need to predict what B might say, and this is indicative that the type of perspective sharing it involves is quite different to the perspective sharing we find in face-to-face conversation.

Recall now the discussion of what we might call the dynamic facts of face-to-face conversation which were derived from Grice's cooperative principle first discussed in Chapter 2 (§4). These dynamic facts were presented as an illustration of the types of information an out-turner in a face-to-face conversation must keep track of in order to be

able to adhere to something like the cooperative principle.[181] The information composing these facts can be derived from the cooperative principle as follows;

(5)

d) what is required of a contribution;
e) what stage the conversation is at;
f) what the accepted purpose or direction of the talk exchange is.

What I suggested when initially discussing (5) is that in producing an utterance an in-turner alters these facts about a conversation. And in order to meet the cooperative principle when her turn begins, an in-turner will be required to track the changes to these subtly different facts. Though the speed of face-to-face conversation also entails that an out-turner can't merely *keep track* of these facts, she must also predict what they might be when her turn begins. What I hypothesise then is that perspective alignment plays an important role in being able to do this. So we might expect that to effectively keep track of facts such as those in (5), requires the prediction of the speech act type, of how an utterance will proceed, and of how it might end.[182] And so in a face-to-face conversation, doing this most effectively relies on an out-turner attempting to at least momentarily take an in-turners current perspective. That a face-to-face conversation requires rapid turn-taking and the continuous attention of interlocutors to the task of sustaining the conversation places upon them demands that aren't paralleled in digital conversation. In digital conversation there is no conventionalised required response speed, there is no necessity to attend to an utterance as it is being uttered,[183] and there is no requirement to predict an interlocutor's speech act or the remainder of her utterance just as she is making it.

2.2 *An arbitrariness worry*

We still haven't reached the point of being able to say that the difference in response speeds in the two types of conversation and the effects these differences have on the process tasks required of participants are necessarily substantial. The overall claim in this chapter is that due to the difference in process coordination between face-to-face and digital conversation there is a difference in the type of cooperation required of these two types of conversation. Perhaps, though, we might still worry it is arbitrary to say that because digital conversation

[181] Again, it is worth noting that these types of facts are constituents of the common ground.
[182] Recalling Levinson's preconditions Section1.4.
[183] Which if we take the typing of the message to be the making of the utterance, would be impossible in most cases.

is more temporally spread that this makes it fundamentally different to face-to-face conversation. We might also wish to argue that because, for example, the necessary tasks required for the continuation of a digital conversation rest only with one participant at any one time, that this doesn't give us sufficient reason to think it is of a fundamentally different nature to face-to-face conversation. Let's consider an example to illustrate this worry.

Correspondence chess is a game of chess in which players remotely send moves to each other (perhaps stretching a single game out over years). Yet despite the obvious temporal difference and the lack of copresence, there doesn't seem much reason to suppose that correspondence chess is somehow interestingly and fundamentally different to a game of *over-the-board* chess. That correspondence chess will be more temporally spread, and that it doesn't require its players to be concurrently attending to a particular game doesn't seem to alter the nature of the game in any interesting way. It may be that there are some advantages in correspondence chess, such as granting players longer to plan their moves, which may be strategically beneficial.[184] On the other hand, it may be that not being copresent with an opponent has other consequences too. For example, having an opponent present as moves are being considered might add extra pressure, and even if opponents are silent there are numerous ways they might be communicating with each other.[185] None of these differences appear to be substantial though, it would certainly seem to be a stretch to claim that beyond the most obvious differences between the two types of chess that there is a genuine distinction. And so if we were to say that it is cooperative to play a game of chess in the sense that both players commit to a goal of playing the other at chess and both fulfil their necessary roles, then it might seem arbitrary to say that over-the-board chess is somehow *more* cooperative than correspondence chess. Applied to the topic of this chapter, then, the argument might be that the structural differences we find between correspondence and over-the-board chess are similar to those we find in written and face-to-face conversation, and as such are equally inconsequential.

[184] A player may even use other means to aid them that wouldn't ordinarily be permitted over-the-board, such as using computer software, or consulting a friend - though this is rather more a conventional difference rather than a necessary one. It would be equally possible to use such means *over-the-board*, though one's opponent might not be too pleased.

[185] And, indeed, there may be more instances of prediction of an opponent's move in an over-the-board game, as white awaits black's moves she might be trying to anticipate what will happen so as to speed up her own turn, especially in a timed match. Though this is not a necessary requirement such as in face-to-face conversation.

One possible disanalogy here, however, is that unlike in a game of chess, conversational partners are not in competition – that is, despite their subsidiary aims and goals within a conversation (which may well be competitive in some senses), a primary aim will be to successfully communicate with each other. This being the case, it is in conversational participants' own interests (so long as they wish to sustain the conversation) that they help each other to complete the tasks required for conversation to proceed as smoothly as possible. Now of course in chess it is essential that each player makes a move if the game is to continue, and this possibly correlates to the required process tasks of face-to-face conversation. Beyond this, however, chess is fundamentally adversarial, and players work alone to decide the best moves for them personally.[186] Therefore, it is of little interest to the players that they help each other with a turn.[187]

On the other hand, even in an adversarial conversation (face-to-face *or* written) - for example, think of two people trying and failing to convince each other on some matter of politics - it is still in the interests of both parties that they understand *what* each other is saying in order to respond to it.[188] And to do this requires that they help each other with the necessary tasks of the conversation. Therefore, the required coordination at the level of the process tasks remain unchanged in a confrontational conversation from that which is required in a case of harmonious talk.[189] In chess, either correspondence or over-the-table, the nature of a move is that a player deliberates alone to make a move which is beneficial to her, and so whether this spreads out temporally and geographically doesn't alter the nature of what is required. Although moves are often reactive to those of an opponent, a 'turn' in chess is generally a solitary activity whether it is an over-the-board or a correspondence game. However, as we shall see next, when conversation becomes spread temporally and geographically, the burden of process coordination changes in important ways. For in face-to-face conversation interlocutors (even in adversarial conversation) will spread this

[186] Though perhaps we might think there is an element of perspective taking if one were to try to anticipate an opponent's moves.
[187] An exception to this might perhaps be if a chess tutor were playing a tutee.
[188] Although we might all be familiar with such exchanges where interlocutors clearly aren't listening to each other, or even if they are, they don't really respond to what each other is saying. Thinking again about paradigm cases of conversation discussed in Chapter 1, we might think such exchanges are at some remove from the paradigm case. The nature of such discussions becomes closer to monologues.
[189] Though contra this, it has been observed that arguments do, to some extent, disrupt interpersonal synchrony, see for example (Paxton & Dale, 2013).

burden between them, whereas in digital conversation the burden falls almost entirely on the in-turner.

2.3 *The burdens of process coordination*

As suggested at the end of Section 1.4, we might wonder whether due to the requirement of continuous attention and the performance of intricate and unceasing cognitive tasks related to prediction, comprehension and language production, that perhaps face-to-face conversation is somehow more demanding than digital conversation. And the fact that we do it at such high speed might make us think that this adds an even further level of difficulty.[190] In most cases of digital conversation, there will be considerably longer intervals between turns. We might think, then, that the absence of the time constraints in digital conversation makes the requirements of it somehow less taxing. So it might seem as if there is a lot more to do (and less time to do it) in face-to-face conversation in order to succeed in communicating. I argue next, however, that there are good reasons to think that the contrary is true. Because although it is essential that interlocutors attend to these concurrent tasks, the nature of face-to-face conversation is such that it is not simply one person carrying out these tasks at any one time. Naturally each conversational partner is responsible for (and is doing) her own thinking and making her own contributions, however the overall task of maintaining the conversation is continuously shared. The continuous attention of the participants in a face-to-face conversation, and the concurrence and simultaneity of the feedback produced, allows the communicative load to be shared. This is because interlocutors in a face-to-face conversation are able to use each other in order to help them formulate contributions to a conversation as and when they are producing them.

I should note here that ability of interlocutors 'to use each other' in order to formulate contributions will also be true in digital conversation to some extent. In a digital conversation between Yasuko and Ichirou, for example, Yasuko can, and usually will, use previous utterances made by Ichirou to help develop her own contributions, and vice versa. Just as they do in face-to-face conversation, digital interactants can draw upon each other's linguistic output to help shape their own and they will also have a common ground that they can both draw upon. However they can't recruit many of the other resources available in face-to-face conversation (such as , for example, the instantaneous feedback of an

[190] Though, of course, that we do this so easily and from such a young age suggests we are clearly well equipped to deal with this speed.

interlocutor) and they often can't rely that an interlocutor is ready, available and focused on the conversation just as they are. Yet it remains true of both face-to-face conversation and digital conversation that in order for them to be communicatively successful interlocutors must still coordinate both in process and in content.

The specific purpose of the process tasks doesn't change depending on the modality used – it is true of both face-to-face and digital conversation that a conversational participant is still required to comprehend the contributions of her interlocutors and produce her own comprehensible contributions in response. So just as it is in face-to-face conversation, it is also the case that for a digital conversation to progress interlocutors must coordinate at the content level, and to facilitate this they must also coordinate at the process level. A key difference between the two types of conversation, however, is that whereas in face-to-face conversation the task of coordinating is concurrently shared, in digital conversation the burden of process coordination rests at any one time primarily with one participant - the in-turner. What I want to argue, then, is that the differences in the burdens of process coordination entails two things. First, although there is less pressure to respond rapidly and to commit continuously to the conversation, when a participant in a digital conversation takes on the role of in-turner, she must carry the weight of the conversation herself. And this makes digital conversation more demanding (more on why so in §§2.3.2-2.3.4). Second, that the burden of process coordination is shared more in face-to-face conversation makes it so that it is an intrinsically more cooperative type of conversation. And these differences in both difficulty and cooperation can help us to understand why we often experience these interactions differently. Let's now consider an example to illustrate the intuition behind these arguments before looking (in §§2.3.2-2.3.4) at the specifics of the burdens of process coordination in face-to-face and digital conversations.

2.3.1 *There's a garage around the corner*

Suppose you and I are out driving in my car. The car suddenly splutters to a halt and I say 'I'm out of petrol.' You tell me there is a garage 200 metres around the corner. I take you to mean that the garage in question is open and sells petrol, and so see this as an excellent opportunity for us to cooperate.[191] I suggest to you that we both push the car to the garage and you agree this is a good plan. How would we do this then? I might suggest we take it in

[191] This is, of course, is taken from one of Grice's examples (1989c, p.32) only here it is put to different use.

turns; maybe I push it for the first 100 metres and you push it for the final 100 metres. You might then point out to me that it would be much more energy efficient, quicker, and generally easier for us both if we push it together for the full 200 metres. You, of course, are right. Importantly, though, either way we do it would be a case of us cooperating. In both scenarios our shared goal is to get the car to the garage and either way we do it we work together to get it there. Though the way in which we cooperate when sharing the load concurrently makes it so that even though we are both pushing the car further, 200 metres instead of 100 metres, when we share the weight of the car between us, we make the action of pushing the car easier. What I think this suggests, then, is if we were to push the car together for the full 200 metres we would be cooperating with each other in a much deeper way than we would be if I push it for 100 metres and then you do the same. Again, on the surface, it might be that the goal and outcome in either scenario is the same, but there is something quite different in how we cooperate to achieve it. And it is the intuition that there is something substantially different in the types of cooperation in these two scenarios that is the basis for some of the rest of my argument.[192]

The point I'm getting to is to suggest that in contrasting face-to-face and digital conversation we see the same divergence in cooperation. The concurrence of the tasks of continuous feedback[193] and interactive alignment[194] allow us to cooperate in face-to-face conversation more efficiently. And that this is what makes face-to-face conversation the more likely setting for entering an *unio mystico*. That in a face-to-face conversation the out-turner is copresent with and concurrently attending to what the in-turner is uttering as she is uttering it, and that the in-turner has some access to the out-turner's uptake and feedback just as she is making her utterance helps her to formulate what she is uttering in a way that makes the language production component of her role as in-turner easier.[195] And the mirror of this is that because the in-turner is copresent and reactive to the out-turner's feedback just as she is making an utterance, this allows the in-turner to develop her utterance in a way that can make the comprehension component of the role of out-turner easier. In the case of digital conversation, however, the burden of communication sits almost squarely with the in-turner. The structure is such that an out-turner has very little to do whereas the

[192] As Jenny Saul points out to me, one important disanalogy to note here is that it may well be the case that pushing a car alone might not be possible, let's suppose the car is suitably small, however, and that you and I are capable of pushing it alone should we need to.
[193] See Ch.2 §3.1.4
[194] See Ch.2 §3.1.5
[195] And this too is another area which links to the common ground.

in-turner is solely tasked with the comprehension of the previous contribution and the formulation of the next contribution. To recall the above example, then, the in-turner is required to push the whole car alone for 100 metres. As I shall argue next, that these tasks must be carried out in isolation not only makes them more demanding, but this also plausibly has consequences for the type of cooperation required to sustain digital conversation.

2.3.2 *The burden of interpretation*

The increased burden on the in-turner works in two interconnected ways. First is something noted by Goldberg (2020), that is; the task of interpretation placed on the in-turner in digital conversation can often be much more complex and cognitively burdensome. One such way Goldberg suggests this is more encumbering for the in-turner is due to the lack of stability in the norms of digital conversation. For example, consider the case of retweets.[196] It isn't clear what type of speech act is being performed by reposting the speech of another without comment. It could be perhaps taken that that one person retweeting another person's tweet is endorsing the tweet reposted.[197] It might also be taken to be a type of quotation of the original tweet.[198] Either way it isn't always clear and so the task of interpreting the speech act-type of a retweet can be difficult.

Perhaps the act of retweeting itself is an oddity if we think in terms of a paradigm case of conversation. Retweeting is not, presumably, something one performs mid-conversation. It may be that an act of retweeting is itself much closer to a broadcast in nature.[199] Or perhaps it might be taken to be a type of conversation starter. But this instability of the norms of digital conversation is not isolated to retweets. Hashtags (#), for example, have multiple uses. They can be used as an indexing feature, as a mean of

[196] Although a 'retweet' is specific to social media platform Twitter, it is, in effect, a more general phenomenon of sharing, usually without comment, the post or utterance of another social media user.

[197] Though a problem with this, as Neri Marsili notes, is that unlike in the case of an endorsement of the following form; 'I agree with x that p.'. The supposed endorsement of a retweet is *defeasible*, it is common to see, for example, the phrase 'retweet not an endorsement' on a user's bio page, or for them to follow up a retweet by stating it is not an endorsement.

[198] The view Marsili argues for is of a retweet being a type of indicative quotation.

[199] Indeed, any public post to a social networking platform that isn't directed at another user specifically might be thought to be of this class. Whatever the specifics of the type of act they are, they are, generally at least, speech acts aimed in the general direction of some audience, though perhaps not specifically at any particular member of that audience. Where they differ from a broadcast, however, is that they may be used as a means of instigating a conversation, or at least an interactive exchange (or exchanges) of sorts.

expressing emotion or mood, or of adding further context to a post.[200] And hashtags can, and are, used in everyday digital conversational exchanges. Now it might be expected that these norms will stabilise over time and as they become more stable perhaps hashtags will become less burdensome to comprehend. Even if it were the case, however, there are still complications and difficulties of comprehension in digital comprehension that we might think are less likely to change.[201] Specifically these are difficulties relating to the asynchrony of digital conversation and the solitary nature of the comprehension involved.

What I contend here, then, is that even if there were a hypothetical situation in which the norms of written speech acts were as equally stable as those we find in face-to-face conversation, it would still remain the case that the task of comprehension in digital conversation will often be more difficult than in its face-to-face correlate. The absence in digital conversation of the time pressures and simultaneity we find in face-to-face conversation creates the conditions in which we might expect it is much more likely that comprehension is a more cumbersome process. Think again about the timescales of face-to-face conversation, if average response speeds are approximately 200 ms, then it would be unlikely an interlocutor might spend, say, 20 seconds considering an utterance made by an interlocutor before making a response.[202] On the other hand, in a digital conversation 20 seconds given to reading an utterance before responding wouldn't seem to be unusual.[203] Now this alone doesn't necessarily make the task more difficult. I could walk a mile at a leisurely pace and it would be an easier task than attempting to sprint the same distance – slower is not necessarily more difficult in and of itself. However, if we keep in mind that the primary purpose of both written and face-to-face conversation is to communicate, then (recalling our pushing the car alone for 100 metres) we might wonder whether delays such as this are indicative of an inefficiency not seen in face-to-face conversation.

For example, it may be the case that participants in digital conversation and face-to-face conversation have equal opportunity to request clarification of some unclear point

[200] This point is another raised by Goldberg (2020). See (Scott, 2015; Wikström, 2014) for more on hashtags and their uses.
[201] There is a further problem of comprehension that relates to the often different intimacy and audience dynamics of the online world; this relates, part, to the rise of trolling and non-human conversational participants such as so called 'propaganda bots' becoming part of our conversational lives, and by the spread of targeted attacks by government agencies, corporate interests and activists.
[202] They may do, of course. The point being it would be unusual.
[203] And indeed, although an empirical matter, we might suppose that delays in response don't have any of the pragmatic connotations of a delay in response such as in face-to-face conversation (see Ch.2).

made by an interlocutor. However, unlike in face-to-face conversation where problems with communication are generally instantly flagged by the out-turner, in a digital conversation it would perhaps seem unlikely that a reader would skim a sentence once and instantly reply 'huh?', for example, without at least giving it a second or third read. If in a conversation between A and B, A assumes B's utterance *u* adheres to something like the cooperative principle and yet A initially struggles to comprehend *u*, then the burden would seem to be on A to at least attempt to comprehend *u* before requesting clarification from B. Again, this need not always be the case, A might simply instantly reply 'huh?' or even '?'. Though without the urgent pressure created by the one second window, A isn't required to flag this trouble instantly, she is afforded more time and opportunity to interpret *u*, and importantly this is done in isolation from B.

Contrast this with face-to-face conversation; there when a communicative problem occurs it is cross-linguistically a norm to respond *immediately* to flag a problem either gesturally or by making an utterance such as 'huh?'[204] So rather than the burden of interpretation of an unclear utterance being felt most acutely by the out-turner (recall interpretation is a task for an out-turner in face-to-face conversation, but an in-turner in digital conversation), the conversational trouble-source is passed back to the utterer of the trouble source for reformulation. So rather than A using time and effort to comprehend B's utterance, A can signal the communication has failed and B is able to attempt to remedy the problem. The reason we might think this is preferable to A trying to work out alone the meaning of B's utterance is that we should expect that it is B who is best placed to resolve the communicative problem; it is B's initial thought that is being offered for consideration after all. Now this might make it appear that by *passing back* the utterance for reformulation the burden falls more heavily on the producer of the unclear utterance, but this too needn't be the case, because related to the difference in the burden of comprehension is the difference in the burden of production. And just as we have good reason to think that comprehending an utterance alone is more difficult than doing it synchronously with another, so too producing an utterance solitarily becomes more difficult.[205]

[204] See for example Enfield (2017, Ch.8)

[205] As with many of the points here, I am speaking in general terms, it need not be the case that the interpretation is always a solitary task in this way. I may write my contribution, send it to you, then after sending it reread it and think it not clear. I might then send you a further message of clarification (a self-initiated repair). The difference is, as always, whereas in face-to-face conversation if something isn't clear, it will become apparent to all of the participants almost instantly – for example, if there is a silent gap of a few seconds after a turn is completed, then for the

2.3.3 *The burden of production*

The increased cognitive burden of digital conversation relates not only to comprehension, we should also expect there is a corresponding increase in the burden of producing an utterance for an interlocutor to comprehend. Although ultimately an empirical conjecture, it strikes me that producing a written utterance will very often be more taxing than producing a spoken or signed utterance. Let's consider next the following observation from Clark;

> When I write my sister a letter, I may take half an hour, pausing halfway through for coffee and revising it several times. She may read it in thirty seconds and reread it. Not only are her actions and mine not synchronized. There may be no point-by-point correspondence between them at all… Writing and reading are no less joint-actions for the lack of synchrony. My actions depend on what I expect my sister to do, and her actions depend on what she thinks I would expect her to do. We still coordinate on content… But I will also design - and redesign, edit, and reedit – my sentences to match the processes I judge she will read them by. I expect her to scan the sentences in order at a certain pace and to do so optimally when I pack information at the right density. Even though our processes are not synchronous, she and I coordinate on them.[206]

So for Clark face-to-face and digital conversations are equally joint actions. I see no reason to disagree with this, the two different scenarios about how we get the car to the garage from Section 2.3.1 are also equally joint actions - that we coordinate them differently doesn't seem to be salient to their status as joint actions. Though the difference in demands placed on us in the car example make them different in some other way – in one case, the burden is concurrently shared and the task is easier as a result, in the other it is passed from one person to another, and this, we might expect, makes the task more difficult. And this is the case too with language production. In a digital conversation, in order to communicate successfully an in-turner will often be required to write, reread, rewrite, reread and edit an utterance to make sure it scans correctly and communicates what it is she wishes to communicate before sending it to be read by an interlocutor. In itself this can

---

conversation to continue it is in the interests of all parties to quickly resolve the problem. This could be through a quick negotiation of what the problem source is, for example. In a digital conversation wherein I just made a contribution to the conversation, there are no obvious ways of determining the success of the communication until an interlocutor responds. And so I am not required to help them with comprehension in the way I would be in face-to-face conversation.

[206] (1996, p.90)

be a time-consuming process (at least relative to making an utterance face-to-face). The nature of such a task, too, is that it is essentially a task performed alone. As mentioned previously, although some of the content of such an utterance may be aided by the previous utterances in the conversation, composing a response is performed in isolation with no further input from an interlocutor until after it has been written and sent.

Contrast this with face-to-face conversation. Any analogue we find there to the 'rewriting' we find in digital conversation is often performed, in part at least, collaboratively and synchronously with others. Face-to-face interlocutors will most often respond within the time constraint of the one second window, and to do so an in-turner will be required to utter something like the first thing that comes to mind. As such, any 'rewriting' will be performed in front of and with a conversational partner. An in-turner need not ask herself whether what she is uttering *might* be understood by the out-turner, the out-turner will very often do this for her by providing simultaneous feedback to help guide the 'rewrite' process. Once her turn begins she shares a thought with an interlocutor who can then help to refine the utterance in real time, if it is so required. The out-turner can help in this task by giving what Goldberg (2020) calls *audience updates*. These updates could be in the form of feedback through any of the communicative channels available, for example an out-turner might look puzzled or nod in understanding. The out-turner can help in other ways too, perhaps by initiating repair for any trouble spots, or by helping to finish an utterance or fill in a blank for the in-turner.[207]

As such, as long as participants remain engaged in the tasks required of them in a face-to-face conversation, then they will be concurrently helping each other to sustain the conversation. So although, as Clark argues, interlocutors in a digital conversation must also coordinate at the process level, the burden of this process coordination will often be placed on the in-turner. The in-turner must work alone to produce her utterances and produce them in a way she estimates will succeed in communicating. And so as Goldberg (2020) observes; if more effort is needed to be put in to digital conversation, then we might expect that we need to derive more joy from it in order to compensate for this. This seems fair enough, and there is more that can also be said along these lines if we recall from Chapter 2

[207] For example, think of a case where an interlocutor forgets a name or noun, and says something like 'Do you remember when we ate that huge sandwich in…erm…', and the other interjects with the place name. Or recall the case in Chapter 2, §3.1.3 where Herb starts to ask about the location of an umbrella, he says; 'where is the other…' before Eve interrupts by saying 'on the back shelf'. See (Goodwin & Goodwin, 1986) for examples of how interlocutors use gestures to help each other when searching for a word.

(§3.1.5) the discussion of interactive alignment. There we saw evidence that when bodily synchronisation increases, subjects more often report that an interaction is experienced as being more harmonious and friendly. So if this is the case, then not only is it the case that coordinating at the process level is more difficult in digital conversation, some of the means of coordinating in face-to-face conversation are in themselves more pleasurable.

2.3.4 *The harmoniousness of alignment*

In Chapter 2 (§3.1.5) it is was noted that interacting groups align in a number of different ways, both behaviourally and cognitively. For example, as conversation progresses eye-movements, hand positions, facial expressions and word choices of interlocutors will often become aligned. This has led some to hypothesise that this alignment at so many different levels makes the task of linguistic alignment easier. The argument for this being that we can cope with the speed of turn-taking in face-to-face conversation by aligning through various channels and that this alignment, in turn, allows conversational partners to distribute the processing load between them.

Clearly many of the potential pathways of alignment used in face-to-face conversation will be absent in digital conversation. Without copresence, visibility and audibility we lose many of the prosodic and paralinguistic features discussed in Chapter 3 (§2.3), as well as the opportunity for bodily alignments. Now of course, this isn't to say that digital conversation is without any means of alignment, naturally interlocutors in such conversations still have at least the opportunity to align linguistically. Participants in a digital conversation also have a common ground (although restricted by their lack of copresence), and they can also use previous utterances of the exchange as a basis for linguistic alignment. So they are certainly able to align and share some of the load. What is clearly different, however, is that the alignment need not, and often won't be, synchronous, and there is evidence to suggest that synchrony itself can be a source of harmoniousness.

Recall, then, the studies by Kerry Marsh and colleagues (2009) discussed in Chapter 2 (§3.1.5). Marsh *et al*. note that when experimental partners were able to bodily synchronise with each other more easily, they report feeling more connected and that their interactions were smoother and friendlier. If this is the case, then we might wonder whether the processes of interactive alignment which allow some of the process load to be shared to enable face-to-face conversation to function at the speed it does might also be contributory to the increase of prosocial attitudes we find more often in face-to-face

conversation. For if in simply carrying out these bodily alignments feels friendly and harmonious then that this opportunity is not available in digital conversation can help us understand why it can appear that digital conversations are often less fulfilling interactions.

2.3.5 *Summary of the difference in process coordination*

In face-to-face and digital conversation, there are different requirements placed on interlocutors in order for them to coordinate at the process level. In face-to-face conversation these tasks are performed synchronously – as an in-turner is making an utterance, an out-turner must attend to the in-turner's utterance, comprehend it, provide feedback on its communicative success and predict its conclusion. She must also use this as a basis for formulating her own upcoming contribution to the conversation. The speeds at which this occurs, and the availability of various communicative channels creates the conditions that allows for participants to concurrently share the process burdens of comprehension and production. In sharing these loads interlocutors are afforded greater opportunity for perspective sharing, indeed we have good reason to expect that this makes the fulfilment of their necessary tasks easier. Through various levels of interactive alignment they are able to synchronise with each other, and this synchronisation itself can generate feelings of friendliness and harmoniousness.

On the other hand, in digital conversation the required process tasks are generally performed asynchronously and in isolation from a conversational partner. Indeed the asynchronous nature of digital conversation is such that at any one time, most of the communicative load is the responsibility of the in-turner. This has consequences for both comprehension and language production. Taking each of these consequences in turn, it will be the case at any one time in a digital conversation the in-turner is tasked with solitarily comprehending a previous utterance. There are a number of problems we might think she might face in doing this. For, example, due to the relative youth of digital conversation the norms of the speech acts used are still unstable, and so comprehending which act is being performed can be more difficult. When we add to this that the communicative bandwidth is reduced, that the aspects of the common ground relating to copresence are absent and perhaps most importantly, that an utterance being responded to will generally have been composed with no audience updates during its formulation, then we should expect that the tasks relating to comprehension will be more difficult. Whereas in face-to-face conversation an utterance is comprehended as it is being made, in digital conversation we might read and reread an utterance a number of times.

Closely related to the difficulty of solitary comprehension is the difficulty of solitary production. When composing a written utterance we might write, rewrite and edit a contribution before sending it to an interlocutor. Contrast this with face-to-face conversation where the response speeds of conversation require that contributions are made rapidly and so often will be much less *finished* than written utterances. Although it might be tempting to think that more time to deliberate on a response in digital conversation is beneficial for reasons of clarity, it certainly need not be the case for reasons discussed above in discussion of the difficulty of comprehension. And when we add to this that an in-turner won't have access to audience updates during the composition of an utterance, the responsibility for this clarity is placed on the in-turner. So whereas in face-to-face conversation the updates of an audience can guide an in-turner as to where there may be communicative problems, in digital conversation the burden is placed on the in-turner to *foresee* any such problems that might arise. And so any comparative process we find in face-to-face conversation to the rewriting of digital conversation is done collaboratively. If an utterance is unclear, an out-turner will quickly make an in-turner aware of it, and then they can work together to solve the communicative problem.

So both face-to-face and digital conversation carry with them a set of necessary tasks required for coordination at the process level and in both types of conversation these tasks serve the same purpose. However, the way interlocutors coordinate in digital conversation is dispersed. The asynchrony of such exchanges makes it so that at any one time the burden of coordination rests with the in-turner. So digital conversation will often feel more difficult when carrying the load. But not only is digital conversation often more taxing, we have good reason to think that the synchrony we find in face-to-face conversation is also conducive to feelings of harmoniousness and friendliness. That face-to-face conversation is fundamentally a synchronous activity, and that interlocutors will interactively align at various levels suggests that face-to-face conversation is more likely to produce feelings of harmoniousness. This all leads us to consider once more the type of cooperation we might find in these two types of conversation.

2.4 *Cooperation*

In the earlier example (§2.3.1) where you and I were pushing a car to the garage around the corner two possible ways of cooperating were considered. First, I would push alone for 100 metres and then you would push alone for the final 100 metres. In another scenario, we would both push the car for the full 200 metres, but we would do so together

synchronously. Intuitively it would appear that even though in both cases we are cooperatively working towards the same shared goal it still appears to be the case that there is something more cooperative about pushing it together. And it is along these lines that I suggest the cooperation in the two types of conversation differ. That we concurrently share the same goal and work towards achieving it by continuously attending to the conversation we sustain our cooperation for as long as we share the goal. If this is the case, then, there is something more cooperative about face-to-face conversation when compared to digital conversation.[208]

In Chapter 2 it was argued that the intricate and continuous nature of coordination at the process level in face-to-face conversation is indicative of a deeply cooperative activity, at least at the process level. For participants to be able to coordinate in the way that they do in face-to-face conversation requires an almost unceasing commitment to the principal goal of sustaining the conversation[209] and this goal must be shared by all participants. The coordination at the process level is thus the enactment of this commitment, and taking cooperation to be the sharing of a goal which cooperative parties work towards achieving, this cooperation at the process level allows conversational participants to meet the preconditions required of them in order to meet something like Grice's cooperative principle. The significance of this, it was argued, is that using the distinction between process and content levels, Grice's cooperative principle is best considered as a means of coordinating at the content level, whereas the precondition to the cooperative principle represents how interlocutors coordinate at the process level in order to meet these content-coordinative aims.

Recall that to meet Grice's cooperative principle (CP);

(CP)

> Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.[210]

[208] And the caveat again, this is not the claim that all face-to-face conversation is more cooperative than all digital conversation, but rather the point is the structure of face-to-face conversation suggests it is necessarily more cooperative at the process level, and this, we might expect, will often affect the content level.
[209] For at least as long as all parties wish for it to proceed.
[210] (1989c, p. 26). Naturally if we are expecting the CP to hold, we might also expect something like the Grice's maxims to be important too. To recap these suggest that a contribution to be just as

face-to-face interactants must satisfy something like the following precondition;

(PCP)

> In the role of out-turner, concurrently with the in-turner producing an utterance; a conversational participant must predict alterations to the purpose or direction of the language exchange, the stage the exchange will be at following the present turn, and begin the set of sub-processes in preparation to produce an utterance.

As was argued in Chapter 2 (§4.6), PCP is formulated with Levinson's preconditions in mind to show how in the role of out-turner there are prediction and production tasks required and that these are dependent on the dynamic nature of the informational elements of CP. For it is those *contents* that the *process* tasks will track. And so just as we might think that two people cooperating in one task might also make them more likely to cooperate in another, we might expect that cooperation at the process level in itself encourages cooperation at the content level. And indeed, in a face-to-face conversation, just as an out-turner is attempting to meet PCP, the expectation is that the in-turner is adhering to something like CP.

Now of course if we are expecting conversational participants in a face-to-face conversation must meet something like the cooperative principle, then we should expect the same in digital conversation. The contrast, however, is that in digital conversation adhering to the cooperative principle doesn't require that they cooperate concurrently, nor does it require any of the perspective-sharing we might expect of face-to-face conversation. So although it is true that in order to sustain a digital conversation, interlocutors would be required to share the same goal to maintain the conversation, and must work towards achieving it, the way they achieve this goal is in isolation from each other. A digital conversation doesn't require a continuously shared goal in the same way as its face-to-face counterpart, the requirement is only that this goal be shared intermittently. And whereas in meeting PCP in face-to-face conversation it is required that an out-turner anticipates the dynamic facts it tracks as they are being altered by the in-turner, in digital conversation this isn't quite the same. So we could propose something like a written precondition, for example;

---

informative as needed (quantity), not knowingly false or undersupported (quality), relevant and not ambiguous, obscure or disorderly (manner).

(WPCP)

> In the role of *in*-turner, upon receipt of an utterance; a conversational participant must determine the purpose or direction of the language exchange, the stage the exchange is at, and begin the set of sub-processes in preparation to produce an utterance.

One difference worth noting here is that WPCP will occur not as an in-turner is making an utterance (as in the case of face-to-face conversation and PCP) but rather once it has been completed. It doesn't rely on prediction, or tracking dynamic facts, but rather it is a process of determining what has already been uttered. Recall in Chapter 21 (§§4.6-4.7) it was argued that this tracking of dynamic facts is indicative of the type of perspective sharing and alignment that occurs in face-to-face conversation, and so in the case of digital conversation we have reason to expect this perspective sharing is different. And so ultimately the cooperation takes a different form.

3 Some benefits of digital conversation

I hope what I have said thus far isn't taken to be a complete denigration of digital conversation. It can, and does, provide a wonderful social outlet for many, be it those who live in remote areas, people unable to leave their house, or for people who find face-to-face conversation difficult, amongst many others. Nothing I have said thus far should lead us to think that digital conversation can't be equally rewarding and satisfying as face-to-face conversation, the point, rather, is to suggest some reasons as to why it often isn't viewed as such when compared to face-to-face conversation. And there is no reason to think that because face-to-face conversation is intrinsically more cooperative at the process level, that this entails that all face-to-face conversations will be somehow more cooperative than digital conversation at the content level. This certainly need not follow. Again, what I suggest is rather that the structure of such interactions suggests that they more often will be. And finally, just because I suggest that the high-level of concurrent coordination, and the sharing of the cognitive load better creates conditions for perspective sharing, this shouldn't be taken to mean this will be true in all cases. We might think that in some cases this probability is reversed, it will be that in some cases the chances of a satisfying and cooperative conversation are increased by being conducted digitally. And in some cases, the opportunities for perspective sharing might be increased by using the written mode. I will

use one particular type of case as an example, that of autistic spectrum disorders (ASD), but I expect many similar examples could be considered to make this same point.

Until recently much research into the perceived difficulties with sociability some people with ASD experience was conducted by looking specifically at the person with a diagnosis, rather than viewing these difficulties as being part of a wider scheme of understanding the nature of conversation as being necessarily social or interpersonal. There is a recent strand of research, however, that looks at how these difficulties might arise not just from people with ASD themselves, but also how those around them react to some of their non-verbal behaviour. For example Sasson *et al* (2017) note (emphasis mine);[211]

> Findings across the three independent studies were remarkably consistent despite using distinct samples and methods. Taken together, they offer strong evidence *that the social presentations of individuals with ASD, particularly their non-verbal cues, including prosody, facial expressions, and body posture, are perceived less favorably and are associated with reluctance on the part of observers to pursue social engagement*. This is particularly important given that individuals with ASD self-reported much greater feelings of loneliness than controls… *Negative first impressions may serve as a barrier to fulfilling this desire for social interaction*, as approach and withdrawal behavior towards novel social partners is based on subjective perceptions regardless of their accuracy. In turn, this may limit opportunities in ASD for developing social connections and friendships.[212]

Interesting here, then, is that some of the features highlighted above are important channels of feedback we see in paradigm cases of face-to-face conversation and are often used to enable interactive alignment. That neurotypical people and people with ASD might struggle to align through such channels might in turn be suggestive of the difficulty, often reported, people with ASD can have with perspective taking. If, as has been argued here, such channels are what help paradigm cases of conversation to function successfully at the speeds that they do, then the negative first impressions neurotypical people have on encountering people with ASD is helpful for understanding why people with ASD can

[211] Another study (Heasman & Gillespie, 2018), looks at how perceptions of Asperger's syndrome (a condition on the autistic spectrum) can be detrimental to social understanding between people with Asperger's and family members.
[212] (2017, p.7)

sometimes struggle with conversations and with perspective taking. And what makes this particularly interesting for current considerations, is an observation in the same paper where it is noted that;

> Our findings show that negative first impressions of adults with ASD occurred only when audio and/or visual information was present, and not when the transcript of their speech content was evaluated.[213]

Which suggests that the rich levels of process coordination found in face-to-face conversation are actually a hindrance to conversation for many with ASD; both from their point of view and from the point of view of their potential interactants. That neurotypical people are less likely to interact with a person with ASD based on audio/visual information makes it so that in the case of ASD we might think that digital conversation, where these features are absent, is actually much more likely to produce fulfilling conversations.

So it certainly need not be the case that in all cases face-to-face conversation provides better conditions to allow the perspective taking and the *unio mystico*, in some cases the requirements of face-to-face conversation will be a hindrance to such a potential interaction. If in digital conversation neurotypical people are more likely to interact with people with ASD, then it would seem that digital conversation, in such examples, is potentially more cooperative than face-to-face conversation would be.

4 Summary

In this chapter I have argued that in one particularly important way face-to-face and digital conversation are different. The requirements of face-to-face conversation are such that (even in cases of disputation) conversational participants are required to coordinate continuously and intricately at the process level and this is not the case in digital conversation. And as a result of this even though both types of conversation require cooperation, the type of cooperation required in face-to-face and digital conversation are notably different. The way that face-to-face interactants cooperate is in sharing the communicative burden continuously, whereas in digital conversation this communicative load is, at any one time, carried by one particular participant. If the process level and content level are indelibly linked (as originally argued in Chapter 2 §4.7), then we might expect that more cooperation at the process level is likely to entail more cooperation at the

[213] (2017, p.7)

content level. As such, we might expect that the structure of face-to-face conversation is such that it will generally produce more harmonious interactions than digital conversation. Finally a scenario was considered where this might not be the case, and in fact, where we might expect the probability of a better conversation is increased by using the written mode. There is still more to say however, because there is one further notable difference between face-to-face and digital conversation that has so far only been gestured towards. This difference also emerges from the different media-specific features, the comparatively reduced personal nature of the type of conversation and is also a product of where digital conversations often take place. The final chapter, then, will act as a case study to show how fruitful it can be to consider the contrasting features of face-to-face and digital conversation by looking at a particular type of digital speech act – *trolling*. There I will consider some of the consequences we see that arise from the fact that many digital conversations are conducted in public, which introduces a different audience dynamic.

5

Digital Conversation: A Case Study on The Speech Act of Trolling

In the previous four chapters I have argued that there are important fundamental differences between face-to-face and digital conversations. These differences relate to how interlocutors in these two types of conversation coordinate their linguistic behaviour in order to sustain their interactions. The upshot of this difference, I argue towards the end of the previous chapter, is that there appears to be a different type of cooperation involved in face-to-face and digital conversations. What I want to do in this final chapter, then, is look more closely at what I take to be one example of the type of things we see emerge from this difference – trolling. The topic of this chapter will be an examination of trolling as a type of speech act that has emerged along with spread of digital conversation. For the reduced cooperation, multiplicity of potential audiences and the reduction of interactive alignment we find in digital conversation relative to face-to-face conversation opens up the possibility for new communicative acts and behaviour. So in this chapter I present what I take to be an example of the type of research that is possible when we focus on the conditions of digital conversation. [214]

The acceptance of trolling as a feature of internet communication, and the clustering together of behaviours that range from playfulness to abusive bullying under the term 'trolling' has interesting moral implications. For it seems that sometimes such behaviour is dismissed as 'mere trolling' and consequently the trolls who perform these acts are granted

[214] I am particularly indebted to many others for this chapter. For formative discussions I am grateful to Jenny Saul, Komarine Romdenh-Romluc and Will Hornett. And for feedback on various versions of some of the material included here I am grateful to Jenny, Komarine, Neri Marsili, Eliot Michaelson and two anonymous reviewers at the *Journal of Social Philosophy*.

a partial exemption from responsibility for the content or consequences of their speech. A result of this is that some acts which might ordinarily be considered bullying, abuse, or threats when uttered outside of the internet can be dismissed as *mere* trolling when carried out on the internet. One popular response to this is to categorise acts of trolling based on perceived intentions or to argue that, for more virulent acts particularly, 'trolling' is not an appropriate label to be applied. Although I agree that, for example, there should be no special exemption granted for racist speech on the basis that it is used in an act of trolling, I argue here that we also shouldn't ignore the structural similarities of the range of speech acts often described as 'trolling'. Application of tools from the philosophy of language and consideration of an act of trolling as a complex speech act can make clear some of these similarities. Once these similarities are made clearer we can then start to understand a little more about *what* trolling is and also *why* it has become such a pervasive feature of internet communication. I suggest that central to acts of trolling is a notion of seriousness and its interplay between the troll, their target and any onlookers to the act of trolling. By viewing trolling in this way we can understand better the rhetorical nature of an act of trolling and why we can often end up facing what I call *the trolling dilemma* – that is, in many acts of trolling the directions of reply are either to respond seriously to an utterance and so be trolled, or to acknowledge it as an unserious act of trolling and so dismiss it as *mere trolling*.

Though far from a complete account of the speech act of trolling, what I present here are the grounds for understanding trolling as a systematically divisive act. In an act of trolling the audience targeted by the troll is a mere prop to be manipulated for the entertainment of the troll and their onlookers. This in-built division between different audiences makes trolling a particularly powerful political device, and the trolling dilemma makes responding to trolling difficult. So when we consider that the act itself is essentially divisive and difficult to respond to effectively, we can start to understand why regimes and movements which thrive on unprincipled division have successfully embraced trolling as a propaganda tool.

1. Trolling preliminaries

Some terminological notes to begin. *Trolling* is carried out by *trolls* on *targets*. A particular target need not necessarily be pre-determined by the troll but is simply anyone who responds in a way amenable to the act of trolling. The target is an audience that trolling is performed *to*, but an act of trolling has an audience (which includes the troll) it is performed *for*. I call this audience *onlookers*.

1.1 *Trolling*

Trolling is a complex linguistic behaviour and the usage of 'trolling' is still in flux. As such, defining it both pithily and satisfactorily is difficult. This is an occasion, though, when etymology is perhaps helpful. 'Trolling' originally referred to a method of fishing whereby bait is placed on a hook, the hook is dangled from a boat, and the angler gently moves the boat downstream waiting for a fish to bite.[215] With a few slight changes we can see how internet trolling is similar. The bait is generally a provocative comment, the hook the social media it is written on, and the troll leaves it dangling waiting for a target to bite.

Trolling isn't an entirely new type of behaviour either. Expressions such as 'baiting', 'codding', 'winding-up', 'teasing' and 'goading' probably capture similar acts (the first two or three are even plausibly fishing-derived terms too). Just like 'trolling', those terms can also sometimes refer to what appears to be amusing, innocuous or playful behaviour, but also cases of abuse and harassment. As I argue later, trolling bears relation to certain types of humour, and it is in these terms we might best characterise it. What we might think makes trolling worthy of close attention, however, is the scale of its proliferation and successes as a communicative tool. It is a conspicuous feature of internet discourse, an implement of corporate marketing[216] and, most importantly, it has become a tool used for disseminating political ideas.[217]

1.2 *Examples*

The following three examples have all been described as trolling. They have been chosen for their differences, and to reflect some of the range of acts that have been termed 'trolling'. I present them here only with some relevant contextual information and analysis follows in subsequent sections.

1.2.1 *KenM*

[215] Such use of 'trolling' dates back to at least Robert Nobbes (1682)

[216] For example Gallucci (2018).

[217] This stretches from the use of large scale so-called 'troll factories' by regimes (see, for example (Gallagher, 2015; Giles, 2019)), to the use of a form of trolling known as 'shitposting' by political parties in election campaigns (for example Read (2019); Stokel-Walker (2019), and far-right terrorists (Evans, 2019).

Kenneth McCarthy posts intentionally confused messages on the internet under the username KenM and is generally regarded as a troll.[218] McCarthy has spoken publicly about his trolling and describes it as "playing a well-meaning moron on the internet." He claims that he does it primarily as a reaction to the harshness of internet discourse.[219] For example, (T1) was posted under a 2015 article titled 'Clinton, Bush fundraising steady amid GOP summer Trump slump'.

(T1):

> Ben Franklin said politicians are like pampers, they both stink and they act like babies.[220]

The comment received the following response from user Karl: "ken hate to tell you pampers have only been around 50 years. true moron."

1.2.2 *Donald Trump*

On 14th July 2019 President Trump posted the following comment on social media platform Twitter.

(T2):

> So interesting to see "Progressive" Democrat Congresswomen, who originally came from countries whose governments are a complete and total catastrophe… Why don't they go back and help fix the totally broken and crime infested places from which they came.

This references four US congresswomen; Alexandria Ocasio-Cortez, Rashida Tlaib, Ayanna Pressley and Ilhan Omar. There are numerous ways in which one might view (T2) as racist, for current purposes though most striking is that it provides a rendition of the "send them back to their own countries" rhetoric eternally popular with racists. Although some of the media responses to Trump's comment correctly condemned it as racist, interestingly some of these same critics noted he was also trolling. In one report the Guardian said; "The presidential trolling may have been meant as a distraction from immigration raids that

---

[218] (Edwards, 2016; Grossman, 2016)
[219] (Edwards, 2016)
[220] The original article on AP News has since been removed, see https://horseysurprise.tumblr.com/image/130477486506 for image, and https://www.reddit.com/r/KenM/ for more examples.

were due in major cities on Sunday."[221] An LA Times editorial on the day of the tweets also said; "He is just trolling, as usual. He is just trying to get a rise out of us. He is baiting us."[222]

1.2.3 *The RIP Troll*

The following example is a case of so-called 'RIP trolling',[223] and was posted under a YouTube video about the 2011 Christchurch earthquake.

(T3)

> I and the rest of the world are pleased your piece of shit family… are dead squashed filthy shit rotting in the ground. Especially those two filthy babies that were squashed REST IN PISS YOU FUCKING RODENT PIECES OF SHIT. [224]

RIP trolls target articles, videos or memorial pages relating to tragedies and deaths. Generally it is supposed that RIP trolls post such comments for the amusement of other RIP trolls. The motivations for RIP trolling and the 'humour' it evokes will be touched upon in §3.3.

2. Trolling, sincerity and seriousness

2.1 *Academic work*

There is some call in academic discussion for the need to distinguish between different types of trolling, for example classifying some acts as 'kudos trolling' and others as 'flame trolling'.[225] So whereas the 'kudos troll' provokes for the entertainment of others, the 'flame troll' is said to be simply carrying out abuse that isn't intended to be humorous. Without knowing the context from which the examples above come from, we might naively think that (T1) is a case of kudos trolling, and (T3) a case of flame trolling, for example. It is plausible that (T3) could, on the face of it, appear to be a case of simple abuse, but as previously mentioned (§1.2.3), and as will be discussed further in §3.3, even (T3) has a component of entertaining an audience. There are clearly merits to drawing distinctions

[221] (Pengelly, 2019)
[222] (LA Times Editorial, 2019)
[223] See McCosker (2014) and Phillips (2011 & 2015, Ch.7) for detailed discussions of RIP trolling.
[224] Cited in McCosker (2014).
[225] (Bishop, 2014; March & Marrington, 2019), for example, both use such terminology, Whitney Phillips (2015) also differentiates types of trolling along similar lines.

between particular acts of trolling in the sense that some acts of trolling are certainly more offensive or damaging than others. However, there is a fundamental problem with trying to draw such a distinction; we might think that any troll is at least, in some way, entertaining themself, and beyond a warped version of utilitarianism I don't see any good moral distinction between abusing someone by saying P to entertain only oneself, on the one hand, and saying P to entertain oneself and some others, on the other. As such I think it very plausible to say that with every act of trolling there is at least one person the troll intends to entertain – themself. When analysing trolling the speech act, therefore, I won't distinguish between acts of trolling in such a way.

There are other arguments we find that suggest that labelling abusive behaviour as *mere trolling* is dangerous as doing so suggests that online abusers are mere trolls, rather than, in many cases, anti-social criminals, and so a term such as 'e-bile' or 'cyberabuse' might be a more apt name.[226] I agree with the aims of such a project; there should be no moral redemption for an act of abuse simply because it was deemed an act of trolling. However, I will argue that paying attention to similarities between the seemingly disparate acts that get called 'trolling' is also an important part of the task of avoiding dismissing abuse as *mere* trolling. By viewing trolling as a type of speech act with some unifying properties, we can understand how it operates as a feature of conversation and how it has spread. We can also understand that there is nothing 'mere' about an act of trolling. It is an intrinsically divisive and manipulative act, regardless of the specific goals of the troll. This doesn't entail an act of trolling is necessarily morally bad, manipulation *is* sometimes justifiable after all. The interesting consequence, though, is that abuse shouldn't be considered somehow more acceptable as a component of an act of trolling ('mere trolling'), it is, in fact, even more questionable when used in such a way. Once we see that an act of trolling is by nature divisive and manipulative, an act of abusive trolling is therefore better thought of not as *mere* trolling, but as abusive *and* divisive *and* manipulative. Just as if I do something wrong and lie about it I would be held liable for two wrongs should the lie be discovered (the lie and the act I lie about), so too using racist speech, for example, doesn't somehow become more acceptable when used as a means of trolling. The moral implications, therefore, are cumulative not diminishing.

2.2 *Defining 'troll'*

[226] See; Jane (2014, 2015) and March & Marrington, (2019).

The next task, then, is to find commonality between these seemingly disparate trolling behaviours. Claire Hardaker (2010) offers a well-cited definition of what it is to be a troll;

(HT)

> A troll is a [computer mediated communication] user who constructs the identity of sincerely wishing to be part of the group in question, including professing, or conveying pseudo-sincere intentions, but whose real intention(s) is/are to cause disruption and/or to trigger or exacerbate conflict for the purposes of their own amusement. (p. 237)

Although useful as a starting point, this is too restrictive for current purposes. Consider the examples (T2) and (T3) above, Trump nor the RIP Troll appear to be making any attempt to construct an identity of wanting to be a part of a group they wish to disrupt, yet both have been described as trolling. Ordinary language usage of 'trolling' has changed since Hardaker proposed (HT), and it no longer seems correct to suggest that a troll attempts to construct an "identity of sincerely wishing to be part of the group", so I'll remove this from consideration. Hardaker also defines the troll rather than the act of trolling, my focus is on the speech act of trolling and so I adjust accordingly. The main point of departure, however, will be to move from talk of 'sincerity' to talk of 'seriousness', I discuss this next.

2.3 *Sincerity and seriousness*

As Hardaker's definition above seems to suggest, and as we might expect considering the role sincerity often plays when analysing duplicitous behaviour, it may be thought that the role of sincerity is essential for an understanding of trolling. As mentioned though, I don't want to discuss trolling in terms of sincerity. The main reason for this being that when thinking about sincerity it seems right that our focus is on the relationship of speaker to utterance.[227] A relatively theory-neutral summary of sincerity might be stated as follows; to

[227] A standard way of thinking about the role of sincerity in communicative practice, for example, is along the lines of that found in Searle (1969). Loosely put, Searlean sincerity is determined by the relationship between an utterance *u* made by *S* and *S*'s beliefs about *u*. So if *S* believes *u* (in the case of assertion), then *S* is sincere in uttering *u*. Conversely, if *S* utters *u* yet doesn't believe *u* (assertion), then *S* utters *u* insincerely. So for Searle a speech act is sincere when the speaker has the mental state expressed by their utterance, and insincere when they don't. The Searlean view meets with counterexamples (Chan & Kahane, 2011; Peacocke, 2000; Ridge, 2006; Stokke, 2018 Ch. 8), for example, however the direction of these challenges generally aims to question what it is to have some mental state in relation to some utterance. I don't here want to challenge any view of sincerity;

utter *u* sincerely is to stand in some relationship to the content(s) of *u*, and to understand sincerity is to understand the nature and type of this relationship. However, what is most interesting about trolling is not the relationship the speaker has to her utterance (not directly, at least), but rather it is the relationship the speaker has to her intended audiences. More specifically, the main point of interest in how a troll intends her multiple audiences to *take* her utterance.

A trolling utterance is generally placed on widely accessible internet-based social media platforms, it is, frequently, a public activity. As with most internet comments it is, potentially at least, therefore viewable by a wide audience. As is suggested in §1, in an act of trolling the audience can be crudely divided into *target* and *onlooker*. Dividing the audience in such a way means the targets are the audience the troll intends to troll, and the onlooker (including the troll) are the audience the troll performs the act of trolling for. If we think it seems plausible to presume that, broadly speaking, the aim of a troll in an act of trolling is to provoke a certain type of response from the target and a particular type of response from the onlooker, then it will turn out that it isn't of direct importance whether a trolling utterance is made sincerely or not. Though this isn't to say that on inspection of individual acts of trolling, and their case-specific moral ramifications, that this utterer-utterance relationship shouldn't be a consideration. The point is simply that a troll can sincerely utter *u* intending to troll a target, but a troll can just as plausibly insincerely utter *y* intending to troll a target. In an act of trolling the troll can say something she believes to be true, something she believes to be false, or even something she has no particularly strong beliefs about - it doesn't matter as long as she gets the response(s) she is aiming for. This type of relationship to the truth of one's own utterance is perhaps closely related, then, to Harry Frankfurt's description of the bullshitter, of which he says;

> [T]he motive guiding and controlling [his speech] is unconcerned with how the things about which he speaks truly are….He does not care whether the things he says describe reality correctly. He just picks them out, or makes them up, to suit his purpose.[228]

---

however, my point is that for the present task, sincerity isn't quite the notion we should be considering.

[228] Frankfurt (2005, p.55)

Whereas in the case of Frankfurt's bullshitter the purpose is to convince an audience of something, we might think that in the case of trolling the primary purpose is to provoke a particular type of action from a target. So we might say that for the purposes of an act of trolling a troll is indifferent to their utterance beyond its utility within the act of trolling. Let us next consider the first two examples in this light.

(T1):

> Ben Franklin said politicians are like pampers, they both stink and they act like babies.

Knowing as we do that KenM is a contrived character, we might expect that the author of (T1) didn't utter it sincerely. He doesn't genuinely believe Franklin said politicians are like pampers.[229] On many accounts of sincerity, then, McCarthy uttered (T1) insincerely, and yet this is generally considered a case of trolling, it seems, however, that he is still aiming for a *particular type of response*. Consider too the second part of the Trump example (T2).

(T2a):

> Why don't they go back and help fix the totally broken and crime infested places from which they came.

It seems plausible that Trump is making a sincere statement in (T2a). We might have good reasons to think he may genuinely wish the congresswomen would "go back" to the places "from which they came". Whether he sincerely thinks this or not, though, is not salient to understanding it as an act of trolling. As an utterance used in an act of trolling it is the intended response(s) that is of most importance. Whether or not Trump is sincere in uttering (T2a) is secondary to the fact that he did utter it, and the nature of what it is he has said (and the use of racist tropes) makes a *particular type of response* likely. With this in mind, what I suggest then is that the *particular type of response* both KenM and Trump are aiming at is related to the notion of *seriousness*. I use this notion of seriousness to help characterise how a speech act of trolling works and so next I'll offer a few thoughts as to

[229] It is possibly a nod to a quote misattributed to Mark Twain about politicians and diapers needing to be changed frequently.

what I take this to be. A good starting point is to consider seriousness and its relationship to joking.[230]

Say S makes a joke to an audience, conventionally we might expect that in making the joke, S intends to make her audience laugh.[231] It could plausibly be the case that S sincerely believes the joke expresses some truth, and as the clichés suggest, *many a true word is said in jest*, and sometimes something is *funny because it's true*. However, there is certainly no requirement that a joke be of this nature. Satire, parody, irony, sarcasm and absurdity are all well-worn jocular devices, and by their nature none of these require that an utterance is made sincerely.[232] Take irony, for example. Traditionally philosophers have treated irony as saying *P* to convey something like *not-P*. A speaker, therefore, can make a joke by making some utterance sincerely or insincerely.[233] So although sincerity might be salient when discussing specific jokes, for a general account it is of little interest. What I suggest *is* of interest when thinking about jokes in general is the way in which the audience *takes* the utterance rather than the sincerity or insincerity of the joker. And this is the case with trolling. Of importance when considering trolling in general is not whether the troll is sincere, but rather the seriousness with which she intends her audiences to take her utterance.

I'll return to discussion of seriousness in §3.2 & §3.2, next though I speak more specifically about trolling and the relationship of troll to onlooker and troll to target. What I want to say is that in uttering *u* a troll intends that the target take *u* to likely be a serious utterance, and that the onlooker take *u* to likely be an unserious utterance. [234]

---

[230] The relationship of trolling to humour is enduring and inescapable. To the type of subcultural troll discussed by Phillips, a form of laughter, 'lulz', is very often the overall aim of any act of trolling. She says; "A corruption (or as the trolls might argue, perfection) of "Laugh Out Loud," lulz celebrates the anguish of the laughed-at victim." (2015, Ch.2).

[231] Or to amuse them in some way, this is very simplistic and not necessarily the case, of course, it may be to humiliate or belittle.

[232] At least in terms of the literal content of the utterance.

[233] I also see no reason why they might not have no sincere/insincere relationship to their utterance. For example a comedian improvising during a set could utter anything that occurs to them with the aim of provoking amusement. There seems no reason to expect that the comedian need have any particular beliefs about the things they say at the point of utterance.

[234] Though I stop short of a precise definition of the speech act of trolling, what I have in mind is something along these lines;

> Linguistic trolling is a perlocutionary act in which A utters *u* and;

3. Perlocutionary Intentions

3.1 *Perlocutionary acts*

As my interest is in trolling as speech act, and as §2.3 discusses the role of intended audience responses (under the guise of seriousness) it seems apt to next consider some of the foundational work on speech act theory, specifically what J.L. Austin says of perlocutions. Austin says; "[s]aying something…produce[s] certain consequential effects upon the feelings, thoughts, or actions of the audience…and it may be done with the design, intention or purpose of producing them".[235] Perlocutionary acts, then, aim at certain effects, and as Austin says later, "'perlocutionary act[s] always include some consequences'; and are 'what we bring about or achieve by saying something, such as convincing, persuading, deterring" (1962, p.108). So if I were to persuade you of something by uttering *u*, it requires that you are persuaded by *u*. The perlocutionary *effect* is that you were persuaded, and my act in persuading you was the perlocutionary act. François Recanati[236] refers to the intention to produce these types of responses as a *perlocutionary intention*, and so a perlocutionary intention is the intention that a perlocutionary effect is produced by some utterance. Considering (HT) and what has been said so far about seriousness, this seems to correlate with an act of trolling. Trivially speaking, a troll will naturally have perlocutionary intentions when making a trolling utterance, and these can be numerous and diverse even within a single act. We can't hope to account for all such intentions so what I aim to do next is to suggest some primary intentions common to acts of trolling.

Typically perlocutionary effects are considered to be such things as convincing, deterring, surprising and misleading. In the case of a specific act of trolling, a troll's

---

i) A's beliefs about *u* are irrelevant to the act of trolling;

ii) A's set of target-directed perlocutionary intentions include that *u* is taken *seriously* by some target(s) B;

iii) A's set of target-directed perlocutionary intentions include that B acts in response to *u* based on ii), and;

iv) A's set of onlooker-directed perlocutionary intentions include that *u* is taken *unseriously* by some onlooker(s) C.

This, then, is fundamentally an unstable type of speech, or perhaps what we might consider to be multi-directional.

[235] Austin, (1962, p. 101)
[236] Recanati, (1986, p.216, fn5.)

perlocutionary intentions might, for example, be to cause disruption, anger, confusion, or to fluster or unnerve a target in some way. However, this will be case-specific, and indeed even specifying one of these perlocutionary intentions for a single act of trolling might prove difficult, for a troll could very plausibly be aiming for any or none of these responses. As such, we need to take a step away from these more traditional types of perlocutionary effect and think of something more general. That it is more general, though, is not to say it is less complex, indeed what I propose is a multifaceted notion. It seems to be that there are two distinct sets of perlocutionary intentions attached to an act of trolling which relate to the troll's relationship to her multiple intended audiences. Recall §2.3, there trolling is discussed in terms of two distinct audiences, the first is the *target* of the act of trolling (the fish the troll intends to catch) the second (which includes the troll) is the *onlooker* (the group that the act of trolling is performed for the benefit of). Let's look now, then, at how the relationship between the troll and these two audiences differ by elaborating on the corresponding sets of perlocutionary intentions.

3.2 *Perlocutionary intentions: Troll-target*

The first set of perlocutionary intentions relates to troll and target. We might expect this first set includes any (or more, or less) of the above mentioned traditional types of perlocutionary intentions (depending on the specific act), plus the effect that the target take the troll's utterance be taken *seriously* (another way we might put this is to say it be taken to be worthy of serious consideration, or to be taken as a serious statement). Regardless of the case-specific contents of this set, it will also include the perlocutionary intention that the target *act* in a specific way. To put this another way, although some specific troll will likely have a number of specific perlocutionary intentions, each of these intentions is secondary to the primary perlocutionary intention to provoke a target into a certain type of response. This *certain type of response*, I suggest, is that a target take the trolling utterance to be an utterance requiring *serious* attention and act accordingly. That is, the intention is for a target to respond to the trolling utterance as if it were a serious statement worthy of a serious response. This doesn't, however, preclude that a target might be uncertain as to the level of seriousness with which they should take the trolling utterance. A target may strongly suspect that the troll is indeed a troll and be suspicious that the troll's intention is to make some target respond in such a way, yet even suspecting this to be the case doesn't entail they won't respond as if the troll has made a serious statement. Take the Trump example (T2), many of those he targets might assume he's trolling - that he has some

intention to make them respond in a way that entertains him and his onlooker audience – yet they might still respond in a way amenable to his trolling.

3.3 *Perlocutionary intentions: Troll-onlooker*

The second set of perlocutionary intentions relates to troll and onlooker. The intentions in this set might include entertaining, amusing, or provoking laughter, for example. It's not immediately obvious that the troll has any specific action-based perlocutionary intentions as is the case with the troll-target set and an onlooker need not play an active role in a trolling interaction. What we can say, though, is that included in this set of perlocutionary intentions is that the onlooker does not take the trolling utterance to be a serious statement in the way the target does. The intention of the troll is that an onlooker view their utterance as something like a joke, or an unserious utterance. This in itself adds a layer of complication that can also help us distinguish seriousness from sincerity. For even if an onlooker takes the utterance to be unserious, as they might do a joke, this isn't to say that they don't endorse it or agree with it. Let's consider (T2a) again.

(T2a):

> Why don't they go back and help fix the totally broken and crime infested places from which they came.

It's certainly plausible that some members of the onlooker audience will believe Trump uttered (T2a) sincerely, and they themselves might even endorse the view expressed in the tweet. However, even if this is the case, it is not incompatible with them taking the utterance to be unserious. Consider again the correlation between joking and unseriousness. A joke can be taken to be amusing by an audience for a range of reasons. A joke might strike us as humorous because it is a shrewd observation, or surprising, or cleverly constructed, for example, but there is also an element of humour that develops from the confounding of social expectation or convention – jokes often play a taboo-breaking social role. It is in this socially confounding way that the onlooker takes some utterance to be unserious. Recall, then, that in §2.3 it was suggested that seriousness, unlike sincerity, doesn't take as its locus the relationship between speaker and utterance, but rather the focus is on the relationship of utterance and audience. So although Trump may well sincerely utter (T2), and some onlooker may endorse his view, there is also no requirement for an act of trolling that either do endorse it, all that is required is the recognition of the way the utterance functions socially. In this case what may be happening

is a challenge to the idea that it is socially unacceptable to say racist things.[237] We can see this type of social taboo-breaking in even more extreme form by looking at (T3) again.

(T3)

> I and the rest of the world are pleased your piece of shit family… are dead squashed filthy shit rotting in the ground. Especially those two filthy babies that were squashed REST IN PISS YOU FUCKING RODENT PIECES OF SHIT.

There is clearly a social norm by which it is rarely acceptable to *speak ill of the dead*. There also seems to be something even more socially distasteful about insulting the mourning families of the deceased, and yet another more shocking layer is added when making light of the death of babies. We don't know much about the disposition of the author of (T3) or their beliefs regarding the comment, so it could plausibly be that (T3) is a sincere utterance and the author could simply be intending to do nothing more than cause distress, for example. However, the fact that this is not an isolated example, and indeed as RIP Trolling is a well-documented phenomenon, we have good cause to think that the author is not *merely* abusing the families of the victims, though they are indeed doing this. If it is an act of trolling, then they are engaged in abuse and also using this abusive speech as a tool in an act of trolling. Phillips (2015, Ch.7) argues that some of the rationale behind RIP trolling is a reaction to a perceived mawkishness in internet responses to tragedies, and utterances such as (T3) are used to respond to this wider social trend. If this is the case, then utterances such as (T3) can be viewed from certain angles as a form of satire. Which again moves us back towards the dual-intentional nature of an act of trolling.

Suppose, then, that the author of (T3) wishes to satirise responses to tragedy found on the internet by making an utterance of polar-extremity to those customarily left on memorial pages. If this is the case, then in this sense (T3) is to be viewed by onlookers as an unserious statement – it's satire.[238] However, clearly the posting of such a comment in a social space wherein most other comments are sympathetic in nature will have a particular effect, and addressed as it is specifically at the families of the deceased, it seems that part of

[237] That it is the actual president making the utterance makes this an even more profound challenge to the norm.

[238] This doesn't exclude that satire can be serious, however if there is such a point to be conveyed by an utterance such as (T3) it would take a lot of presupposition for such an interpretation. That is, faced with simply what is said in (T3) alone, a lot of background knowledge is required to arrive at the understanding of it as satire.

the 'joke' is that it is such an offensive thing to say that a target will likely have a strong emotional reaction to it. Unless they accept it as satire, then they might seem to have little choice but to respond seriously to the utterance even if they suspect it is an act of trolling. The problem for the target, then, is that how to respond is not clear. This, I suggest, is the trolling dilemma.

4. The trolling dilemma

Central to the spread and success of the phenomenon of trolling and its complex socio-linguistic role is something we might call *the trolling dilemma*.[239] The dilemma faced is in how to respond to an act of trolling. As such, it doesn't really arise when an act of trolling is successful in the sense that the target is unaware that they are being trolled and responds seriously to a trolling utterance. We do see the dilemma, however, in almost all other cases wherein a potential target is suspicious they might have encountered a troll. Consider again Trump and the RIP troll. Many readers encountering (T2) or (T3) might be suspicious that these are trolling utterances. In being suspicious in such a way, the reader might also suspect that the intentions of Trump and the RIP troll are for the target to take these utterances seriously, in some sense, and respond accordingly. [240] Responding to the trolling utterance in such a way, though, allows the troll to successfully troll – the respondent thus becomes the target and does precisely what the troll wants. However, to not respond as if trolling utterances are serious statements. and simply dismiss them as acts of trolling, can therefore allow a president to use racist tropes, or an RIP troll to direct abuse at mourners without challenge. In effect what this means is that a president using racist tropes can be dismissed as engaging in *mere* trolling.

If the division I suggest that exists between target and onlooker reflects the structure of an act of trolling, there might be another unpalatable conclusion to be drawn too. In not responding we might, by default, be an onlooker too. If we assume Trump is *just joking* or the RIP troll is *merely satirising* in order to troll, then we take their utterances to

[239] Thanks to Komarine Romdenh-Romluc for discussion that led me to think about this.
[240] Interestingly enough, the availability of an utterer's intentions in an act of trolling seems to differ from many other types of manipulative speech, such as misleading, where it is essential for the success of the act that the intention to manipulate remains hidden. For example consider the *covert collateral acts* such as discussed by Bach and Harnish. These acts "are performed with intentions that are intended not to be recognized." (1979, p. 101). A similar notion too is used by Saul (2018) who suggests that dogwhistles are a strain of *covert perlocutionary act* which are defined as an act that doesn't succeed if the perlocutionary intention is recognised. Even recognising the troll's *trolling intentions*, however, doesn't provide escape from the dilemma.

be *unserious*. We need not endorse (T2) or (T3), and we might find them repugnant, but we take them to be *unserious* in the sense that we don't take them to be utterances worthy of serious response. This could even be the case in the example (T1), though with much less worrying moral implications.

(T1):

> Ben Franklin said politicians are like pampers, they both stink and they act like babies.

It's plausible that a reader of (T1) might assume it to be a joke; it's verifiably wrong and makes little sense. (T1) did receive what appears to be serious responses though; recall it was pointed out to KenM that pampers didn't exist during Franklin's time. Again we have no knowledge of what the respondent thought of (T1), but even if they were suspicious that (T1) was not intended to be a serious remark, not knowing for certain might still be reason enough to respond in a way amenable to the troll's troll-target intentions. Suppose the respondent was someone with an acute concern about the well-documented spread of false information on the internet. This being the case, then, even an example as seemingly innocuous as (T1) might push someone towards the dilemma. From that particular point of view the dilemma is: should one respond seriously to a suspected act of trolling, and so become a target in a successful act of trolling, or should one let the demonstrably false claim go unchallenged?

That such a dilemma arises out of an act of trolling, I suggest, is part of the reason why trolling has been embraced by political actors in recent years, particularly those wishing to propagate divisive violent ideologies. The ambiguous dual-identity of an act of trolling – as speech that can be responded to either seriously or unseriously - allows those with racist views, for example, to make public their racist ideas, but with enough ambiguity that an effective response is difficult to formulate. That in such a case the trolling dilemma can give at least the appearance that there is no good response - the troll either succeeds in trolling when taken seriously, or racist views remain unchallenged or dismissed as *mere* trolling when taken unseriously – provides the racist statement with a status not befitting its actual content. That is, it can appear like a view or argument that has no adequate response. However, in the case of trolling, it isn't the argument itself that carries this dilemma, but rather it is the ambiguity built-in to the rhetorical device (an act of trolling)

used to deliver it.[241] There is another closely related side to the appeal of trolling to these groups too, and this is the topic I discuss next; the cloak of humour.

5. The cloak of humour

So far the characterisation of the speech act of trolling has been based around the notion of seriousness. This notion, I claim, is closely related to joking. As such it is worth considering the role of humour in trolling in a little more detail. That trolling is often malignant, bullying and abusive, and is increasingly used as a tool of propaganda, might make it seem odd to suggest that a form of humour or unseriousness underpins it. However, when we remind ourselves that humour can be used to dehumanise, poke fun and bully, and can also be used to mark out 'us' and 'them', then it need not be thought to somehow dampen the force of an act of trolling. A racist 'joke' is still a racist remark, just as a sexist 'joke' is still sexist. That it is shrouded in the cloak of humour makes it no less potent. In fact, there is an argument to be made that it becomes even more forceful in this guise. Indeed consider what Hobbes describes in his famous description of what has come to be known as the *superiority view of humour*.

> I may therefore conclude, that the passion of laughter is nothing else but sudden glory arising from some sudden conception of some eminency in ourselves, by comparison with the infirmity of others… It is no wonder therefore that men take heinously to be laughed at or derided, that is, triumphed over.[242]

Although Hobbes's account doesn't seem to fit with many forms of humour, it captures something of the nature of trolling. If the target is merely a prop in an act of trolling, then in Hobbes's terms, the troll's aim is to triumph over the target, and they do this for the entertainment of the onlooker. If this is correct, then what seems to be built-in to the act of trolling is a natural division between onlooker and target. The target becomes a mere figure of fun, and as I will discuss in §5.2, this has some serious consequences. Before

[241] The issues this bring to mind are not modern, we can see correlation in ancient discussions of rhetoric. Think of Plato's observation in the *Protagoras* of Protagoras's use of argument as spectator sport (335a4-8). Also Aristotle's rejection in *Rhetoric* II.24 of the sophists claims to be making the 'weaker argument the stronger' (1402a23–5), in which Aristotle explains how the methods of Protagoras used in supposedly doing this were rejected once it was realised they relied on rhetorical manoeuvring. On this note, it is worth directing you to the excellent Rachel Barney (2016) account of trolling delivered in the style of Aristotle.

[242] Hobbes, (1845, Ch. IX.13)

considering this though, I want to discuss how trolling acts can provide a level of deniability to trolling utterances.

5.1 *Figleaves*

There are good reasons why we might want to avoid dismissing Trump using racist tropes as *mere* trolling, however this type of deniability is invoked in Trump's defence (even implicitly by his critics). That this type of defence might seem available gives us reason to consider that in some respects trolling operates in a similar way to a racial figleaf, such as discussed by Jenny Saul.[243]

A successful racial figleaf operates by offering a block on the inference from an utterer making a racist comment to the conclusion that the utterer is racist ('I'm not racist but…,' being a familiar example). If it is accepted that Trump is trolling when uttering racist tropes, then it leaves open the possibility to read his utterance unseriously. So if it is viewed by an onlooker as merely a means to 'wind-up' those who *do* take it seriously, or to force them to confront the trolling dilemma rather than taking it as a serious utterance, then this changes what it becomes acceptable to say.[244] As Saul puts it; "[i]f the audience accepts that the figleaf blocks the concern about racism arising from the utterance of racist sentence R, then R becomes seen as something one can say without being racist". Similarly, if someone were to accept that by uttering (T2) Trump is *merely* trolling, this too might provide a block to the concern of racism that Trump's utterance of the racist (T2) might ordinarily entail. A conclusion from that being that if (T2) is uttered in an act of trolling, then (T2) is not necessarily racist. So whereas with the type of figleaf characterised by Saul a speaker will offer some qualifying remark, in an act of trolling the act itself operates in this way. Again, this should give us cause for concern and points again to the problem posed by the trolling dilemma. For if we dismiss (T2) as an act of *mere* trolling, we give credence to the notion that (T2) can be uttered (without caveat) in a non-racist way (or perhaps less strongly, in a 'less-racist' way), we are in danger of accepting it as an effective figleaf. The dilemma is such though that if we do respond to Trump's racism, we then become props in Trump's act of trolling.[245]

---

[243] Saul (2017, p.116)

[244] Trump did also offer what Saul terms a *diachronic figleaf* following the original utterance, tweeting 'I don't have a racist bone in my body'.

[245] Indeed, claims of 'joking' can often be used to block other types of inference, for example a threat. Consider the phone call between Rudy Giuliani and a journalist on 14th November 2019 in which

5.2 *Humour and propaganda*

The use of humour in propaganda is not new. Humorous media such as cartoons and caricature have been used effectively in recent history as a means of disseminating political messages. What seems to set trolling apart though is that not only can it be used to disseminate ideas, but the people those ideas reach are then equipped to use trolling as a means to disseminate them further – even if they are 'just joking'. So even if a troll is initially drawn solely to the seeming irreverent and taboo-breaking 'humour' of using, for example, racial slurs in an act of trolling, and even if they don't take themselves to sincerely hold racist beliefs, when they use them in an act of trolling they are still spreading racist ideas. In doing this the troll is still contributing to a divisive climate both in the content of what they say and also with the type of speech act they are using. For it seems that making unpalatable ideas appear in some way less serious can detract from their insidiousness yet can still be used to broadcast them further. Not only can it be used to broadcast them, it can make them more appealing; humorousness and unseriousness are attractive qualities after all.

The last point to make, then, is that trolling has become an effective method of spreading authoritarian and far-right ideas in recent times, and based on the previous discussion, we might have some good reasons to expect trolling to be such a natural vehicle for this type of idea. One of Jason Stanley's overall points in his analysis of the tactics used by fascists to gain power is that they seek to separate populations.[246] Trolling as an act does this *regardless* of the content of the utterance. It is built-in to an act of trolling that there be a target and an onlooker, and that the target is objectified – the target is a prop in an act of trolling and their worth is as a mere source of amusement. This in itself need not be socially divisive, such as in the case of (T1), which seems so absurd as to not be genuinely divisive. However, even innocuous-seeming (T1) includes targets to be laughed at, and even (T1) can make a reader confront the trolling dilemma. So when already divisive political ideas are loaded into the act of trolling, the divisiveness of the act itself, in combination with the divisiveness of the content, makes for a powerfully disruptive act. If I'm right about the humour component of trolling too, there is further cause for division. To the onlooker, a target is lacking humour, they don't get the 'joke'. So there is division between those who

---

Giuliani responded to a question as to whether he was worried if Trump might "throw him under the bus" to avoid impeachment, Giuliani responded "I'm not, but I do have very, very good insurance", his lawyer, also on the phone call, then interjected to say "He's joking". Kirchgaessner, (2019).

[246] Stanley (2018).

'get the joke' and those who don't. The power of that as an enforcer of disharmony shouldn't be underestimated. Humour is not to be taken lightly, and if trolling is a form of humour, then this makes it no less of a serious matter.

6. Conclusion

What I present here is intended as a foreshadowing of the direction of research I think the work in previous chapters points us towards. Once we understand some of the fundamental differences in the communicative environment digital conversations take part within, we can understand more clearly how it is that new linguistic behaviours can arise. Speech acts such as trolling, though possible in face-to-face conversation, spread online due to some of the particularities we find in digital communicative spaces. That interlocutors are not required to coordinate and cooperate in the same way in digital conversations as they are face-to-face interaction we might suppose creates an environment in which misleading behaviours such as trolling can flourish.[247] Recalling some of the features of conversation discussed in Chapter 2 and Chapter 4, whereas face-to-face we tend to be co-present with our interlocutors, in digital conversation we tend to be geographically distant. One result of this is that we often don't know who our audience is precisely. Add to this too some of the medium-specific differences such the contrast between the evanescence and recordlessness of face-to-face utterances and the permanent recorded nature of written digital utterances. A result of this is that digital utterances have potentially much wider and more disparate audience than those made face-to-face. Digital conversation therefore has a generally much more public character and with this comes a potential multiplicity of audiences. What I think this should lead us to consider (though is beyond my scope here) is how this multiplicity of audiences can lead to other types of novel linguistic behaviour. If the troll, as described above, exploits the multiplicity of audiences to perform different speech acts relative to different audiences, then this might give us cause to consider a novel class of speech act beyond simply trolling.

[247] Such environments are also conducive to other types of misleading behaviour. Take for example the arguments made in Pepp, Michaelson & Sterken (2019a, 2019b) as to why although one type of 'fake news' predates the internet, the particularities of the digital environment mean that there is another distinct type of information spread that relies on the available tools of dissemination online that has been contributory in the creation of a new type of 'fake news'.

Conclusion

Communication, of which conversation is just one part of, is clearly a rich and sprawling object to study. The recent technological developments in communication technology have added even more layers to how we can and do communicate with each other. One could never hope to do justice to its intricacies and variations in a single dissertation but what I hope to have done here is convince you that by focusing on one seemingly small aspect of communication – how we coordinate in conversation – that there is a lot we can learn about how we communicate. More particularly I hope that it has provided some insight to one of the main differences between the types of conversations we've been having for thousands of years and the newer types of conversations made possible by recent technological innovation.

I hope too that this hasn't appeared to be casting a negative light on digital conversation. Such an interpretation would be understandable. It could be read that I portray the paradigm type of face-to-face conversation as being a harmonious, meaningful and cooperative endeavour, whereas digital conversation is made out to be a chaotic, burdensome and uncooperative facsimile of face-to-face conversation. Some of this interpretation might be taken from the way that I set up the discussion. In Chapters 1 & 2 I present as default face-to-face conversation and then contrast digital conversation in Chapters 3 & 4 in terms of some ways in which it deviates from this default. Indeed, the very particular aspects I focus on leave open the inference that because digital conversation is lacking some of what I consider to be positive features of face-to-face conversation, then it must in turn be inferior. That however is certainly not what I take away from this. I take these particular differences to be interesting and informative but one of the things I have come to realise more clearly over the course of this project is that the differences we find between the two types of conversation don't suggest any general inferiority on the part of one when compared to the other. Rather what the spread of digital conversation has done, in fact, is increase the communicative possibilities available to us. So whereas we shouldn't neglect appreciation of the very unique powers of face-to-face conversation as a means of

communication, nor should we neglect the unique power digital conversation has in extending the possibilities of how we can communicate.

I briefly discuss in Chapter 4 Section 3 one benefit the particular structure of digital conversations might bring as a means to help people with ASD not be excluded from a wider conversational life. And this type of benefit will certainly extend far beyond such cases. For people living geographically remotely from others, people living with social anxieties, or communicative or sensory disorders, digital conversation can in some cases provide a much more harmonious, meaningful and cooperative type of interaction than face-to-face conversation can. Digital conversation can in other ways also allow for conversation where perhaps no conversation would otherwise have existed.

For any of us with friends and family not living in close enough proximity to meet for conversation communication technology provides the opportunity to exchange texts, emails or view each other's' posts on social media and so provide us with interactions that otherwise wouldn't have been possible. It allows groups of geographically disparate people with a particular similar interest to meet and discuss it. It has provided a platform and meeting place for marginalised groups and provided some opportunity to address conversational injustices. And there are other ways it has added depth to our communicative lives. As a means of day-to-day coordination often digital communication is often much more useful- face-to-face conversation is useless as a means of explaining to a friend that you will be late to meet them at the café. So although in comparison to face-to-face conversation digital conversation can seem to be a more solitary endeavour. When compared to an absence of conversation though, clearly it isn't more solitary, so ultimately it can create new opportunities for conversations to develop. And as part of my reason for choosing conversation as the central focus of my dissertation is a love of conversation, I take this to be generally a good thing.

In order to conclude I'd like to highlight two particular strands of future research that I think emerge from the discussions here. The first is along primarily empirical psycholinguistic lines. As I tried to make clear in Chapter 2, there is a prolific movement in psycholinguistics that over recent years that has provided us with a fascinating picture of how face-to-face conversation manages to be so successful. It is this body of work that I particularly draw upon to make the contrast between face-to-face and digital conversation in Chapter 4. The way that I draw this contrast, however, is by taking the findings from empirical studies on face-to-face conversation and hypothesising as to why we might think

that some of the requirements of face-to-face conversation don't hold for digital conversation. I then use these hypotheses to form the basis of my discussion on the implications this might have for the nature of cooperation in digital conversation. One area of empirical research, then, would look to test the validity of these underlying hypotheses.

The other strand of research that would be productive is along lines that would be of interest to philosophers of language, theoretical linguists and social philosophers. In the final chapter I moved away from the direct comparative project of earlier chapters and focused more closely on the type of phenomena that arise in digital conversation. It is in this type of direction that I think future research in the area would be most usefully directed. The growth of digital conversation and the peculiarities of the communicative environments they take place in have created the opportunities for new types of communicative behaviour to develop. So we see the emergence of new types of speech act such as trolling, sharing and reposting. There are also new forms of what we might call *mechanised speech* to consider, that is speech that can be created at the click of a button such as a like, a retweet or forwarded email which bypasses many of the psycholinguistic processes (such as those discussed in Chapters 2 & 4) ordinarily required for language production. The emergence and spread of such new types of linguistic behaviour in digital conversational spaces make these rich areas worth pursuing in order to develop new aspects to our theories.

A further area of research should look into the effects of the structures of digital conversational spaces on the notion of conversational control (such as discussed in Chapter 2 (§4.1-4.3). Whereas there I assume that digital conversations can very often have similarly synocratic levels of control as face-to-face conversations, as is also noted in the discussions in Chapter 4 (§1.2.3) many conversations (face-to-face and digital) will have more autocratic structures in which the control over the discussion rests predominantly with one interlocutor (echoing work such as found in Langton (1993), Kukla (2014), Green (2017), Bianchi (2019), this is a type of injustice). An interesting dimension to this debate stems from the structures of social media. So one area of interest to look at is how the potential to organise in these communicative spaces can provide a place for marginalised communities to redress some of the conversational injustices they face by each adding a to a collective conversational control redress.

There are also issues that arise in digital conversation which will put to the test some of our *extant* language theories. Although much has been said about the topic in the

philosophy of mind and of AI, currently there seems to be very little in the language literature about how we deal with the utterances of bots and AIs. Anyone familiar with digital conversational spaces will be aware that many of the participants in digital conversations are not human. Be they be chatbots, trollbots, or automated social media accounts. Fundamentally these are machines to which we might be hesitant to ascribe the type of complex intentions which we generally ascribe to each other. However it is also the case that they appear able to make assertions or ask questions. This being the case there are questions to be asked of theories which posit complex speaker intentions or beliefs as being requisite components for performance of such speech acts.

References

Ammonius. (1996). *On Aristotle On interpretation 1-8* (Blank, D. L. Ed.). London : Duckworth, 1996.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Aristotle. (1984). *The Complete Works of Aristotle: The Revised Oxford Translation* (J. Barnes, Ed.). Princeton, N.J: Princeton University Press.

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, *4*(6), 1236–1239.

Austin, J. L. (1962). *How to Do Things With Words*. Oxford: Clarendon Press.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press, 1988.

Bach, K. (1994). Conversational Impliciture. *Mind & Language*, *9*(2), 124–162.

Bach, K. (2006a). *From the strange to the bizarre: Another reply to Cappelen and Lepore*. Retrieved from http://online.sfsu.edu/~kbach

Bach, K. (2006b). *Minimalism for Dummies : Reply to Cappelen and Lepore*. Retrieved from http://online.sfsu.edu/~kbach

Bach, K. (2006c). The excluded middle: Semantic minimalism without minimal propositions. *Philosophy and Phenomenological Research*, *73*(2), 435–442.

Bach, K. (2012). Context Dependence (such as it is). *Continuum Companion to the Philosophy of Language*, 153. Retrieved from http://userwww.sfsu.edu/~kbach/Bach.ContextDependence.pdf

Bach, K., & Harnish, R. M. (1979). Linguistic Communication and Speech Acts. In *Linguistic Communication and Speech Acts*.

Baddeley, A. (1986). Working memory. In *Working memory*. New York, NY, US: Clarendon Press/Oxford University Press.

Baddeley, A. (1990). *Human memory : theory and practice*. Hove : Erlbaum.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208.

Baddeley, A., & Hitch, G. J. (1974). Working memory. In (Bower, G. Ed.), *Psychology of Learning and Motivation* (pp. 47–90). New York: Academic Press.

Barthel, M., & Levinson, S. C. (2020). Next speakers plan word forms in overlap with the

incoming turn : evidence from gaze-contingent switch task performance. *Language, Cognition and Neuroscience*, *3798*.

Bartlett, F. C. (Frederic C. (1932). *Remembering : a study in experimental and social psychology*. Cambridge [etc.]: Cambridge etc. : Cambridge University Press, 1932.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380.

Bianchi, C. (2019). Asymmetrical Conversations. *Grazer Philosophische Studien*, *96*(3), 401–418.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behav Brain Sci*, *7*(2), 173–188.

Bishop, J. (2014). Representations of "trolls" in mass media communication: a review of media-texts and moral panics relating to "internet trolling." *International Journal of Web Based Communities*, *10*(1), 7.

Bloomfield, L. (1935). *Language*. London: George Allen and Unwin.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*.

Borg, E. (2004). *Minimal Semantics*. Oxford: Oxford University Press.

Borg, E. (2012). *Pursuing Meaning*. Oxford: Oxford University Press.

Borg, E. (2019). Explanatory Roles for Minimal Content. *Noûs*, *53*, 513–539.

Borg, E., & Connolly, P. (forthcoming). Exploring Linguistic Liability. In *Oxford Studies in Philosophy of Language Vol.2* (Sosa, D. & Lepore, E. eds). Oxford: Oxford University Press.

Bradford, A., Sagara, K., & Zeshan, U. (2013). Multilingual and multimodal aspects of "cross-signing" – A study of emerging communication in the domain of numerals. Poster presented at *Theoretical Issues in Sign Language Research Conference, University College London, Deafness, Cognition and Language Research Centre*.

Bratman, M. E. (1992). Shared Cooperative Activity. *The Philosophical Review*, *101*(2), 327–341.

Brentari, D., Falk, J., Giannakidou, A., Herrmann, A., Volk, E., & Steinbach, M. (2018). Production and comprehension of prosodic markers in Sign Language imperatives. *Frontiers in Psychology*, *9*(MAY), 1–14.

Brown-Schmidt, S., & Duff, M. C. (2016). Memory and Common Ground Processes in Language Use. *Topics in Cognitive Science*, *8*(4), 722–736.

Byun, K. S., de Vos, C., Bradford, A., Zeshan, U., & Levinson, S. C. (2017). First Encounters: Repair Sequences in Cross-Signing. *Topics in Cognitive Science*, 1–21.

Byun, K. S., de Vos, C., Bradford, A., Zeshan, U., & Levinson, S. C. (2018). First Encounters:

Repair Sequences in Cross-Signing. *Topics in Cognitive Science*, *10*(2), 314–334.

Cappelen, H., & Lepore, E. (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell Publishing Ltd.

Cappelen, H., & Lepore, E. (2006a). *Kent Bach on Minimalism for Dummies*. Retrieved from http://folk.uio.no/hermanc/docs/KBM.pdf

Cappelen, H., & Lepore, E. (2006b). Reply to Bach. Retrieved from http://folk.uio.no/hermanc/docs/BachPPR.pdf

Carston, R. (1991). Implicature, Explicature, and Truth-Theoretic Semantics. In (Davis, S. Ed.), *Pragmatics* (pp. 155–181). New York: Oxford University Press.

Carston, Robyn. (2008). Linguistic communication and the semantics/pragmatics distinction. *Synthese*, *165*(3), 321–345.

Chan, T., & Kahane, G. (2011). The trouble with being sincere. *Canadian Journal of Philosophy*, *41*(2), 215–234.

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In Resnick, L. B., Levine, J. M. & Teasley, S. D. Eds.), *Perspectives on socially shared cognition*. APA Science Volume Series

Clark, H. H., & Marshall, C. (1978). Reference diaries. *Theoretical Issues in Natural Language Processing*, 57–63.

Cohen, J. (2013). Indexicality and the Puzzle of the Answering Machine. *The Journal of Philosophy*, *110*(1), 5–32.

Cohen, J., & Michaelson, E. (2013). Indexicality and the answering machine paradox. *Philosophy Compass*, *8*(6), 580–592.

Connolly, N. (2017). I'm Here Now, But I Won't Be Here When You Get This Message. *Dialectica* , 71 (4), 603-622

de Saussure, F. (1959). Course in General Linguistics. In (Baskin, W. Ed.), *Columbia University Press*.

Dijksterhuis, A. P., & Bargh, J. A. (2004). The Perception-Behavior Expressway: Automatic Effects of Social Perception on Social Behavior. *Advances in experimental social psychology*, Vol. 33 (p. 1–40).

Dostoevsky, F. (1993). *A Writer's Diary: Volume One 1873-1876* (Lantz, K. Ed.). Evanston, Illinois: Northwestern University Press.

Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*, *20*(3), 249–268.

Edwards, P. (2016). The world's greatest internet troll explains his craft. Retrieved August 1, 2019, from: https://www.vox.com/2016/5/6/11597960/ken-m-internet-troll

Enfield, N. J. (2006). Social Consequences of Common Ground. In (Enfield, N. J. & Levinson, S. C Eds.), *Roots of human Sociality: Culture, Cognition and interaction* (pp. 399–430). Oxford: Berg.

Enfield, N. J. (2017). *How We Talk: The Inner Workings of Conversation*. Basic Books.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., … Torreira, F. (2013). *Huh? What? – A first survey in 21 languages*. Retrieved from: http://www.oapen.org/record/630828

Evans, R. (2019). Shitposting, Inspirational Terrorism, and the Christchurch Mosque Massacre. *Bellingcat*. Retrieved October 9, 2019, from: https://www.bellingcat.com/news/rest-of-world/2019/03/15/shitposting-inspirational-terrorism-and-the-christchurch-mosque-massacre/

Faulkner, P. (2018). Fake Barns, Fake News. *Social Epistemology Review and Reply Collective*, *7*(6), 16–21. Retrieved from http://social-epistemology.comshorturlhttps//wp.me/p1Bfg0-3Y4

Fillmore, C. (1981). Pragmatics and the description of discourse. In (Cole, P. Ed.), *Radical Pragmatics* (pp. 143–166). New York: Academic Press.

Fodor, J. D. (1998). Learning to Parse? *Journal of Psycholinguistic Research*, *27*(2), 285–319.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. *Proceedings of the 1st International Conference on Speech Prosody*, 83–88.

Gallagher, P. (2015, March 27). Putin's army of pro-Kremlin bloggers. *Independent*.

Gallucci, N. (2018). 13 of the absolute best and worst brand trolls in recent history. Retrieved July 1, 2019, from MashableUK website: https://mashable.com/article/best-brand-trolls-social-media

Garfinkel, H. (1967). *Studies in Ethnomethology*. Englewood Cliffs, N.J.: Prentice-Hall.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, *6*(JUN).

Gathercole, S. E., & Baddeley, A. (1993). *Working memory and language*. Hove: Hove : Lawrence Erlbaum Associates, 1993.

Gelfert, A. (2018). Fake news: A definition. *Informal Logic*, *38*(1), 84–117.

Giles, K. (2019). Putin's troll factories. *The World Today*, *71*(4), 47–48.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS ONE*, *10*(3), 1–24.

Goffman, E. (1967). Interaction ritual: Essays on face-to-face behavior. In *Interaction*

Ritual: Essays in Face-to-Face Behavior.

Goldberg, S. C. (2020). The Promise And Pitfalls Of Online "Conversations." Retrieved February 12, 2020, from *Royal Institute of Philosophy London Lectures* website: https://www.youtube.com/watch?v=DfxhOmRrDcE&feature=youtu.be&fbclid=IwAR3rotxAmSmjSifbR4wXB9dPf7j1fqnKrM5cwTtTiPKi2jfQyt-mr4Eqxtg

Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, *62*(1–2), 51–76.

Gould, W. R. (2018). Go ahead, talk to yourself. It's normal —and good for you. *NBC*. Retrieved January 21, 2020, from: https://www.nbcnews.com/better/health/talking-yourself-normal-here-s-how-master-it-ncna918091

Gray Hardcastle, V., & Flanagan, O. (1999). Multiplex vs. Multiple Selves: Distinguishing Dissociative Disorders. *The Monist*, *82*(4), 645–657.

Green, M. (1999). Illocutions, implicata, and what a conversation requires. *Pragmatics and Cognition*, *7*(1), 65–91.

Green, M. (2017). Conversation and common ground. *Philosophical Studies*, *174*(6), 1587–1604.

Grice, H. P. (1989a). Further Notes on Logic and Conversation. In *Studies in the Way of Words*. London: Harvard University Press. 41-57.

Grice, H. P. (1989b). Indicative Conditionals. In *Studies in the Way of Words*. London: Harvard University Press. 58-85

Grice, H. P. (1989c). Logic and Conversation. In *Studies in the Way of Words.* London: Harvard University Press. 22-40

Grice, H. P. (1989d). Presupposition and Conversational Implicature. In *Studies in the Way of Words.* London: Harvard University Press. 269–282.

Grice, H. P. (1989e). Retrospective Epilogue. In *Studies in the Way of Words*. London: Harvard University Press. 339-386

Grice, H. P. (1989f). Utterer's Meaning and Intentions. In *Studies in the Way of Words.* London: Harvard University Press. 86-116

Griffin, Z. M., & Bock, K. (2000). What The Eyes Say About Speaking. *Psychological Science*, *11*(4), 274–279.

Grossman, S. (2016). The 30 Most Influential People on the Internet. Retrieved July 10, 2019, from *Time* website: https://time.com/4258291/30-most-influential-people-on-the-internet-2016/

Habgood-Coote, J. (2019). Stop talking about fake news! *Inquiry (United Kingdom)*, *62*(9–10), 1033–1065.

Halliday, M. A. K. (1963). Intonation in English Grammar. *Transactions of the Philological Society*, *62*(1), 143–169.

Halliday, M. A. K. (1989). *Spoken and written language* (2nd ed.). Oxford : Oxford University Press.

Hansen, M. B. M. (2008). On the availability of "literal" meaning: Evidence from courtroom interaction. *Journal of Pragmatics*, *40*(8), 1392–1410.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, *6*(2), 215–242.

Harris, C. B., Keil, P. G., Sutton, J., Barnier, A. J., & McIlwain, D. J. F. (2011). We remember, we forget: Collaborative remembering in older couples. *Discourse Processes*, *48*(4), 267–303.

Heasman, B., & Gillespie, A. (2018). Perspective-taking is two-sided: Misunderstandings between people with Asperger's syndrome and their family members. *Autism*, *22*(6), 740–750.

Heldner, M. (2011). Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *The Journal of the Acoustical Society of America*, *130*(1), 508–513.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568.

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in Psychology*, *6*(October).

Hirsch, E. D. (1977). *The philosophy of composition*. Chicago : University of Chicago Press.

Hirst, W., & Manier, D. (2008). Towards a psychology of collective memory. *Memory*, *16*(3), 183–200.

Hirst, W., & Rajaram, S. (2014). Toward a social turn in memory: An introduction to a special issue on social memory. *Journal of Applied Research in Memory and Cognition*, *3*(4), 239–243.

Hobbes, T. (1845). *The English works of Thomas Hobbes of Malmesbury*. London: London : Longman, Brown, Green, and Longman ..., 1845.

Hockett, C. F. (1958). *A course in modern linguistics*. New York: New York : MacMillan, 1958.

Hoek, D. (2018). Conversational exculpature. *Philosophical Review*, *127*(2), 151–196.

Holler, J., Kendrick, K. H., & Levinson, S. C. (2017). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin and Review*, 1–9.

Horton, W. S., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*(1), 1–35.

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the Memory-Based Processing Approach

to Common Ground. *Topics in Cognitive Science*, *8*(4), 780–795.

Huebner, B. (2016). Transactive Memory Reconstructed: Rethinking Wegner's Research Program. *Southern Journal of Philosophy*, *54*(1), 48–69.

Hurley, S. (2001). Perception and action: Alternative views. *Synthese*, *129*(1), 3–40.

Hurley, S. (2008). The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences*, *31*(1), 1–58.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*(OCT), 1–16.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144.

James, W. (1890). *The Principles of Psychology*. *1*.

Jane, E. A. (2014). "Your a ugly, whorish, slut": Understanding E-bile. *Feminist Media Studies*, Vol. 14, pp. 531–546.

Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online hostility. *Ethics and Information Technology*, *17*(1), 65–87.

Kendrick, K. H., & Torreira, F. (2015). The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes*, *52*(4), 255–289.

Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition*, *123*(1), 1–20.

Kentner, G., & Vasishth, S. (2016). Prosodic focus marking in silent reading: Effects of discourse context and rhythm. *Frontiers in Psychology*, *7*(MAR), 1–19.

Kirchgaessner, S. (2019, November 14). Rudy Giuliani says Trump will stay loyal to him but jokes that he has "insurance." *The Guardian*. Retrieved from https://www.theguardian.com/us-news/2019/nov/14/rudy-giuliani-donald-trump-insurance-loyal

Kirchhoff, M. D. (2016). Composition and transactive memory systems. *Philosophical Explorations*, *19*(1), 59–77.

Krantz, M., George, S. W., & Hursh, K. (1983). Gaze and mutual gaze of pre-school children. *The Journal of Psychology*, *113*, 9–15.

Kroll, B. M., & Vann, R. J. (1981). *Exploring speaking-writing relationships : connections and contrasts*. Urbana, Ill.: Urbana, Ill. : National Council of Teachers of English, c1981.

Kukla, R. (2014). Performative force, convention, and discursive injustice. *Hypatia*, *29*(2), 440–457.

LA Times Editorial. (2019, July 14). Trump is truly America's Bigot-in-Chief. *Los Angeles Times*. Retrieved 10 July 2019 from:

https://www.latimes.com/opinion/editorials/la-ed-trump-aoc-squad-ilhan-bigoted-tweets-20190714-story.html

Lang, M., Shaw, D. J., Reddish, P., Wallot, S., Mitkidis, P., & Xygalatas, D. (2016). Lost in the Rhythm: Effects of Rhythm on Subsequent Interpersonal Coordination. *Cognitive Science*, *40*(7), 1797–1815.

Langton, R. (1993). Speech Acts and Unspeakable Acts. *Philosophy and Public Affairs*, *22*(4).

Lefebvre, C. (2004). *Issues in the study of pidgin and creole languages*. John Benjamins Publishing Company.

Levinson, S. C. (1983). *Pragmatics*. Cambridge [Cambridgeshire]: Cambridge Cambridgeshire.

Levinson, S. C. (2006). On the human "interaction engine." In (Enfield, N. J. & Levinson, S. C. Eds.), *Roots of human Sociality: Culture, Cognition and interaction* (pp. 39–69). Oxford: Berg.

Levinson, S. C. (2016). Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends in Cognitive Sciences*, *20*(1), 6–14.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*(JUN), 1–17.

Lewis, D. K. (1979). Scorekeeping in a Language Game. *Philosophical Logic*, *8*(1), 339–359.

Lewis, D. K. (2002). *Convention : a philosophical study*. Oxford: Oxford : Blackwell, 2002.

MacFarquhar, N. (2018, February 18). Inside the Russian Troll Factory: Zombies and a Breakneck Pace - The New York Times. Retrieved February 24, 2020, from NY TImes website: https://www.nytimes.com/2018/02/18/world/europe/russia-troll-factory.html

Magee, J. (1989). *Boethius on signification and mind*. Leiden : Leiden .

Maiese, M. (2016). Dissociative identity disorder and ambivalence. *Philosophical Explorations*, *19*(3), 223–237.

Maitra, I. (2012). Subordinating Speech. In (Maitra, I. & McGowan, M. K. Eds.), *Speech and Harm: Controversies Over Free Speech*. Oxford: Oxford University Press.

March, E., & Marrington, J. (2019). A Qualitative Analysis of Internet Trolling. *Cyberpsychology, Behavior, and Social Networking*, *22*(3), 192–197.

Mari-Beffa, P. (2017). Is talking to yourself a sign of mental illness? An expert delivers her verdict. Retrieved January 21, 2020, from The Conversation website: https://theconversation.com/is-talking-to-yourself-a-sign-of-mental-illness-an-expert-delivers-her-verdict-77058

Marques-Quinteiro, P., Mata, A., Simão, C., Gaspar, R., & Farias, A. R. (2019). Observing Synchrony in Dyads. *Social Psychology*, *50*(3), 174–184.

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social Connection Through Joint Action and Interpersonal Coordination. *Topics in Cognitive Science*, *1*(2), 320–339.

Marsili, N. (2020), Retweeting: its linguistic and epistemic value, *Synthese* (October 2019).

McCosker, A. (2014). Trolling as provocation: YouTube's agonistic publics. *Convergence*, *20*(2), 201–217.

McDougall, W. (1939). *The group mind : a sketch of the principles of collective psychology with some attempt to apply them to the interpretation of national life and character* (2nd ed., ().Cambridge : Cambridge University Press.

McKinney, R. A. (2016). Extracted Speech. *Social Theory and Practice*, *42*(2), 258–284.

Merleau-Ponty, M. (2012). *Phenomenology of Perception* (Landes, D. A. Ed.). Oxon: Routledge.

Metcalf, A. (2011). *OK: The Improbable Story of America's Greatest Word*. New York: Oxford University Press.

Mira d'Ercole, M., van Zanden, J. L., Baten, J., Rijpma, A., & Timmer, M. P. (2014). How Was Life? In *How was life? Global Well-being since 1820,* OECD: Paris.

Mogan, R., Fischer, R., & Bulbulia, J. A. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *Journal of Experimental Social Psychology*, *72*(March), 13–20.

Moore, M. K., & Meltzoff, A. N. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science, New Series*, *198*(4312), 75–78.

Nairne, J. S. (2002). REMEMBERING OVER THE SHORT-TERM: The Case Against the Standard Model. *Annual Review of Psychology*, 53–81.

Nobbes, R. (1682). *The compleat troller : or, The art of trolling, with a description of all the utensils, instruments, tackling, and materials requisite thereto : with rules and directions how to use them : as also a brief account of most of the principal rivers in England*. London: T. James for Tho. Helder.

Parret, H., & Verschueren, J. (Eds.). (1992). *(On) Searle on conversation*. Amsterdam: John Benjamins Publishing Company.

Pavitt, C. (2003a). Why we have to be reductionists about group memory. *Human Communication Research*, *29*(4), 592–599.

Pavitt, C. (2003b). Why we still have to be reductionists about group memory. *Human Communication Research*, *29*(4), 624–629.

Paxton, A., & Dale, R. (2013). Argument disrupts interpersonal synchrony. *Quarterly Journal of Experimental Psychology*, *66*(11), 2092–2102.

Peacocke, C. (2000). Conscious Attitudes, Attention, and Self- Knowledge. In Wright, C., Smith, B. C. & Macdonald, C. (Eds.), *Knowing Our Own Minds*.

Pengelly, M. (2019, July 15). “Go back home”: Trump aims racist attack at Ocasio-Cortez and other congresswomen. *The Guardian*. Retrieved from https://www.theguardian.com/us-news/2019/jul/14/trump-squad-tlaib-omar-pressley-ocasio-cortez

Pepp, J., Michaelson, E., & Sterken, R. K. (2019a). What’s New About Fake News? *Journal of Ethics and Social Philosophy*, *16*(2), 67–94.

Pepp, J., Michaelson, E., & Sterken, R. (2019b). Why we should keep talking about fake news. *Inquiry (United Kingdom)*, *0*(0), 1–17.

Perry, J. (2003). Predelli’s threatening note: Contexts, utterances, and tokens in the philosophy of language. *Journal of Pragmatics*, *35*(3), 373–387.

Phillips, W. (2011). LOLing at tragedy: Facebook trolls, memorial pages and resistance to grief online. *First Monday*, *16*(12). Retrieved from https://journals.uic.edu/ojs/index.php/fm/article/view/3168/3115

Phillips, W. (2015). *This is why we can’t have nice things : mapping the relationship between online trolling and mainstream culture*. Cambridge, Massachusetts: Cambridge, Massachusetts : MIT Press.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02).

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.

Pietroski, P. M. (2018a). *Conjoining Meanings: Semantics Without Truth Values*. Oxford: Oxford University Press.

Pietroski, P. M. (2018b). Précis of Conjoining Meanings: Semantics Without Truth Values. Retrieved from: http://www.terpconnect.umd.edu/~pietro/research/PrecisForCM.pdf

Plato. (1997). *Plato: Complete Works* (J. M. Cooper & D. S. Hutchinson, Eds.). Hackett Publishing Co.

Pomerantz, A. (1978). Compliment Responses: notes on the co-operation of multiple constraints. In Shenkein, J. (ed), *Studies in the Organization of Conversational Interaction Vol. 1*. Academic Press.

Pratt, M. L. (1986). Ideology and Speech-Act Theory. *Poetics Today*, *7*(1), 59–72.

Predelli, S. (1998). I Am Not Here Now. *Analysis*, *58*(2), 107–115.

Predelli, S. (2011). I Am Still Not Here Now. *Erkenntnis*, *74*(3), 289–303.

Quan-haase, A., & Wellman, B. (2005). Local Virtuality in an Organization: Implications for Community of Practice. *Communities and Technologies 2005*, (January), 215–238.

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). *Alignment in multimodal interaction : an integrative framework*. 1–28.

Read, J. (2019, October 22). Tories accused of trolling Remainers by using Comic Sans in pro-Brexit post. *The New European*. Retrieved from https://www.theneweuropean.co.uk/top-stories/tories-accused-of-trolling-remainers-by-using-comic-sans-in-brexit-post-1-6335865

Recanati, F. (1986). On Defining Communicative Intentions. *Mind & Language*, *1*(3), 213–241.

Reddish, P., Tong, E. M. W., Jong, J., & Whitehouse, H. (2019). Interpersonal synchrony affects performers' sense of agency. *Self and Identity*, *00*(00), 1–23.

Ren, Y., & Argote, L. (2011). Transactive memory systems 1985–2010: An integrative framework of key dimensions, atecedents, and consequences. *The Academy of Management Annals*, *5*(1), 189–229.

Rennung, M., & Göritz, A. S. (2016). Prosocial consequences of interpersonal synchrony: A Meta-Analysis. *Zeitschrift Fur Psychologie / Journal of Psychology*, *224*(3), 168–189.

Ribáry, G., Lajtai, L., Demetrovics, Z., & Maraz, A. (2017). Multiplicity: An explorative interview study on personal experiences of people with multiple selves. *Frontiers in Psychology*, *8*(JUN), 1–10.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, *18*(5), 407–413.

Ridge, M. (2006). Sincerity and expressivism. *Philosophical Studies*, *131*, 487–510.

Rini, R. (2017). Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal*, *27*(2), E-43-E-64.

Roberts, F., & Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America*, *133*(6), EL471–EL477.

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes*, *48*(5), 331–354.

Romdenh-Romluc, K. (2002). Now the French are invading England! *Analysis*, *62*(1), 41–46.

Roser, M., & Ortiz-Ospina, E. (2020). Literacy. Retrieved from *Our World in Data* website: https://ourworldindata.org/literacy

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735.

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., & Grossman, R. B. (2017). Neurotypical Peers are Less Willing to Interact with Those with Autism based on Thin Slice Judgments. *Scientific Reports*, *7*(October 2016), 1–11.

Saul, J. M. (2017). Racial Figleaves , the Shifting Boundaries of the Permissible , and the rise of Donald Trump. *Philosophical Topics*, *45*(2), 97–116.

Saul, J. M. (2018). Dogwhistles, political manipulation, and philosophy of language. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New Work on Speech Acts* (pp. 360–383).

Schafer, J. (1981). The Linguistic Analysis of Spoken and Written Texts. In B. M. Kroll & R. J. Vann (Eds.), *Exploring speaking-writing relationships : connections and contrasts*. Urbana, Ill.: Urbana, Ill. : National Council of Teachers of English, c1981.

Schegloff, E. A. (1992). To Searle On Conversation: A note in return. In H Parret & J. Verschueren (Eds.), *(On) Searle on conversation* (pp. 113–128). Amsterdam: John Benjamins Publishing Company.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, *53*(2), 361–382.

Schegloff, E. A., & Sacks, H. (1973). Opening up Closings. *Semiotica*, Vol. 8, p. 289.

Schiffer, S. (1972). *Meaning* ([Paperback). Oxford: Oxford : Clarendon, 1972.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research*, *35*(2), 189–213.

Scott, K. (2015). The pragmatics of hashtags: Inference and conversational style on Twitter. *Journal of Pragmatics*, *81*, 8–20.

Searle, J. R. (1969). *Speech acts : an essay in the philosophy of language*. Cambridge: Cambridge : Cambridge University Press, 1969.

Searle, J. R. (1990). Collective Intentions and Actions. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication*. MIT Press.

Searle, J. R. (1992a). Conversation. In H Parret & J. Verschueren (Eds.), *(On) Searle on conversation*. Amsterdam: John Benjamins Publishing Company.

Searle, J. R. (1992b). Conversation Reconsidered. In H Parret & J. Verschueren (Eds.), *(On) Searle on conversation* (pp. 137–147). John Benjamins Publishing Company.

Sicoli, M. A., Stivers, T., Enfield, N. J., & Levinson, S. C. (2015). Marked Initial Pitch in Questions Signals Marked Communicative Function. *Language and Speech*, *58*(2), 204–223.

Sidelle, A. (1991). The Answering Machine Paradox. *Canadian Journal of Philosophy*, *21*(4), 525–539.

Simon, D. (2002). *The Wire*. HBO TV Series.

Sperber, D., & Wilson, D. (1986). *Relevance Communication and Cognition*. Wiley-Blackwell.

Spradley, T. S., & Spradley, J. P. (1985). *Deaf Like Me*. Washington DC: Gallaudet University.

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.

Stalnaker, R. (1999). Introduction. In *Context and content* (pp. 1–28). Oxford: Oxford University Press.

Stalnaker, R. (1999 [1970]). Pragmatics. In *Context and content* (pp. 31–46). Oxford: Oxford University Press.

Stalnaker, R. (1999 [1974]). Pragmatic presuppositions. In *Context and content* (pp. 47–62). Oxford: Oxford University Press.

Stalnaker, R. (1999 [1978]). Assertion. In *Context and content* (pp. 78–95). Oxford: Oxford University Press.

Stalnaker, R. (1999 [1998]). On the representation of context. In *Context and content* (pp. 96–114). Oxford: Oxford University Press.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy 25*, 701–21.

Stalnaker, R. (2014). *Context*. Oxford: Oxford University Press.

Stanley, J. (2018). *How fascism works: the politics of us and them*.

Steinthal, H. (2013). *Geschichte der Sprachwissenschaft bei den Griechen und Römern : Mit besonderer Rücksicht auf die Logik*.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Stokel-Walker, C. (2019, October). "They're doing this badly on purpose": Why the Tories' latest online ads look so ugly. *The New Statesman*. Retrieved from https://www.newstatesman.com/politics/media/2019/10/they-re-doing-badly-purpose-why-tories-latest-online-ads-look-so-ugly

Stokke, A. (2018). *Lying and insincerity* (First edit). Oxford: Oxford : Oxford University Press, 2018.

Striano, T., Henning, A., & Stahl, D. (2006). Sensitivity to interpersonal timing at 3 and 6 months of age. *Interaction Studies*, *7*, 251–271.

Striano, T., & Liszkowski, U. (2005). Sensitivity to the context of facial expression in the still face at 3-, 6-, and 9-months of age. *Infant Behavior and Development*, *28*(1), 10–19.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology and Behavior*, *7*(3), 321–326.

Theiner, G. (2013). Transactive Memory Systems: A Mechanistic Analysis of Emergent Group Memory. *Review of Philosophy and Psychology*, *4*(1), 65–89.

Tollefsen, D. P. (2006). From extended mind to collective mind. *Cognitive Systems Research*, *7*(2–3), 140–150.

Tollefsen, D. P., & Dale, R. (2012). Naturalizing joint action: A process-based approach. *Philosophical Psychology*, *25*(3), 385–407.

Tollefsen, D. P., Dale, R., & Paxton, A. (2013). Alignment, Transactive Memory, and Collective Cognitive Systems. *Review of Philosophy and Psychology*, *4*(1), 49–64.

Tuomela, R. (2000). *Cooperation: A Philosophical Study*. Dordrecht: Springer.

Tuomela, R. (2005). Two Basic Kinds of Cooperation. In D. Vanderveken (Ed.), *Logic, Thought and Action* (pp. 79–107).

Tuomela, R. (2011). Cooperation as joint action. *Analyse Und Kritik*, *2011*(1), 65–86.

Tye, M. (2005). *Consciousness and Persons: Unity and Identity*. Cambridge, Mass.: MIT Press.

Vachek, J. (1973). *Written language : general problems and problems of English.* The Hague: The Hague : Mouton, 1973.

Vachek, J. (1989). *Written language revisited* (P. Luelsdorff, P. A. Luelsdorff, & P. (Philip) Luelsdorff, Eds.). Amsterdam : Amsterdam .

von Fintel, K. (2008). What Is Presupposition Accommodation, Again? *Philosophical Perspectives*, *22*, 137–170.

von Zimmermann, J., & Richardson, D. C. (2016). Verbal synchrony and action dynamics in large groups. *Frontiers in Psychology*, *7*(DEC), 1–10.

Vygotskii, L. S. (1986). *Thought and Language* (A. Kozulin, Ed.). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=49350&site=ehost-live

Walker, M. B., & Trimboli, C. (1982). Smooth Transitions in Conversational Interactions. *The Journal of Social Psychology*, *117*, 305–306.

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behaviour* (pp. 185–208). New York: Springer.

Wegner, D. M. (1995). A Computer Network Model of Human Transactive Memory. *Social Cognition*, *13*(3), 319–339.

Wegner, D. M., Giuliano, T., & Hertal, P. T. (1985). Cognitive Interdependence in Close Relationships. In W. J. Ickes (Ed.), *Compatible and Incompatible Relationships* (pp. 253–276). New York: Springer.

Weldon, M. S. (2000). Remembering as a social process. *Psychology of Learning and Motivation*, *40*, 67–120.

Wikström, P. (2014). Srynotfunny: Communicative functions of hashtags on twitter. *SKY Journal of Linguistics*, *27*(Cmc), 127–152.

Wilcke, C. (2000). Wer las und schrieb in Babylonien und Assyrien: Überlegungen zur Literalität im Alten Zweistromland. *Sitzungsberichte Der Bayerischen Akademie Der*

Wissenschaften. Philosophisch-Historische Klasse, *2000*(Jahrgang 2000, Heft 6).

Zeshan, U. (2015). “Making meaning”: Communication between sign language users without a shared language. *Cognitive Linguistics*, *26*(2), 211–260.

Zeshan, U. (2019). Task-response times, facilitating and inhibiting factors in cross-signing. *Applied Linguistics Review*, *10*(1), 9–29.