

UCLA

UCLA Electronic Theses and Dissertations

Title

Situated Conversation: The Role of Attention in Communication

Permalink

<https://escholarship.org/uc/item/7868j3b3>

Author

De Leon, Christian Martin

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Situated Conversation:
The Role of Attention in Communication

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Philosophy

by

Christian Martin De Leon

2022

© Copyright by
Christian Martin De Leon
2022

ABSTRACT OF THE DISSERTATION

Situated Conversation:
The Role of Attention in Communication

by

Christian Martin De Leon

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2022

Professor Samuel John Cumming, Co-Chair

Professor Gabriel Jae Greenberg, Co-Chair

Regular, everyday conversations do not occur in a vacuum. The physical and social context in which a conversation takes place is utilized through the course of that conversation—it is *situated*. The ways in which a context can be used vary. Sometimes we implicitly and unconsciously utilize context in interpreting others. While at a restaurant, you would effortlessly interpret “What are you going to have?” not as a general question disconnected from the current situation, but as a question about that particular restaurant’s menu. Other times we overtly utilize context to aid in communication. While out on a walk in the city, I might point to an ice cream cone that has fallen onto the street and say “Somebody’s having a bad day”, and it would be clear what I am talking about.

I argue in this dissertation that the psychological category of *attention* is key for understanding how situated conversations operate. In contrast to viewing communication as a process of pure information exchange, I argue that communication is helpfully analyzed as a process of attention management. Our utterances, gestures, and facial expressions do not

serve to merely convey information, but serve to coordinate our attentional states.

I begin by focusing on a particular communicative phenomenon that I call *rich demonstration*. It is, roughly, a deictic (pointing) gesture that functions to communicate an entire thought, in contrast to fixing the reference of a demonstrative expression such as ‘that’. Just by making something salient, one can seemingly *say something*. I argue that rich demonstration is a speech act in its own right, and provide a formal semantic framework for modeling its communicative effect in discourse.

I then proceed to analyze *salience* itself. The notion has proven useful for understanding a number of communicative phenomena, including reference resolution, certain kinds of pragmatic inference, quantifier domain restriction, and more. Focusing on cases of reference, I argue that the linguistic notion of salience is best understood in terms of the psychological notion of *mutual attention*. I discuss competing options, including mutual knowledge and varieties of joint attention that are weaker than mutual attention, and argue that only mutual attention suffices for establishing conversational salience. The picture that emerges is one on which conversational moves function in part to coordinate the attentional states of interlocutors.

I take advantage of these considerations to argue that discourse information states are genuinely structured—they represent different types of information that can be updated independently. I argue that such structural models of conversation cannot be reasonably reduced to unstructured models, such as the Common Ground model, which flatten discourse information into one kind.

This dissertation therefore develops a view of conversation as attention management. This is a perspective on situated conversation on which communicative acts are both restricted by facts about attention and function to coordinate the attentional states of interlocutors.

The dissertation of Christian Martin De Leon is approved.

Jessica L. Rett

Joshua David Armstrong

Gabriel Jae Greenberg, Committee Co-Chair

Samuel John Cumming, Committee Co-Chair

University of California, Los Angeles

2022

*To my parents,
for showing me the joy
in learning,
in kindness,
and in talking with others.*

TABLE OF CONTENTS

Introduction	1
1 Speech Acts Without Speech	6
1.1 The Phenomenon	8
1.2 Rich Demonstrations as Speech Acts	12
1.2.1 An Informative Point	13
1.2.2 Discourse Effects	15
1.2.3 Goats and Herod’s Intentions	17
1.3 Enriching Demonstrations	21
1.3.1 Indicating Situations	22
1.3.2 Forming a Coherent Message	28
1.3.3 Building a Speech Act	35
1.4 Conclusion	36
2 Pointing to Communicate	37
2.1 Extracting Situations from Context	38
2.1.1 Indicating Situations	38
2.1.2 Indexicality	40
2.2 Coherent Discourse Moves	42
2.2.1 Moves on the Record	42
2.2.2 Adding Discourse Structure	48
2.2.3 Worked Cases	53

2.3	Future Research	60
2.4	Conclusion	64
3	Conversational Salience and Mutual Attention	66
3.1	Why Model Salience	68
3.2	Conversational Salience, Psychological Attention	73
3.3	Salience and Varieties of Joint Attention	76
3.3.1	Overlapping Attention	77
3.3.2	Weak Common Attention	79
3.3.3	Strong Common Attention	81
3.3.4	Mutual Attention	86
3.4	Knowledge of Attention	89
3.4.1	Salience in the Common Ground	91
3.4.2	A Privileged Common Ground	95
3.5	Conversation as Attention Management	97
4	Common Ground and the Structure of Context	100
4.1	How Common Ground is Used	102
4.2	Pseudo-Structure	109
4.2.1	Meta-Propositions	114
4.3	Grounding Structural Facts	118
4.3.1	Truth in Virtue of Acceptance	119
4.3.2	Acceptance in Virtue of Truth	120
4.4	Benefits of Structure	124

4.4.1	Non-Ideal Conversation	124
4.4.2	Cooperative Social Lives	125
4.5	Conclusion	126
	Bibliography	128

LIST OF FIGURES

3.1	Levels of Joint Attention	77
4.1	Unstructured Order of Explanation	119
4.2	Structured Order of Explanation	121

ACKNOWLEDGMENTS

I'd like to begin by thanking my advisors, Gabe Greenberg and Sam Cumming. Beginning graduate school, I had not planned on being a philosopher of language, but I quickly found myself wanting to learn more about the sub-field. I asked Gabe if he had recommendations for how to go about this. Though neither of us knew it, that was the day he became my advisor. In addition to connecting me with the rest of my future committee, Gabe immediately started mentoring me with the combination of enthusiasm, kindness, and sharpness that anybody who works with him is lucky to experience. Out of our many, many meetings, I cannot think of a single one where I did not leave feeling energized about the work ahead. Whether I needed help talking through data, planning a writing schedule, or structuring an argument, Gabe was ready with encouragement, toughness, and philosophical clarity.

Sam has been an invaluable resource for both thinking through big-picture ideas and working out very specific details, but I think I was usually wrong about which one was really happening during most of our meetings. It's quite fun when this happens, because there comes a point after the meeting when you realize the previously unnoticed benefit of your conversation. I often find myself looking over Sam's comments on my old—sometimes totally abandoned—drafts because I know that I didn't fully absorb them the first time.

I'm very grateful to have Jessica Rett and Josh Armstrong as the other members of my committee. I want to thank Jessica for introducing me to formal semantics and pragmatics, always complicating what I thought were simple data, and teaching me that there's no such thing as a sentence without context. I want to thank Josh for being my go-to resource for learning about cognitive science and for helping me navigate the choppy waters where philosophy, linguistics, and psychology meet.

Something that all of my committee members have in common is that, beyond being extraordinarily helpful, they're just super fun to talk to. I'm so grateful to have gotten the chance to nerd-out with them about how weird and cool communication is.

I have been influenced by and benefited from many faculty at UCLA. I particularly want to thank Tyler Burge, Dylan Bumford, Katie Elliott, Pamela Hieronymi, David Kaplan, Yael Sharvit, Seana Shiffrin, and Sean Walsh for having a decisive impact on the kind of researcher, teacher, and colleague that I aim to be.

I want to thank my teachers and advisors from IUPUI—Luise Morton, Greg Keller, and John Tilley—for seeing potential in me as a young undergraduate and investing the time and effort to help me get into graduate school.

I doubt that there's very much in life that is more valuable than friendship. My colleagues-turned-friends are what really made my graduate school experience what it was. I am so grateful to Zach Biondi, Tristen Cardwell, Milo Crimi, Gabe Dupre, Andrew Flynn, Catherine Hochman, Gabby Johnson, John Kardosh, Amber Kavka-Warren, Kevin Lande, Jürgen Lipps, Esther Nikbin, Torsten Odland, Jonah Wolf Ragir, Ayana Samuel, and Aaron West for being a stellar community of friends and interlocutors.

I want extend my heartfelt appreciation to those whose personal companionship has especially influenced who I am today. Ian Boon is one of the sharpest and quickest thinkers that I know. His ability to clearly distill complex arguments and linguistic analyses has taught me a great deal, and he has been a true role-model as a philosopher of language. Kim Johnston has been the person I've counted on to give it to me straight about most things, and I'm better for it. For questions big and small, there aren't many people I trust as much as Kim. Bill Kowalsky brings his full intellectual resources to practically every topic, and I deeply admire the resulting enthusiasm in every conversation. Few things are as fun or rewarding as doing a deep-dive on something with Bill, be it philosophical or not. Alex Pelaez has an astonishing work ethic that is a joy to participate in. She knows how to prioritize what's important in life and how to make the most of every second. Joining in has been thoroughly fulfilling. Ekin Zeytinoglu has had my back in lots of different ways. His advice on topics both philosophical and personal is always keen and to the point. One of these days we'll find a topic to co-author about.

I'll conclude by thanking my family, whose love and support has been with me my whole life. I want to thank my brother, Matthew, for allowing me to talk his ear off about philosophy and for engaging with me in many hugely helpful conversations over the years. I'm grateful to my grandmother, Jessene, for always being ready to talk and laugh about our different regionalisms. Quiero agradecer a mi abuelita, Socorro, por su apoyo y confianza entusiasta. Finally, I thank my parents, Juan and Tish, who provided all I could ask for, and to whom this dissertation is dedicated. Thank you for your unending love.

VITA

- 2019 C.Phil in Philosophy, University of California, Los Angeles
- 2016 M.A. in Philosophy, University of California, Los Angeles
- 2015 B.A. in Philosophy (Highest Honors), Indiana University-Purdue University, Indianapolis

PUBLICATIONS

“Pointing to communicate: the discourse function and semantics of rich demonstration”. (Forthcoming) In *Linguistics & Philosophy*.

“A semantics of face emoji in discourse”. with Patrick Grosz, Gabriel Greenberg, and Elsi Kaiser. (Forthcoming) In *Linguistics & Philosophy*.

“Situating rich demonstrations in discourse”. (2019) In *Proceedings of the 22nd Amsterdam Colloquium*. pp.543–552.

INTRODUCTION

While out at a restaurant, as we're finishing our dessert, you might point at the portion of cake left on my plate and raise your eyebrows earnestly. I might then slide the plate toward you and say "Hurry, though" as I glance toward an employee stacking some chairs. I understand your intent to ask for some cake, you understand my action as giving it to you, and you understand that the reason I said to hurry is that I think the restaurant is closing soon. Our back and forth succeeds because our conversation is *situated*—the words we speak, the gestures we use, and the facial expressions we make all have their particular impacts on our dialogue because of our physical and social context.

This dissertation is about communication as a situated phenomenon. I argue that the psychological category of *attention* plays a key explanatory role in situated communication. In contrast to viewing communication as a process of pure information exchange, I argue that communication is helpfully analyzed as a process of attention management. Our utterances, gestures, and facial expressions do not serve to merely convey information, but serve to coordinate our attentional states.

This dissertation can be divided into two parts. In the first, Chapters 1–2, I discuss a particularly striking phenomenon that makes vivid how important attention management can be in normal conversation.¹ I call it *rich demonstration*. One richly demonstrates when they make something salient in some way, typically with a deictic (pointing) gesture, and in so doing communicate an entire thought. Rich demonstrations can occur parallel to speech, but to isolate the phenomenon I focus on cases in which a gesture's contribution to discourse is easily separable from any surrounding speech. I discuss several examples, but the following is the central one that anchors the discussion through both chapters:

¹The bulk of Chapter 2, together with parts of Chapter 1, is to be published as "Pointing to communicate: the discourse function and semantics of rich demonstration" in *Linguistics & Philosophy*.

- (1) My parents are going to be $\overbrace{\text{furious}}^{\text{ARM TURN: tat}}$

Here, the speaker turns their arm as they say “furious”, revealing a fresh tattoo. Intuitively the gesture communicates the reason why the speaker’s parents are going to be upset: because of this tattoo. This is importantly different from the reference-fixing function traditionally associated with deixis, as in (2):

- (2) $\overbrace{\text{This is}}^{\text{ARM TURN: tat}}$ going to make my parents furious.

The gesture here does not communicate very rich content, but instead helps to fix the referent of the demonstrative “this”. The motivating observation is that the same physical movement, the turn of an arm, can intuitively make a different, more complex, discourse contribution depending on the context.

In Chapter 1 I argue that rich demonstration is a genuine speech act in its own right. On the basis of its functional profile in conversation and the similarities it bears to assertion, I conclude that rich demonstration should be considered a complete discourse contribution. This is in contrast with views that would take rich demonstration to be informative but not communicative in a theoretically robust sense. I then argue for a framework that can explain how a deictic gesture can come to perform the speech act of rich demonstration. The framework takes advantage of *discourse coherence theory*, which studies the conceptual connections between speech acts. I argue that there exist conventions about how rich demonstrations can be used and that these conventions guide how a rich demonstration is interpreted.

I give the details of this framework in Chapter 2. Having argued that rich demonstrations are complete, meaningful discourse contributions, I turn to providing a formal semantics designed to model those contributions. On the formal treatment I provide, deictic gestures can indicate entire situations, not just objects. The complexity of a situation coupled with the

semantic content of the operative coherence relation enables the retrieval of the proposition that was intentionally communicated.

The upshot of this first part is that complete speech acts do not need to rely on linguistic vehicles to be performed. Instead, actions which are properly thought of as acts of attention management suffice for performing speech acts. Just by making something conversationally salient, one can communicate an entire thought. Importantly, the mechanism by which a thought is conveyed is one of directing attention.

The second part, Chapters 3–4, explores in more detail the notion of *managing attention*. In Chapter 3, I aim to analyze the linguistic notion of *conversational salience* in terms of attention, where *attention* is a specific kind of psychological state.

It is common in semantic and pragmatic explanations to appeal to something's being salient in a conversation. Intuitively, the right meanings of (3) and (4) are recovered successfully thanks to the fact that some particular thing is salient during the conversation.

(3) *While at The Louvre, admiring the Mona Lisa:*

I can't believe he was such a talented painter as well as a brilliant inventor.

(4) *Right after hearing a large crash of dishes from the kitchen:*

Whoa, are you okay?

'He' refers to Leonardo da Vinci in (3), and the speaker's concern in (4) is clearly in relation to whatever caused the loud crash. These facts are naturally explained in terms of what is salient in the conversation—the painting by da Vinci in the first case, and the sound in the second. Whereas the first half of the dissertation focuses on the effects of making something salient, Chapter 3 investigates what salience itself is. I argue that it is best understood in terms of *mutual attention*, a psychological state wherein two people have the same thing as the object of their cognitive attention and they each secondarily attend to the other's primary act of attention. My claim is that conversational salience is not a notion sui generis to

linguistic theory, but rather can be understood in specific psychological terms. In particular, I argue that (a) weaker kinds of attentional state do not suffice for establishing something as conversationally salient, and (b) conversational salience is not established thanks to the propositional attitudes held by the conversation's interlocutors. Instead, a complex kind of joint attention with another, *mutual* attention, is the psychological notion related to conversational salience.

Looming large throughout the dissertation is the highly influential framework of Robert Stalnaker (1978, 2002, 2014). He theorizes about the *common ground*, an abstract state of shared information between two or more people. For Stalnaker, the common ground is the core factor in explaining both particular communicative phenomena as well as in explaining the nature of communication itself. The common ground is the set of all propositions that a conversation's participants mutually accept. As we communicate with another, we add more and more propositions to our common ground, thereby growing our store of background presuppositions. Central to this picture is the notion of a *manifest event*—an event so obvious that we cannot help but accept it as having taken place, and so something that automatically becomes part of our common ground. If a goat were to walk into the room while we were having a conversation, its sudden presence would automatically be presumed to be common ground, to borrow the classic example from Stalnaker (1978).

I argue in Chapter 1 that the common ground framework alone does not explain why rich demonstration is possible or how it functions. Though it traditionally is thought to explain how coordination on manifest events is achieved without explicit acknowledgement, it lacks an explanation of how an intentional act of attention direction can result in a fully communicative conversational move.

In Chapter 3, I argue that conversational salience is not properly understood as a feature of the common ground. Something's being conversationally salient is not a matter of our having the right propositional attitudes, as would be required for a Stalnakerian analyzing salience. Instead, the psychological category that is relevant to salience, attention, is less

cognitively complex.

These considerations culminate in Chapter 4, wherein I argue that models of communication that utilize informational structure are not reducible to the common ground. It is a common practice in formal semantics and pragmatics to model a discourse as a structured information state, with different elements of different types that can potentially be updated independently. Often the common ground is one of these elements, alongside a stack of salient potential referents, a hierarchical list of questions guiding the discussion, and more. According to a position that is explicitly endorsed by some and tacitly accepted by others, this practice is nothing more than a notational convenience. A discourse's information state is actually nothing more than the common ground, and any apparent structure is explained in terms of common ground coordination, the thought goes. On the contrary, I argue that the details required by this unstructured approach to context are implausible and undermine central motivations behind the common ground to begin with. Instead, models of discourse that incorporate structure do so with good reason; genuine structure explains how different kinds of information are separately tracked by conversationalists as well as how they can interact. Those interactions are what enable attention management to play such a robust role in communication—we coordinate on what from our surroundings is salient in the conversation, which unavoidably impacts what we mean.

CHAPTER 1

Speech Acts Without Speech

When Herod presented a charger holding the head of John the Baptist to Salome, he intended that she come to believe that the Saint was dead. Grice (1957) claimed that the presentation, despite Herod's intention, was intuitively not a case of *non-natural* meaning; it did not mean that John the Baptist was dead in the way that the bell's ringing means that the bus is full. Instead, Grice thought, the situation itself—the head on the charger—*naturally* meant that John the Baptist was dead, in the way that spots mean measles. His reasoning was that the situation in front of Salome made the death perceptually obvious. So the act of drawing attention to it could not communicate that John the Baptist had died.¹ Swayed by this line, one might conclude that directing attention (by pointing, showing, or otherwise indicating) cannot be communicative in any robust sense.

Indeed, traditional pictures of the role of deixis in communication are *referentialist*—they would say that deictic gestures such as pointing with an index finger can only fix or assist in fixing the referential content of a demonstrative linguistic expression. Standard examples of such expressions since Kaplan (1989b) are “this”, “that”, “there” and so on. These expressions are indexical—their referential content can vary between contexts—and demonstrations that accompany them supply their referents. While theories vary regarding the precise reference-fixing mechanisms behind demonstrative expressions, most share the assumption that the role deictic gestures play in communication is limited to fulfilling

¹Strictly, Grice was not concerned with conditions for communication, only meaning. These are related, but distinct. See Neale (1992, Sec.6) for a helpful discussion of Grice's precise aims.

demonstrative reference.²

I intend to argue for an opposing view: directions of attention can be, and often are, communicative. On my view, pointing to something can be a speech act in itself. In fact, I ultimately hope to show that Herod's revelation is an example of just such a case. I call this kind of speech act a *rich demonstration*. It is, roughly, where a direction of attention (e.g. with a pointing gesture) is used to *say something* in a conversation. It is because deictic gestures can be used to *say things*—a view for which I argue in Sec.1.2—that I believe they can constitute full speech acts. I will argue in Sec.1.3 that this is possible because these gestures indicate not objects but entire situations, raising them to salience in conversation. Such manipulation of salience can be used to express propositional contents. Not just any content can be expressed; particular communicative conventions govern the range of possible interpretations. This is what explains how contents are expressed at all.

If I am right, then *speech act* is a misnomer, and rich demonstration is a case of a speechless speech act. It is of course trivial that not all speech acts are acts of speech. One needs only to consider signed languages to realize this. But the intended, more interesting conclusion is stronger: speech acts need not arise from acts of linguistic meaning. Rich demonstrations arise from acts of indication.³

I introduce cases of rich demonstration in Sec.1.1, then in Sec.1.2 argue that they should be analyzed as full speech acts. I show that rich demonstration functions in conversation much like *assertion*. On the basis of core features shared by each I conclude that rich demonstration, like assertion, is a speech act in its own right. In particular, both assertions

²See Elbourne (2008), King (2001, 2008), Maier (2009), Roberts (2002), and Zeevat (1999) for analyses that depart from Kaplan (1989b) in significant ways, involving the interaction with communicative intent, similarity to presuppositions and quantifiers, and more.

³MacFarlane (2011) observes that the speech act of assertion (which, I will show, is highly similar to rich demonstration) itself is not necessarily linguistic, and mentions assertions made with conventionalized gestures (emblems) to exemplify the point. This makes sense because emblems are conventionally associated with contents of their own. But my proposal is that there are non-linguistic acts that (*a*) lack any prior connection to propositional content, yet (*b*) nevertheless constitute complete speech acts.

and rich demonstrations function to coordinate agents on what they mutually accept as true.

The puzzle then is in how deictic gestures can accomplish this complicated function. Despite lacking anything like a compositional semantics, a deictic gesture can ultimately result in informing an audience of some propositional content. I propose in Sec.1.3 that tools from *coherence theory* (Asher and Lascarides 2003; Hobbs 1979, 1990; Kehler 2002) can help to explain this. I posit interpretive rules that dictate how directions of attention can combine with speech to form a coherent message. According to these rules, a rich demonstration offers an *explanation of* or *elaboration on* some proposition from preceding discourse. Rich demonstrations, then, are *relational* speech acts (Asher and Lascarides 2003).

The picture is this: just as one can perform an assertion to convey a proposition, one can perform a rich demonstration to convey one as well. Assertions arise out of utterances that have internal features whose semantics determine the proposition conveyed. But rich demonstrations arise out of deictic gestures that have no such features. Instead, the act of drawing attention to a situation constitutes performance of a rich demonstration in virtue of conventions for establishing coherence in conversation. These conventions in turn determine the content conveyed by the speech act. The resulting analysis relies on the complexity of indicated situations and the limited conventions for how they can conceptually relate to prior discourse.

1.1 The Phenomenon

To introduce rich demonstrations, I will begin by contrasting two cases. Throughout the chapter I will use lines to mark the beginning and end points of deictic gestures, to represent their timing in relation to the utterance of the sentence. So the gesture in (5), for example, begins as the speaker is saying “this” and ends on “is”. Gestures will be labelled “FORM: x” where FORM is the form the gesture takes (e.g. an extended index finger, an open-palm

hand, or even a head nod) and x is the object or scene indicated.⁴ In the following cases, the speaker makes the exact same gesture: they turn their arm, revealing a fresh tattoo on the inner forearm.

(5) $\overline{\text{This is}}^{\text{ARM TURN: tat}}$ going to make my parents furious.

(6) My parents are going to be $\overline{\text{furious}}^{\text{ARM TURN: tat}}$

The speakers in (5) and (6) *prima facie* communicate the same thing to their audiences: that the speaker's parents will be upset with them for getting the indicated tattoo. But that message is communicated in different ways; the gestures are doing different things in each case. In (5), the demonstration is referential in the sense described above: it fixes or assists the interpretation of the demonstrative 'this', depending on your theory. The rest of what we understand is given by what the speaker actually asserts. But in (6), the speaker does not explicitly mention the existence of any reason for the parents to be upset, only that they will be. Furthermore there is no demonstrative expression whose reference the gesture can fix or assist in fixing. Nevertheless we easily interpret the speaker of (6) as communicating that their parents will be upset because of the new tattoo. To put the point in a theoretically neutral way: unlike in (5), we know about the reason for the parents' being upset in (6) thanks to the gesture. To put it a little less neutrally: the gesture in (6) seems to be contributing a proposition to the discourse, whereas the gesture in (5) merely supplies a referent. Compare (6) with (7):

- (7) a. My parents are going to be furious.
 b. I got a new tattoo.

⁴As will become clear, the surface features of rich demonstrations, such as hand shape or movement speed, do not matter much for the purposes of this chapter. I therefore limit the notation to rough timing and general description of the gesture.

The first assertion in (7) is the same as the one made in (6). As before, the speaker successfully conveys, without explicitly mentioning it, what will cause their parents to be furious. Furthermore in both cases we know the cause of their being upset because of the speaker's second action, after the initial assertion. The key difference is that the second step of (7) is itself an assertion—a speech act that obviously conveys propositional content—whereas the corresponding step of (6) is merely the turn of an arm.⁵

Grice did not believe that Herod communicated *John the Baptist is dead*. I take up the reason for this Sec.1.2.3, but note that he (1969) did take a similar case to be genuinely communicative (more accurately: a case of non-natural meaning). I agree, and think it is a case of communication because it is a case of rich demonstration. (8) is an adapted version. In it, *B* replies to *A*'s question by rolling up their jeans to show a bandaged leg and saying nothing, intuitively conveying that they cannot accept the invitation due to injury.

- (8) a. *A*: Do you want to go play some squash?
b. OPEN HAND: leg

When *B* responds just with a gesture, they successfully communicate that they are unable to accept *A*'s invitation.⁶ Just by drawing attention to something in the vicinity, *B* communicates a complete thought.

Hunter et al. (2018) present a version of the following case, where the speaker nods toward a large scratch on the wall while speaking:

⁵Hunter et al. (2018) make a similar observation regarding the analogy to full assertions. I present a case of theirs in (9) and discuss their theory in Sec.1.2.3.

⁶Grice considered this to be a case of non-natural meaning because the communicated content is not perceptually obvious (*B* communicates a rejection of an offer, and plausibly a reason. But they don't communicate that the leg is bandaged).

- (9) a. $\frac{\text{NOD: scratch}}{\text{I moved the table into the living room this morning.}}$
 b. I had to buy some new paint.

The speaker of (9) does not nod to merely inform the hearer that the wall is scratched. Instead, they intentionally use the nod to inform the hearer of what happened while moving the table—the speaker scratched the wall with it. That information is necessary for understanding the point of the second utterance and how the need to buy paint arose from the events earlier in the day.

In (10), the speaker points toward an unopened box of cookies on the kitchen counter.

- (10) a. $\frac{\text{POINT: cookies}}{\text{I went to the grocery store earlier.}}$
 b. ...if you want any.

The speaker in (10) does not merely present one fact and a disconnected situation in the world. They instead draw attention to something in the surroundings for two purposes. First, they communicate that they bought the cookies while on the trip. Second, the *biscuit*-conditional (Austin 1956; DeRose and Grandy 1999) that the speaker uses to offer cookies succeeds because ‘any’ is restricted by the gesture.

(11) takes place in a context where Maria is helping Carlos hunt for a new apartment. While walking up to a new complex, Carlos gets Maria’s attention and nods toward a community pool visible behind the building.

- (11) a. C : NOD: pool
 b. M : You would get such a bad sunburn.

Maria comments on what would happen in a very specific hypothetical scenario, one in which Carlos rents the apartment and decides to frequent the complex’s pool. She can only

feliculously make that comment given that she understands Carlos to be raising that as a live option with his gesture. Note that if Carlos were to replace the gesture with an utterance of “There’s a pool!”, a functionally identical discourse would unfold.

1.2 Rich Demonstrations as Speech Acts

The cases above help clarify the phenomenon. Rich demonstrations arise from deictic gestures and have more complex contributions to discourse than do typical directions of attention. Gestures that result in rich demonstrations can take a number of forms (index finger, open hand, head nod, etc.). Their timing with respect to speech seems to matter little, if at all—the gesture can come before, during, or after accompanying speech. No content is “read off of” the gesture’s internal shape or movement, as is the case with iconic gestures (Kendon 2004). The content contributed to discourse by a rich demonstration is highly context-sensitive, unlike conventionalized emblems (Kendon 2004; McNeill 1992).⁷

I argue in this section that drawing attention to something can constitute performance of a speech act—this is what a rich demonstration is. I will argue for this on the basis of the similarities between rich demonstration and assertion, a paradigm speech act. Both when one richly demonstrates and when one asserts: *(a)* one intentionally informs their interlocutor of something, *(b)* that new information determines how the conversation can proceed, and *(c)* one becomes publicly committed to that new information.⁸

In making this argument, I emphasize the the importance of *function* for an action’s status as a speech act (Murray and Starr 2018, 2020). The approach characterizes speech acts as

⁷An example iconic gesture: sweeping the hand high to low in an arc while saying “the vase broke” to communicate that it broke by falling. An example emblem: shrugging the shoulders to communicate *I don’t know*.

⁸These features are all, of course, related. That assertions are intentionally informative (quality *(a)*) is a central reason for theorists like Lewis (1979) and Stalnaker (1978) why they have their regular effects on future discourse (quality *(b)*). For theorists like Searle (1969), the act of assertion can succeed in informing (quality *(a)*) because it is fundamentally an act of committing oneself to the truth of some proposition (quality *(c)*)—but see (Pagin 2004, 2009) for critical discussion.

those actions that coordinate the social attitudes of agents in particular conventionalized ways. My argument is that rich demonstrations coordinate agents in ways deeply analogous to assertion. Namely: both function to coordinate agents on the propositions they mutually take to be true.

My position is then firmly situated within the tradition of discourse dynamics, or dynamic pragmatics (Lewis 1979; Roberts 2012a; Stalnaker 1978).⁹ Researchers in this tradition theorize about the finite taxonomy of speech acts, which typically includes at least assertions, questions, and commands. Speech acts are classified according to how they affect context. My objective in this section is to show that rich demonstrations affect context much like the speech act of assertion. I will begin by taking up rich demonstration’s informational contribution to conversation (quality *(a)*) then move to its effects on future discourse (quality *(b)*). I will conclude the section by discussing an objection from an intentionalist perspective on communication, to which speaker commitments (quality *(c)*) are relevant.

1.2.1 An Informative Point

To begin with quality *(a)*: Both rich demonstration and assertion are acts that are intentionally informative. Assertions are standardly taken to *inform*—they offer new information in a conversation that is supposed to be taken up by all of the participants—and to commit the speaker to the newly conveyed information. In the terminology of Stalnaker (1978), the *characteristic effect* of assertion is to add information to the the *common ground*: the collection of mutually accepted propositions in conversation.¹⁰ Rich demonstrations do the same thing. A preliminary observation makes this notion salient: deictic gestures can follow

⁹Theories of speech acts in this tradition build on the classic frameworks of Austin (1962) and Searle (1969) and characterize speech acts according to their ability to change the context (Gazdar 1981). What this amounts to varies between theorists (see Sbisà 2002 for discussion), but the broad perspective is that speech acts are those actions that modify a communicative context in regular ways.

¹⁰A question, by contrast, characteristically affects context differently. It might divide the common ground into distinct parts for the conversationalists to choose between (Green 2016; Roberts 2012a).

overt propositional discourse markers such as ‘because’ or ‘and’, as in (12):

- (12) I moved the table into the living room this morning and ^{NOD: scratch}

Here, a deictic gesture takes the place of a declarative clause, which would have been the vehicle for an assertion. That this is possible is a reason to believe that the gesture functions in the same way. In fact, the most natural cases of rich demonstration involve antecedent intonational cues that would normally precede a linguistic continuation.¹¹ I take these observations as preliminary evidence that rich demonstrations *slot in* to conversation like assertions by making the same kind of contribution, a propositional one.

More directly, the informative nature of rich demonstration can be seen by comparing cases of rich demonstrations with cases of assertion. Recall the tattoo case (6) and its purely linguistic version in (7), both repeated here:

- (7) a. My parents are going to be furious.
b. I got a tattoo.

In each case, the hearer comes to learn about the reason for the parents to be upset. The turn of the arm in (6) is the source of that communicated content just as the second utterance is the source in (7b). It is the reason why the listener knows what the speaker is ultimately saying. But the act performed by the gesture is more than just the causal source of the information. The act has a *function* to convey the information. That is, it is no accident that the arm turn is the source of the information—that’s why it was performed: it is a reliable guide to the information. The function of the arm turn in (6) is to inform the audience of the reason why the parents will be furious. Because of this, the speaker is

¹¹For example: *i*) the assertion in (6) would be said in exactly same way as its counterpart (7a) in the linguistic version, and *ii*) the linguistic clause in (12) would be said in the same way as the first clause of “I moved the table into the living room this morning, and scratched the wall”.

committed to specific content: that their parents will be upset as a result of the tattoo. In each case, though the speaker does not explicitly say *the reason will be the tattoo*, we take them to be intentionally providing the reason and therefore take them to be committed to that content.

That rich demonstrations are used to convey information can be seen in all of the cases above. If performance of the gesture in (6) would have failed to inform us of why the parents would get upset (for whatever reason), it likely wouldn't have taken place at all—what reason could there be to make the gesture? This observation can be made of all of the examples from Sec.1.1: If revealing the bandaged leg in (8b) couldn't get across the message that a game of squash was off the table, a more effective communicative action would be required. If drawing attention to the scratched wall in (9b) weren't enough to convey what happened while moving the table, the details of the event would have to be described in some other way. In each case, the speaker intentionally directs the attention of the hearer for some larger communicative purpose.

1.2.2 Discourse Effects

Rich demonstration is an act that, like assertion, is informative and functions to commit a speaker to the conveyed information. In addition, rich demonstration is like assertion in that they each influence the direction of conversation—they each open up and shut off possible replies. This is their shared quality (*b*). For example, a reply of “Why?” from the hearer would be perfectly acceptable before the rich demonstration of the tattoo but not after. But a more interesting followup is the following:

- (6) a. My parents are going to be ^{ARM TURN: tat} furious
b. But they have tattoos themselves!

Here, the original speaker overall conveys the same content as before: they got a new tattoo,

and that is going to make their parents furious. The contrastive marker ‘but’ signals that, for the hearer, there is conceptual conflict between the parents having tattoos and something from prior discourse. But strictly this does not conflict with the parents being upset nor with the speaker having a tattoo. Instead, it conflicts with the inferred causal connection between those facts—why would they be upset *because* of the tattoo? The followup only makes sense if we understand the original speaker to be communicating, through their packaged actions, the reason for the parents to become upset. The hearer’s objection *they have tattoos themselves* does not contrast with just any inference they’ve made, but one about what the speaker is conveying. Put more directly: the response must contrast with something the speaker *meant*. Recall (9), in which the speaker continues their message after the rich demonstration:

- (9) a. I ^{NOD: scratch} moved the table into the living room this morning.
 b. I had to buy some new paint.

Were the speaker to say the exact same thing, but without nodding their head, the followup about buying new paint wouldn’t make much sense. The speaker however knows that their mistake while moving the table has been successfully signalled, and so can say without issue that they bought paint. The act of richly demonstrating opens up the possibility of the followup in (9b).

The overall point is that in these cases, followups are licensed that can only be understood if we take seriously the communicative role of rich demonstration. The regular changes to context enforced by rich demonstrations are like as those enforced by assertions.

But the rich demonstration does not affect context in exactly the same way as does assertion. This is why rich demonstration is not the same thing as assertion. An important difference, for example, is in the at-issue status of the information conveyed. Roughly, at-issue content labels a communicative act’s “main” content (Potts 2005), while not-at-issue content is more sidelined. Appositives are a standard example of not-at-issue content:

(13) My brother, who studied film, said that the movie's great.

Here, the at-issue content is that my brother said that the movie is great, while the not-at-issue content is that my brother studied film. How to theoretically characterize and diagnose the at-issue/not-at-issue distinction is controversial.¹² But a standard test is direct rejection: roughly, if “That’s false” can target some propositional content, then that content is at-issue.¹³ But following the rich demonstration in (6) with “That’s false” would seem to target the parents getting upset, not the reason that was offered as to why. That the cause of their being upset cannot be targeted by rejection is evidence that it is not-at-issue. The information conveyed by a rich demonstration, then, is different in this respect from information that is directly asserted. But this does not undermine rich demonstration’s status as fully communicative any more than it undermines an appositive’s status as such.

1.2.3 Goats and Herod’s Intentions

I have argued that rich demonstration is a kind of speech act. The reason to think this is that it shares core features with assertion, in particular that each they function to coordinate agents on what information they treat as true. I now want to consider an objection one might have to the approach I’ve presented. A skeptic might think hold that rich demonstration is not a communicative phenomenon. Instead, they might hold that the explanation behind these cases is simply that agents reason about scenes in the conversational vicinity—something that clearly is not a speech act. The objection comes from an intentionalist perspective inspired by Grice (1957).¹⁴ From such a perspective, communicative acts are those that are performed with the right kind of intention (typically directed at the hearer).

¹²See Anderbois et al. (2013), Murray (2014), Simons et al. (2010), Snider (2017), Syrett and Koev (2014), and Tonhauser (2012).

¹³Saying “That’s false” in response to (13) would constitute a denial of what my brother said about the movie, not a denial of whether he studied film.

¹⁴See also Bach and Harnish (1979) and Cohen and Perrault (1979).

The reason, Grice thought, why Herod could not communicate that John the Baptist was dead by presenting his remains was that the right intention could not be formed. The indicated scene made the Saint's death obvious, and so Herod could not intend for Salome to come to believe *John the Baptist is dead* on the basis of the mere recognition of his action. One might claim that all cases of rich demonstration fit this pattern: the scenes indicated make the relevant information obvious, and so in performing a rich demonstration, one cannot form the appropriate communicative intention. So it would not matter that rich demonstrations share core features with assertion, as I have argued. They lack the intentionalist grounding that makes assertion communicative; what I have called *rich demonstration* is not communicative, but rather just a way of helping others reason about the world.

The objection can be made sharper. One who takes this position might offer a story about how it is that “rich demonstrations” are useful without being properly communicative. Stalnaker (1978, 2002, 2014) discusses *manifest events*—events surrounding a conversation that are so obvious that we cannot help but come to mutually accept them. To borrow one of his examples: if while you and I were talking, a goat were to suddenly walk into the room, we would both recognize it and recognize that the other recognized it. Our common ground of mutually accepted propositions would therefore be updated to include the content that *a goat just walked in*. Thanks to this, we could even perform speech acts that presuppose the goat's presence (“How did that thing get in here?”). Despite the informative nature of the goat's walking in (or of our perceiving the goat), nobody would claim that it is like an assertion or any other kind of speech act. Instead, the theoretical apparatus of the *common ground* helps to explain how we can come to mutually accept new facts despite their not being communicated. From the perspective of this extension of the intentionalist line, rich demonstrations are not communicative, but they can still be the causal source of our mutually accepting something nonetheless.

Two mistakes fuel this objection. First: the information conveyed by a rich demonstration is far more complex than what is perceptually available. The way in which rich

demonstration exhibits quality (*a*) is not by informing of something perceptually obvious. When one richly demonstrates, they do not simply show that the indicated scene obtains, but rather they indicate the scene for some discursive end. The tattoo is shown to give the reason why the parents will be upset—a fact that is not available to be observed. The bandaged leg is revealed for a reason: to respond to an invitation. The scratch on the wall and the cookies on the table are pointed out not just to show that they exist, but to give more information about a series of events earlier in the day. In each case, conceptual connections are conveyed, not just bare facts about the world. *X is the cause of Y*, or *X happened during Z*, for example. These connections are not just available for inference on the basis of observation. The inferences must be made on the basis of what the speaker is communicating as a whole. Because these contents are not perceptually available, it is easy to form the appropriate communicative intentions. For example, the speaker could intend that the hearer recognize the explanatory fact (*this is why*) on the basis of recognizing their very communicative intention. So the intentionalist's strict necessary (and, on some versions, sufficient) condition for an act's being communicative is actually satisfied. Furthermore, this complicates the Stalnakerian story of how agents coordinate their attitudes. What they coordinate on is no longer merely what is plainly in front of them, but an abstract conceptual connection between two facts.

The second mistake is even more serious. Even if the intentionalist can make sense of how rich demonstrations are informative without being communicative (which I've just argued they cannot do), they still fail to account for the commitments a speaker takes on in virtue of performing a rich demonstration. Imagine the speaker only said "my parents are going to be furious", but did not followup with an utterance or gesture of any kind. If in that case the hearer happened to notice the tattoo on the speaker's arm, they might correctly guess that it would be the reason behind their eventual fury (provided the hearer had enough relevant background knowledge). If the objection were well-founded, then there would be no significant difference between this and my central case of rich demonstration. This is

because the core phenomenon in both cases would be the same: the hearer reasons about their surroundings. But there is a significant difference between the cases. In the case of mere noticing and inferring, the speaker has not publicly committed themselves to the fact that their parents will be upset *because of the tattoo*. After all, they could say “I bought a motorcycle” after the hearer noticed the tattoo, undoing the hearer’s hard inferential work. But the rich demonstration precludes such a possibility. The speaker commits themselves to the proposition that their parents will be upset because of the tattoo. In this respect, quality (*c*) mentioned above, rich demonstration is again strikingly similar to assertion. Not only does the rich demonstration convey complex (perceptually unavailable) content, but it brings with it *a commitment* to that content’s truth. This is the key difference between rich demonstration and mere goat-like manifest events.

It is because of this that I take Herod’s revelation to be genuinely communicative (a rich demonstration). Herod did not want Salome to learn of John the Baptist’s death just because. He had a reason for drawing her attention to that fact. Salome had demanded that her stepfather have the Saint killed, and he had promised her John the Baptist’s head. In revealing the charger, he intentionally informed her that he had followed through, and her demand had been met—something that could not be perceptually obvious. Not only was this the information he conveyed, but he became committed to it. By showing the head, Herod became publicly committed to the proposition that he had satisfied his promise. He successfully communicated this thanks to a rich demonstration. The presentation drew attention to the scene before them, and it functioned to take advantage of it toward a larger communicative goal: to elaborate on the status of the salient promise between them. I now turn to explaining how this is possible.

1.3 Enriching Demonstrations

That rich demonstration is a kind of speech act adds further evidence to the growing literature regarding the meaningful impact gesture has in conversation. My position is consistent with general research into gesture (Kendon 2004, 2017; McNeill 1992, 2005) and its formal semantics/pragmatics (Ebert and Ebert 2014; Ebert et al. 2020; Lascarides and Stone 2009a,b; Schlenker 2018; Stone and Stojnić 2015). But these conclusions go further. Not only do gestures combine with speech to form meanings as described by many contemporary theories, but rich demonstrations independently, qua speech acts, contribute their own meanings in conversation. The challenge, then, is explaining how this occurs. How does a gesture indicate something from the context, but result in the contribution of content that is propositional? More concretely, the challenge is in explaining where the communicated content comes from. The proposition that the speaker’s parents in (6) will be upset because of a tattoo isn’t “out in the world” waiting to be pointed to. The tattoo is.

The purpose of this section is to provide an explanation of how a simple act of directing attention can become a full assertion-like speech act. On my view, the mechanisms behind the inference of causation in (7) (the case’s linguistic version) are the very ones behind the interpretation of (6) (the rich demonstration), with a few important differences. In each case, the second action (the utterance or gesture) is intuitively self-contained in terms of its communicative function. Despite this, the content of the communication (what’s learned as a result of the move) conceptually relates to what was communicated previously in a regular, understandable way: by *explaining* it.

The analysis I offer has two parts. First, I suggest a broader notion of what is indicated by a deictic gesture than what is assumed in the referentialist tradition. On my view, the deictic gestures from which rich demonstrations arise indicate *situations*—roughly, perceptually available scenes. Second, interpretive conventions of coherence (Asher and Lascarides 2003; Hobbs 1979; Kehler 2002) provide a interpretive framework for generating the right

propositional inferences involving indicated situations. The analysis, then, is that a rich demonstration is performed when the conventions of coherence enforce a conceptual connection between an indicated situation and the content of a prior communicative move. The account's explanation of the phenomenon hinges on the detailed nature of contextual situations and on the limited range of propositions that can be inferred.

1.3.1 Indicating Situations

There are two reasons I wish to highlight for thinking that rich demonstrations semantically interact with situations, instead of merely with objects. First: complex networks of properties are just as relevant as particular objects for understanding the discourse contribution of a rich demonstration. In this respect, I follow Hunter et al. (2018) in emphasizing the importance of extra-linguistic situations in discourse interpretation. Consider (14), where the speaker sweeps their hand, indicating an entire room littered with empty beer bottles, food remains, some broken picture frames, etc.:

(14) My parents $\frac{\text{SWEEP: mess}}{\text{are going to kill me.}}$

It is the whole scene, straight out of a 90's movie, that the speaker utilizes in their message. No particular element of the scene alone accounts for why the speaker is worried, but rather everything together. That is, in order to glean the right message, we must first understand the speaker as indicating the right thing. That thing is a complex scene, composed of many different details, each of which contributes to but does not alone suffice in guiding the interpretation. The information contained in the situation as a whole is crucial for generating the appropriate interpretation of the rich demonstration. Note that this kind of point is not unique to rich demonstrations. That situations allow for informational interactions of the kind needed here is a point central to situation semantics (e.g. Barwise and Perry 1983). These cases here further illustrate, though, the ways in which situations can become

integrated into conversation.¹⁵

The kind of situations I have in mind follow an ontology in the style of Kratzer (1989, 2002), who treats situations as particulars with mereological structure that are the building blocks of worlds. Any two distinct worlds will differ in at least one of their component situations: they each will contain at least one situation not contained in the other. These situations roughly function as scenes or areas in a world. There will be, for example, the situation of the dog barking outside my apartment in some worlds but not others.

Situations happen to correspond to propositions—all worlds containing the situation where the dog is barking will be part of the set that is the content of “the dog is barking”. But not all propositions correspond to situations. The worlds making up the propositions [[The dog is barking or not]] and [[The earthquake explains the vase’s falling]] will not have any particular unifying situation. In the case of the former proposition, situations are not disjunctive. In the case of the latter, the abstract *explanation* fact is not the kind of thing characterized by a situation. This is a very rough gloss, but the intention is to enable a mechanism for interacting with parts of worlds (indicated by deixis) that has a separate function from the informational one of propositions.¹⁶ In a normal case, a gesture will indicate a situation that has many sub-situations composing it. These sub-situations (and their sub-situations) provide the detail needed to account for the networks of properties that are relevant for interpreting discourses like (14).

The second reason to think that rich demonstrations indicate situations involves their flexibility of interpretation. I earlier emphasized the similarities between cases of rich demonstration and their purely linguistic counterparts. But their differences are just as important to my analysis. Consider again (6) and (7), this time adding the contextual background information that the tattoo depicts a pentagram:

¹⁵This is related to a point made by Stojnić et al. (2013), who present a situation-based analyses of indexical reference resolution. While I focus on how an indicated situation can semantically impact a discourse, they focus on how a discourse can incorporate a situation without explicit indication via a gesture.

¹⁶For a complete treatment of situation semantics, see Devlin (2006).

- (7) a. My parents are going to be furious.
b. I got a tattoo.

With this added information, there are now two possible precisifications of the message communicated in (6). One is the interpretation I have been assuming so far: the parents would be upset by the speaker's getting any tattoo. On the newly available interpretation, the speaker's parents might have been fine with other tattoos; it is the image depicted by this tattoo that will upset them. But the added fact about the tattoo's content does not similarly change (7). Without showing the tattoo, this second interpretation is unavailable. Even if an interpretation on which the tattoo's content (whatever it is) is the culprit, that is distinct from the interpretation on which *this specific depiction* is to blame.

I suggest that this complexity is owed to the fact that these deictic gestures indicate situations. The speaker in (6) communicates that the indicated tattoo, with all its detail, location on the body, quality, and so on, will for some reason or other upset their parents. The richness of the indicated situation allows for many different justifications for this, which may of course be explored in conversation ("Because it's a pentagram or because the lines are jagged?").

In the case of the tattoo, most of the reasonable interpretations involve possible properties of the tattoo, so while situations may offer a compelling explanation of the flexibility, the case may not decisively favor situation-indication over object-indication. But the same kind of consideration in other cases does favor indicating situations:

- (11) a. *C*: NOD: pool
b. *M*: You would get such a bad sunburn.

In order for Carlos to reasonably raise the possibility of regularly going to the pool, we must take the gesture to indicate more than just the pool itself; more should go into the scene.

For example: that the pool is not closed off, that it belongs to this apartment complex and not the neighboring one, that there are plenty of lounge chairs around, and so on are all situational facts that might go into why Maria interprets the gesture as she did. If the scene involving the same pool were differently detailed—say, if there were trash laying all over the ground—we would take Carlos to be communicating something very different.

In this next case, imagine the speaker is talking to their partner, who just got home, about their child. The speaker motions toward an area with a broken vase and a football nearby:

(15) We're going to $\frac{\text{POINT: vase+ball}}{\text{have to ground her again.}}$

Here, neither the broken vase nor the football alone is sufficient to get across the message that the child broke the vase with the football and therefore must be grounded. The intuition is that the speaker is picking out the scene of the broken vase and nearby football, and that they do so with communicative intent. Analyzing rich demonstrations as indicating whole situations is the natural way to account for this.¹⁷

In arguing that situations are indicated by the gestures that perform rich demonstrations, I do not want to claim that situations from context are antecedently elements of the discourse. This is roughly the approach that I take Hunter et al. (2018) to endorse. They explicitly discuss cases that I consider rich demonstrations, and like me (Sec.1.3.2) present a coherence-based explanation of how rich demonstrations are interpreted. Our analyses differ, however, with respect to what kinds of things can introduce content into a discourse. Our philosophical conclusions differ about what rich demonstration's status as a speech act means for the nature of linguistic representation. Hunter et al. endorse a claim that they call the *nonlinguistic*

¹⁷The approach outlined here does not encounter difficulties often associated with situation semantics. The criticisms from Soames (1985) of the semantic framework offered by Barwise and Perry (1983), for example, only takes hold if a possible worlds semantics is foregone. The position I offer here allows for possible worlds and situations to coexist.

discourse unit hypothesis, according to which worldly events can contribute discourse content to a conversation:

Nonlinguistic Discourse Unit Hypothesis (NDU)

A nonlinguistic event e can affect the interpretation of a discourse by contributing the content of an entire discourse unit (i.e. an instance of a proposition), which enters into rhetorical relations with other, linguistically-specified discourse units. In so doing, e changes the very structure or logical form of the discourse, and it can do this without any explicit expression signalling its relevance. Rather, its relevance is inferred through the kind of reasoning used to infer rhetorical connections between linguistically expressed contents. (pp. 5–6)

I strictly maintain that speech acts are the building blocks of conversation—information from the world can be integrated into discourse only through this one path. This is the basis for my argument regarding informative function in Sec.1.2.1. In contrast, they take on the project of assigning semantic content to worldly events and integrating that content directly into discourse, without any act to necessarily introduce them. They say:

In our view, semantic structures composed entirely of what are traditionally classified as discourse moves (including, perhaps, discourse dependent nonlinguistic moves) are just a subclass of the kinds of structures that we can use such moves to build. In fact, we think that the kinds of semantic structures built up from coherence relations need not involve any discourse moves at all. Suppose Peter looks out into the garden and sees his cat, Lupin, staring at a pile of leaves. The leaves suddenly move, and Lupin pounces. Peter goes to investigate and finds a baby whip snake. He now understands why Lupin was staring at the leaves and why the leaves rustled; he also understands that Lupin’s pounce was a result of the leaf movement. Yet, neither the snake nor the cat intended to communicate anything, and certainly the snake didn’t intend its presence to explain Lupin’s

behaviour and Lupin didn't intend his pounce to be a result of the leaf movement. Nevertheless, both the result and meta-level explanation are inferred. (pp. 18–19)

The analysis that Hunter et al. present is designed to handle conversations that take place during a *Settlers of Catan* game. Their system reflects, for example, when a player says something as a result of something that happened in the game.

I reject NDU, and do not believe it is needed to accommodate the data that Hunter et al. focus on. The fact that complex inferences (about e.g. causal relationships) are regularly drawn in everyday life need not be explained by the explanations behind discourse interpretation. It seems to me to be an empirical question whether a causal inference from perception has the same psychological basis as the causal inference made when one is told two facts. Depending on how one conceives of coherence relations, it may be impossible for them to be the kinds of inferences available outside of communication, as is required by NDU.¹⁸ Some theorists view coherence relations as social *conventions*.¹⁹ Others view coherence relations as patterns of abduction for understanding intended messages.²⁰ In each case, it would seem highly implausible that perceptual inferences are underwritten by the rules of coherence.²¹

I agree that non-linguistic events can become semantically relevant to discourse, but disagree with the claim that nothing must be done to enforce an event's relevance. Instead, some kind of conversational move must be performed. For cases in which there is no overt assertion or gesture to draw attention to an event, I contend that there is something that makes the non-linguistic events conversationally salient. In many cases, this will be thanks to the coordinated/mutual attention of the conversation's participants. Acts of gaze-following,

¹⁸See Asher and Lascarides (2003, Ch. 3.3-3.6) for a discussion of some different approaches to coherence.

¹⁹For example: Cumming et al. (2017), Lepore and Stone (2015), Stojnić (2018), and Stojnić et al. (2017).

²⁰For example: Hobbs (1979), Hobbs et al. (1993), and Kehler (2002).

²¹I'm grateful to an anonymous reviewer for helping to clarify the logical space of how my analysis might relate to NDU.

for example, can delineate between communicatively relevant and irrelevant surroundings.

My priority is on modeling how communication happens. I take cases of rich demonstration to show not that events or situations can form discourse structures, but that the communicative actions that indicate them (gestures) can. Without people to have conversations, cats would still pounce on snakes, but discourse structures would be no more. We do not consider the height of the Empire State Building to be an element of discourse structures simply because we can discuss it or because its actual height impacts what we say. Similarly, we should not consider worldly situations to be discourse elements just because we can use them effectively in communication.

1.3.2 Forming a Coherent Message

Taking stock: I have argued that rich demonstration is a speech act that shares core features with assertion. It arises from the performance of a gesture, and that gesture indicates a situation from a conversation's context. Though the gestures that trigger them merely indicate situations, rich demonstrations inform the audience of propositions that are not obviously true. How does drawing attention to a scene result in such complex communication? The final ingredient to my analysis bridges the gap: conventions of establishing coherence in conversation meaningfully connect indicated situations to previously expressed content.

My analysis is based in discourse coherence theory (Asher and Lascarides 2003; Hobbs 1979; Hobbs et al. 1993; Kehler 2002). That is, my approach assumes that there are background conventions of interpretation that conceptually relate contents to one another to generate further (implicit) content. I contend that the range of conceptual relations rich demonstrations can convey is quite narrow. Typically, a rich demonstration *explains* or *elaborates on* something from preceding conversation. In essence, the picture is that some fact or situation will be salient in conversation, and rich demonstrations say *this situation explains why* or *this situation adds more detail*. Because the range of possible connections is so limited by convention, no complex pragmatic derivation is required to interpret a rich

demonstration. The complexity offered by indicated situations combined with the simplicity of inferential pattern explain how we arrive at the interpretations we do. I'll start by describing the coherence approach generally, then move to my specific proposal regarding rich demonstration.

1.3.2.1 The Framework

Coherence theory investigates the conventions for inferring meaningful connections between discourse segments. A classic example from Hobbs (1979) and a variant on it from Kehler (2002) are particularly illustrative of the kind of connection being studied:

(16) John took a train from Paris to Istanbul. ??He likes Spinach.

(17) John took a train from Paris to Istanbul. He has family there.

The unstated, but certainly communicated, content in (17) is that John went to Istanbul to visit his family. Though (16) is questionable, notice that we find ourselves looking for a way for the second sentence to justify the first, exactly as in (17). The reason for this, from the perspective of coherence theory, is that both cases are structured identically—according to an *Explanation* relation—but only the latter pragmatically satisfies the semantic requirements that relation imposes. *Explanation* is one of many relations that structure discourses.²² Normal relations are *Explanation*, *Narration*, *Contrast*, *Elaboration*, and so on. I follow Asher and Lascarides (2003) in conceiving of coherence relations as *relational speech acts*—communicative acts that relate contents from discourse. Each of these imposes additional content onto a conversation, above and beyond what's expressed by the explicitly performed speech acts. For example, *Narration* relating two propositions would require that the second

²²The number varies depending on the flavor of coherence theory. Hobbs (1979) and Kehler (2002) posit only around ten, Rhetorical Structure Theory (RST, Mann and Thompson 1988) utilizes around two dozen, and Segmented Discourse Representation Theory (SDRT, Asher and Lascarides 2003) defines over thirty. For simplicity I will mostly use the relations discussed by Kehler (2002), but nothing hinges on that.

describes something that happened after the first: “Jorge set the book down. He went to make some food.” conveys that Jorge set the book down *before* making the food. The grammatical features of the segments of a conversation determine which coherence relations are operative, where the category of *grammatical features* is broadly construed to include lexical, syntactic, morphological, and information-structural features.

Coherence theory has traditionally been utilized only to understand the structure of discourses whose parts are purely linguistic, but it can naturally be extended. My approach follows theorists such as Hunter (2018), Lascarides and Stone (2009a,b), and Stone and Stojnić (2015) in conceiving of gestures as the kinds of things that cohere with other discourse moves. They consider iconic, not deictic, gestures, but the analyses are complimentary. Put simply, my view is this: just as chains of linguistic segments can be used to form coherent messages, so too can chains involving both linguistic segments and gestures. A rich demonstration is a speech act that relates a proposition from earlier discourse to one that corresponds to a deictically indicated situation.

1.3.2.2 The Conventions

I would now like to show that rich demonstrations cannot be used to communicate just anything. Instead, the range of possible expressions is surprisingly limited and seemingly arbitrary. This is evidence that there are real conventions of coherence at work. Rich demonstrations, I claim, by default must stand in a *subordinating* coherence relation to prior discourse, with a strong bias for *Explanation* or *Elaboration*, the core semantic subordinating relations.

In conversation, one can assert that X explains or elaborates on Y , and one can do so by richly demonstrating. Importantly, I only claim that this restriction holds for how rich demonstrations relate to *preceding* discourse. As I will show, the potential followups to a rich demonstration seem unconstrained.

Though the term lacks a precise widely agreed-on definition, *subordinating* coherence relations are those that are present in order to fulfill an already established goal in the discourse. They contrast with coordinating relations, which introduce new topics and goals.²³ Among the standardly recognized subordinating relations are *Explanation* and *Elaboration*. Intuitively, when I communicate so as to explain something previously mentioned, I do not shift the topic of conversation, but pursue it. Similarly, when I elaborate on something previously mentioned, I do not change, but maintain the same narrative path. Standard coordinating relations are *Parallel* and *Contrast*. Drawing a parallelism in conversation shifts it to something similar, but new. Contrasting two situations or topics turns the focus around to something opposite in theme. When the speaker of (6) turns their arm to show the tattoo, they do so in order to explain why their parents are going to be furious. When the speaker of (10) nods to the cookies on the table, they elaborate on what happened while going shopping.

These are very natural cases, and illustrate that rich demonstrations are at least capable of standing in subordinating relations to preceding moves. To argue that there is a bias for subordination, I will consider cases in which (i) a deictic gesture follows an assertion, (ii) pragmatic reasoning requires that the the gesture relate to the assertion by a coordinating relation, but (iii) the assertion-gesture pair is still questionable. The reason why the cases are questionable is that subordinating relations are preferred, even at the expense of pragmatic felicity.

I begin with a case that naturally should admit of a *Contrast* reading (*Contrast* is coordinating (Kehler 2002)). Consider the following in a context where the speaker and hearer are visiting New York and walking through a busy crowd. In each, the speaker motions with an open hand toward a person ahead who's walking slowly.

²³This is the most natural way of introducing the distinction between subordination and coordination, in terms of the goals and topics in discourse (Asher 1993; van Kuppevelt 1995). But see Asher and Vieu (2005) for a discussion of complications.

(18) *Contrast?*

??I always heard that people in New York walk fast ^{OPEN HAND: slow}

(19) *Contrast*

^{OPEN HAND: slow} I always heard that people in New York walk fast

Upon reflection, the leading gesture is far preferable to the trailing gesture. This is surprising if rich demonstrations are not sensitive to subordinating/ordinating discourse structure. *Contrast* imposes a semantic requirement that is symmetric: if a situation or fact *a* conceptually conflicts with another *b*, then obviously *b* conceptually conflicts with *a*. This kind of conflict is pragmatically allowed in the above cases, but only naturally arises in (19). I contend that this is preliminary evidence that rich demonstrations are sensitive to subordination/ordination, at least when relating to prior discourse. Note that (18) becomes acceptable with the addition of an explicit contrast marker.

(20) I always heard that people in New York walk fast, but ^{OPEN HAND: slow}

Without strictly enforcing a structure with *Contrast* operative, the discourses naturally sound odd, even though a contrasting relation is readily pragmatically available (if it were not, even (20) would be infelicitous). It seems that the default interpretation is an *Explanation* reading, even at the cost of acceptability. Importantly, my claim about the subordination bias only extends to the relation between rich demonstrations and prior, not posterior discourse.

The bias for subordination to prior discourse is even more evident with respect to *Result*. *Result* imposes the same semantic constraints on a discourse as *Explanation*; they only differ with respect to argument order and subordination/ordination status. That is, *Explanation*(*X*, *Y*) will impose the same contentful restriction on a discourse as the structure *Result*(*Y*, *X*). Roughly, this will be that *X* causes *Y*. Given that the two relations

impose the same contentful restrictions, we should (ignoring subordination/coordination) expect pairs of cases whose only difference is in segment ordering to be acceptable. But this is not what we find. Instead, only *Explanation* cases are acceptable.

(21) *Result?*
 OPEN HAND: vase
 ??Leila threw a football

(22) *Explanation*
 OPEN HAND: vase
 Leila threw a football

The distinctions between these cases are subtle, but instructive. In both cases, only one causal chain makes sense: that the football-throwing caused the broken vase (not the other way around). To force this strongly preferred interpretation, the discourse moves in (21) should be related by *Result* and the moves in (22) should be related by *Explanation*. Neither case has any obvious interfering factors to coerce one coherence relation or the other. Nevertheless, (22) is preferred. This is because of the subordinating bias brought by the rich demonstration.²⁴ Note that the trailing gesture becomes more natural with the addition of ‘and’, which requires a coordinating relation (Txurruka 2003), overriding the default bias of the rich demonstration.

(23) *Result*
 OPEN HAND: vase
 ??Leila threw a football, and

This is the same phenomenon as in (20): a discourse marker forces a coordinating reading that would otherwise be unavailable. There are also variants wherein the grammar of the

²⁴This hypothesis only bears on the infelicity of (21). I think the reasons for the *Explanation* reading in (22) are not necessarily owed to conventions about rich demonstration, but to other structural features of the discourse and to the strong pragmatic preference.

leading sentence is changed so that the case is more natural, as in (24). But this does not mean that they allow for the coordinating *Result* relation.

- (24) *Elaboration*
??Leila was throwing a football around OPEN HAND: vase

Importantly, (24) is a case of *Elaboration*, not *Result*. The football-throwing situation is not the same in (24) as it is in the other cases, and this is due to the change from the stative to progressive aspect. To put it simply, the situation described in the cases above involve one throwing event, but the situation described in (24) involves many throwing events. Expanding the situation to Layla's progressive, not-one-off activity means that the broken vase situation can be permissibly contained within it. This is why, for example, we can naturally say (25):

- (25) Layla broke the vase while throwing a football around.

That situations can be conceptually nested in this way allows for *Elaboration* to be operative while *Result* is blocked. The latter coherence relation is ideal for two separate situations, and the former is ideal for situations that stand in a part-whole relation.

These cases illustrate rich demonstration's sensitivity to established conventions of discourse coherence. I have argued that they, absent overriding factors, must attach to prior discourse by a subordinating relation. *Explanation* and *Elaboration* are strongly preferred because the gestures that generate rich demonstrations do not supply propositions. Instead, *Explanation* and *Elaboration* can fill in the necessary content.

1.3.3 Building a Speech Act

To emerge from the details and return to the big picture: I've argued for an analysis that explains how an act of drawing attention to a scene in the vicinity can count as a speech act. Deictic gestures are, as always, ways of drawing attention to something in the vicinity. But they indicate whole situations from a conversation's context. The act of doing so can constitute a speech act thanks to conventions of coherence that establish means of connecting contents. These rules restrict the kinds of contents a rich demonstration can express: explanations or elaborations. Altogether, a rich demonstration is the connection of an indicated scene to a previously expressed proposition according to the narrow conventions of coherence. I will return to the tattoo case one last time to show how this manifests:

I'll refer to the proposition expressed by the linguistic assertion as ' P ' and to the situation indicated by the deictic gesture as ' T '. P is the set of worlds where the speaker's parents are furious. T is the tattoo-situation: the scene revealed by the turn of the arm. The rich demonstration is the speech act of relating P to the proposition that supports T (i.e. that conjunctively describes all of the information of the scene)—call it ' Q '. The rules of coherence that I have posited say that Q relates to P according to either *Elaboration* or *Explanation*. Pragmatically, *Explanation* is preferred (i.e. *Elaboration* doesn't make much sense).²⁵ *Explanation*(Q, A) requires the inference that something about Q causes A : something about the speaker's having their tattoo must stand in a causal relation to the parent's upset attitude. This is exactly the interpretation we naturally arrive at, and the one that would be targeted by the potential followup "But they have tattoos themselves!".

²⁵I leave open the possibility (and suspect it is true) that the *Explanation* interpretation is preferred to *Elaboration* on the basis of morpho-syntactic features of the asserted sentence, not just pragmatically. Ascertaining this would require in-depth empirical research.

1.4 Conclusion

Rich demonstrations, I have argued, are deployments of deixis that result in speechless speech acts. They share core features with assertion. This fact bolsters their status as speech acts but must be explained. I have offered an account of how it is possible that fully communicative speech acts can arise from directions of attention. Deixis is capable of indicating entire situations from context. Those situations are utilized in narrow interpretive patterns to cohere with surrounding conversation. Thanks to the flexibility offered by situations and the conventions of interpretation, directing attention can communicate.

CHAPTER 2

Pointing to Communicate

I argued in the previous chapter that rich demonstrations such as (26) are genuine speech acts, and argued for an approach to understanding their discourse contribution on which deictic gestures can indicate situations, and such indication can bear coherence relations to surrounding speech acts.

(26) My parents are going to be ^{ARM TURN: tat} furious

In this chapter I present and argue for the details of this approach. In virtue of being a speech act, rich demonstration counts as a complete *discourse move* in the sense of the tradition in dynamic semantics and pragmatics. Accordingly, in this chapter I provide a formal treatment of rich demonstration detailing how it dynamically affects a discourse's information state as well as how its content is inferred. The analysis I provide makes use of an update semantics in which situations make up contexts (Kratzer 1989, 2002) and coherently relate to surrounding discourse in predictable ways (Asher and Lascarides 2003; Hobbs 1979; Kehler 2002).

The resulting picture builds on ongoing research in the super linguistics of gesture. It adds a new kind of evidence for the now classic claims of Kendon (2004) and McNeill (1992), who emphasize the various deeply important roles gesture has in communication. It is designed to capture a gesture's discourse effects, instead of more local sentence-modifying effects (Ebert and Ebert 2014; Ebert et al. 2020; Schlenker 2018; Schlenker and Chemla 2018). It also directly extends the theoretical claims of Hunter (2018) and Lascarides and Stone (2009a,b),

who argue that iconic gestures are interpreted according to coherence-theoretic mechanisms. On my view, the same can be said of deictic gestures.

I will begin in Sec.2.1 by discussing the role of situations on my analysis and providing a simple definition of a model. I will then in Sec.2.2 introduce a formal representation of a conversation state, which can represent the semantic impact that coherence relations have when they hold between multiple discourse moves. These two ingredients together are what enable performance of a rich demonstration. In Sec.2.3 I present more complex empirical data and discuss directions for future research.

2.1 Extracting Situations from Context

2.1.1 Indicating Situations

In the previous chapter I argued that the vehicles of rich demonstration, deictic gestures, are helpfully analyzed as indicating situations.¹ Recall the observation that rich demonstrations lend themselves to a kind of flexibility of interpretation. In a context where the speaker has a pentagram tattooed on their arm, (26) makes available two different explanations of why the parents will be upset.

On one, it is the fact that the speaker got a tattoo that is upsetting. On another, it is the fact that the speaker got *a pentagram* tattoo that is upsetting. One might even imagine other reasons, for example that the tattoo is low quality. The upshot is that the rich demonstration does not isolate any of these possibilities, but rather communicates the general thought that the tattoo, for some reason or other, will be upsetting. This is easily

¹Strictly, deictic gestures need not be the only vehicles of rich demonstration. In my view, the proposal put forward here and in Chapter 1 could apply to anything that would count as a *direction of attention*. I focus on deictic gestures as a starting point because this is where the phenomenon most naturally occurs, but one can imagine cases of non-gestural deixis that behave similarly. If instead of turning their arm the speaker in (26) showed a sign that read “I got a tattoo” or even said “look at my arm”, I think a rich demonstration would have been performed. I’m grateful to Philippe Schlenker for raising questions about this and for the former example, and to an anonymous reviewer for raising the same point.

captured by conceiving of the rich demonstration’s discourse contribution as relating the scene of the tattoo, including all of its details, to the proposition that the speaker’s parents will be upset.

The details relevant for this chapter can be fleshed out formally, to a first approximation, with the following definition of a model:

Model (v1) = $\langle \mathcal{S}, \mathcal{W}, \mathcal{L}, \llbracket \cdot \rrbracket \rangle$

A model is a tuple of the set of all possible situations $\mathcal{S} = \{s_1 \dots\}$, the set of all possible worlds $\mathcal{W} = \{w_1 \dots\}$, the set of well-formed meaningful linguistic segments $\mathcal{L} = \{l_1 \dots\}$, and an interpretation function $\llbracket \cdot \rrbracket$ such that:

- Situations have mereological structure. They can form part-whole relations (\leq) with respect to one another such that:
 - \leq is reflexive: $\forall s \in \mathcal{S} : s \leq s$
 - \leq is transitive: $\forall s, s', s'' \in \mathcal{S}$ if $s \leq s'$ and $s' \leq s''$ then $s \leq s''$
- Worlds are maximal situations: $\mathcal{W} = \{s \in \mathcal{S} : \neg \exists s' \neq s \in \mathcal{S} : s \leq s'\}$
- $\llbracket \cdot \rrbracket$ is a function from phrases to propositions: $\llbracket \cdot \rrbracket : \mathcal{L} \rightarrow \mathcal{P}(\mathcal{W})$

My intent is to elucidate the interaction between rich demonstrations and other communication at the level of discourse. So for simplicity, I set aside the compositional semantics of natural language and consider entire linguistic segments,² which are assigned content—propositions modeled as sets of worlds—by the interpretation function.

This system, in roughly the style of Kratzer (1989, 2002), treats situations as particulars that make up worlds. For any two worlds w_1, w_2 , therefore, if $w_1 \neq w_2$ then they differ in at least one of their component situations: $\exists s : s \in w_1 \wedge s \notin w_2$. These situations roughly

²“Segments” because \mathcal{L} should include anything that can function as an assertion, such as a meaningful sentence fragment.

function as scenes or areas in a world. There will be, for example, the situation of the dog barking outside my apartment in some worlds but not others.

Situations happen to correspond to propositions—all worlds containing the situation where the dog is barking will be part of the propositional set that is the content of “the dog is barking”. But not all propositions correspond to situations. The worlds making up the propositions [The dog is barking or not] and [The earthquake explains the vase’s falling] will not have any particular unifying situation. In the case of the former proposition, situations are not disjunctive. In the case of the latter, the abstract *explanation* fact is not the kind of thing characterized by a situation. This is a very rough gloss, but the intention is to enable a mechanism for interacting with parts of worlds (demonstrated by gestures) that has a separate function from the informational one of propositions.³ This is the mechanism that will ultimately enable the reconstruction of the assertion-like effect of rich demonstration.

2.1.2 Indexicality

The arguments from Chapter 1 begin to show why rich demonstrations should be considered properly indexical. The reason is not complicated—the content and effect of a rich demonstration are crucially impacted by whatever from the surrounding context the gesture indicates. Very different discourses arise from (27), for example, if the speaker holds up a broken pair of glasses vs. a broken smartphone:

(27) $\frac{\text{HOLD: } x}{\text{Guess I'll be spending more money than I budgeted for this week.}}$

The speaker communicates in each variant what they’ll be spending money on. This crucially depends on what it is they hold up. That is, the interpretation of the entire message requires an element from context that is indicated by the gesture. We can note more intricate cases of indexicality, though. Recall (10), repeated here:

³For a complete treatment of situation semantics, see Devlin (2006).

- (10) a. I went to the ^{POINT: cookies}grocery store earlier.
 b. . . .if you want any.

As noted before, here the deictic gesture fills two roles, one for communicating content (the rich demonstration), and the more classical one for restricting the domain of ‘any’ (perhaps involving ellipsis). It might be claimed that this is a more-or-less classical case of indexicality and nothing more. The idea would be that the gesture functions to restrict ‘any’, and it is pragmatically inferred on the basis of the offer of cookies that they were bought at the store. So the inference is generated in the same way as (28):

- (28) I went to the grocery store earlier, if you want any cookies.

But the explanation of the inference in (28), whatever it turns out to be, cannot work in both cases. To see this, consider a variant of the message in (10) without the trailing conditional offer:

- (29) I went to the ^{POINT: cookies}grocery store earlier.

The same inference is generated, suggesting that its causal source is the gesture, since in (29) there’s no indexical phrase to plausibly account for the interpretation.

My formal treatment of rich demonstrations utilizes evolving conversation states, or *discourse records* (Lewis 1979; Thomason 1990a). Discourse records are labeled ‘ $K_1 \dots K_n$ ’ and track the various information types in discourse. I will notate a rich demonstration made with respect to a conversation state K as $D(\delta)^K$, where δ is an indexical gesture that indicates some situation in the current context, and ‘ D ’ designates the deixis as a rich demonstration. Included in a conversation state K is its context c , which is the situation in which the conversation takes place. It will have numerous other situations as parts. It will include situations detailing classical elements of context such as the speaker and location.

It will additionally include various parts of the surrounding physical context available for indication, such as the adorable Chow Chow that walks by as we chat in the park.

To account for this kind of indexicality, I modify the definition of the interpretation function $\llbracket \cdot \rrbracket$ slightly. The situational content of a rich demonstration is indexed according to an assignment function g from K : $\llbracket \delta \rrbracket^{g_K}$. g requires that δ be assigned a sub-situation of c : $\llbracket \delta \rrbracket^{g_K} \leq c_K$. This may be a situation where, for example, the speaker has a tattoo, where there is a large scratch on the wall, where the speaker's glasses are broken, etc. as required contextually.

2.2 Coherent Discourse Moves

2.2.1 Moves on the Record

I characterize rich demonstrations as complete discourse moves, which means that they are capable of coherently relating to other complete discourse moves. In this subsection I will give a gloss of the theoretical role the notion of *discourse move* plays and connect it with discourse coherence theory (Asher and Lascarides 2003; Hobbs 1979, 1990; Kehler 2002).⁴

I have already noted one reason for thinking of rich demonstrations as full discourse moves. It has to do with the gesture's attachment patterns—its timing with speech does not matter much for its status as a rich demonstration. Additionally though, when the gesture occurs post-speech, it can follow an overt propositional discourse marker such as 'because' or 'and', as in (30):

(30) My parents are going to be furious, because ARM TURN: tat

⁴These considerations are related to those I discuss in Chapter 1. However the point here is slightly weaker—I take it that one could hold that something (e.g. rich demonstration) is a discourse move while rejecting that it is a speech act (a conclusion I argued for in the previous chapter). But the reverse is not true, anything that is a speech act in the sense discussed previously will automatically be a discourse move according to the considerations I discuss here.

In fact, the most natural cases of rich demonstration involve antecedent intonational cues that would normally precede a linguistic continuation.⁵ I take these observations as evidence that rich demonstrations *slot in* to discourse as full units.

This leads to the second, more important kind of reason for thinking of rich demonstrations as entire moves in a discourse. It has to do with the intuitive effects they have in conversation. A *discourse move* is a somewhat theoretical notion, but I hope to not use it in a way that is tied to any one theory. A discourse move can be thought of as an action one performs with a special communicative intent.⁶ Individual words used in utterances, unconscious ticks, and deictic gestures that fix the referent of a demonstrative do not count as having a special communicative intent associated with them in this sense. The purpose behind those things (if there is any) is to complete some fuller discourse move—e.g. to complete an utterance that is to function as an assertion. This gloss is of course vague. My intention behind saying “I took a long road trip last year” might be to fill in the gaps of some larger story, and so the intent in some sense may be to facilitate some other intention, yet it should be a discourse move of its own. Nevertheless, there is an equally good sense in which the utterance is made with an intention aimed at changing the discourse in a recognizable way.

Given this overall gloss on what it is to be a discourse move, I think it is clear that rich demonstrations count as full moves unto themselves. They are performed with communicative intentions that do not seem parasitic on the intentions behind other moves. Instead, they are done so that they themselves can accomplish something communicative. I therefore follow other recent super linguistic work in analyzing gestures as acts capable of coherently relating to other speech acts (Hunter 2018; Hunter et al. 2018; Lascarides and Stone 2009a,b; Stone and Stojnić 2015). Each of these prior theories focuses on iconic or emblematic ges-

⁵For example: the assertion in (26) would be said in exactly same way as its counterpart in the purely linguistic version.

⁶I am not making any meta-semantic claims. I use this gloss only as a way of tracking the concept of a discourse move, not as a description of what grounds an action’s status as a discourse move (or as communicative). For my purposes, it does not matter whether discourse moves constitutively involve intentions. Since they tend to pattern with intentions, this is a good way of understanding the idea.

tures, but I extend this line of research by including rich demonstration—whose vehicle is a deictic gesture—among the set of possible discourse moves. Given this classification, I update the definition of a model to delineate move-types A and D , for assertion and rich demonstration respectively. Each move type takes an argument— A must take a linguistic segment and D must take a deictic gesture:

$$\mathbf{Model (v2)} = \langle \mathcal{S}, \mathcal{W}, \mathcal{L}, \llbracket \cdot \rrbracket, \mathcal{M}, \rangle$$

A model is a tuple of the set of all possible situations \mathcal{S} , the set of all possible worlds \mathcal{W} , the set of well-formed meaningful linguistic segments \mathcal{L} , as defined in v1. A model also includes the set of possible move types \mathcal{M} , and an interpretation function $\llbracket \cdot \rrbracket$ such that:

- Possible moves include assertions and rich demonstrations. Assertions are of segments from \mathcal{L} :
 $\mathcal{M} = \{A(l), D(\delta)\}$ such that $l \in \mathcal{L}$ and δ is a deictic gesture.
- $\llbracket \cdot \rrbracket$ complex interpretation function either from linguistic segments to worlds, so that $\llbracket l \rrbracket^g \subseteq \mathcal{W}$, or from indexical gestures in context to situations, so that $\llbracket \delta \rrbracket^g \in \mathcal{S}$

The types of moves making up \mathcal{M} are ways of updating conversation states according to whatever the content expressed is. In classifying rich demonstrations as full discourse moves, I am aligned with Hunter et al. (2018). However, as I noted in the previous chapter, we differ in how we conceive of discourse itself. Whereas Hunter et. al. aim to blend discourse facts and general observation, I maintain a strict separation between the moves that have been made in a conversation (and their dynamic effects), and what composes the surrounding context.

Rich demonstrations, along with other moves, function to alter conversation states. Conversation states track information throughout discourse in the following way:

Conversation State (v1) $K = \langle M, CS, c, g \rangle$

A conversation state (or discourse) is a tuple of:

- An ordered list of moves made, whose types are from $\mathcal{M} : M = \langle m_1 \dots m_n \rangle$
- The Context Set CS , which is the set of worlds from \mathcal{W} considered live possibilities by the participants: $CS = \{w_1 \dots w_n\}$
- The current context $c \in \mathcal{S}$
- A standard assignment function g .

M is the simplest element of the discourse. It is merely a record of the moves as they happened in order. For my purposes, this will be a chain of assertions, and rich demonstrations. The Context Set, following the tradition of Stalnaker (1978), represents the possibilities the discourse participants take to be live options for the purposes of conversation. The current context c is a situation consisting of all situations that can be indexically utilized at any given point, and the assignment function fixes the content of indexicals (rich demonstrations) on the basis of contextual factors such as the angle of the vector projected by a pointing gesture.⁷ The moves recorded in M define a series of updates to a conversation state K . I notate generic moves ‘ $m(x)$ ’ where m is either an assertion A or a demonstration D and x is either a linguistic segment l or a gesture δ , accordingly. Whether a move is an assertion or a rich demonstration, at its core it has the same type of effect:

Discourse Update (v1)

A move $m(x) \in \mathcal{M}$ updates a conversation’s move list and shrinks the CS to include only those worlds that contain or satisfy the move’s content: $K_n[m(x)] = K_{n+1}$:

- $M_{n+1} = M_n + m(x)^{K_1}$

⁷This could be spelled out in further formal detail along the lines of Lascarides and Stone (2009a).

- $CS_{n+1} = CS_n \cap \{w : \llbracket x \rrbracket^{g_{n+1}} \leq w\}$

A move $m(x)$ must be one of the move-types defined in M of the model, which for the purposes here means it is either an assertion or a rich demonstration. In the case of an assertion $A(l)$, this means that the assertion is recorded among the list of moves and that the CS becomes a subset of $\llbracket l \rrbracket$ in the familiar Stalnakerian way. For a demonstration $D(\delta)$, the move is recorded and the CS shrinks to include only worlds that have the indicated situation $\llbracket \delta \rrbracket$ as a part.⁸ We can see how this fits together in the following simple example:

- (31) — K_0
- I moved your yoga mat. — K_1
 - It's in the hall closet. — K_2

The speaker's first assertion updates the discourse-initial state K_0 with a newly recorded move and updates the CS so that all worlds reflect that the speaker moved the yoga mat. This new state, K_1 , is then updated in the same way to become K_2 , with a new move and smaller CS .⁹

K_1

M_1	=	$\langle A(31a)^{K_0} \rangle$
CS_1	\subseteq	$\llbracket (31a) \rrbracket^{g_1}$

⁸The definition of CS update works for both assertive and demonstrative update because worlds are maximal situations. When x is a linguistic segment $l \in \mathcal{L}$, the only worlds w such that $\llbracket l \rrbracket \leq w$ will be the very worlds in $\llbracket l \rrbracket$. So in the case of assertion, the definition becomes equivalent to: $CS_2 = CS_1 \cap \llbracket l \rrbracket$. For a demonstration, x is a gesture δ that indicates a situation $\llbracket \delta \rrbracket$, and the CS must only include worlds that have $\llbracket \delta \rrbracket$ as a part.

⁹In these representations of conversation states, c and g are omitted, as their updates are backgrounded and their contents impact the rest of K but not vice-versa.

K_2

M_2	$=$	$\langle A(31a)^{K_0}, A(31b)^{K_1} \rangle$
CS_2	\subseteq	$\llbracket (31a) \rrbracket^{g_1}, \llbracket (31b) \rrbracket^{g_2}$

The CS row lists propositions of which the CS is a subset, since it is just a set of worlds. The end state K_2 is one that records the pair of assertions along with the states against which they were made and the current discourse knowledge that’s a subset of the learned content.

Note that this minimal system is only designed to track simple informational update. This is to elucidate what I take to be the core feature of rich demonstration, that it can be informative like an assertion. A limitation of this high level of abstraction is that the at-issue status of updates is not captured. How to theoretically characterize and diagnose the at-issue/not-at-issue distinction is controversial.¹⁰ But a standard test is direct rejection: roughly, if “That’s false” can target some propositional content, then that content is at-issue. But following the rich demonstration in (26) with “That’s false” would seem to target the parents getting upset, not the reason that was offered as to why. That the cause of their being upset cannot be targeted by rejection is evidence that it is not-at-issue.¹¹ The framework I am sketching here does not explicitly encode these differences, but it is general enough that it could incorporate various ways of modeling that at-issue/not-at-issue distinction.¹²

More to the point: these resources by themselves are insufficient for capturing the core effects of rich demonstrations. The definition of demonstration update requires, as it should, that the CS be compatible with the indicated scene. But rich demonstrations are rich

¹⁰See Anderbois et al. (2013), Murray (2014), Simons et al. (2010), Snider (2017), Syrett and Koev (2014), and Tonhauser (2012).

¹¹An anonymous reviewer suggests that a rich demonstration’s contribution is possibly at-issue when it is post-speech. I am open to this possibility, but am not yet convinced of anything decisive about the at-issue status of rich demonstration. Further research into the timing of deictic gestures is needed.

¹²For example, Questions Under Discussion (Beaver et al. 2017; Roberts 2012a; Simons et al. 2010) could be added as an element of the conversation state K , which would then provide a way of drawing the distinction.

because they communicate more than what is perceptually available. With respect to the tattoo case, for example, the current definitions only permit recording that the speaker has a tattoo, but not that it explains anything. We therefore need a mechanism for predicting inferences of the kind of content regularly communicated using deictic gestures. Toward this end, I turn to discourse structure.

2.2.2 Adding Discourse Structure

I argued in Chapter 1 that an analysis of rich demonstration should be based in discourse coherence theory (Asher and Lascarides 2003; Hobbs 1979; Hobbs et al. 1993; Kehler 2002). That is, my approach assumes that there are background conventions of conversation that conceptually relate rich demonstrations to surrounding discourse and that generate predictable inferences.

To implement this formally, my analysis does not necessitate the machinery posited by other theories such as SDRT. My focus is on the fact that relations hold between moves and can bring new implicit information into a discourse. Toward this end, I add a set of coherence relations \mathcal{C} to the definition of a model:

$$\mathbf{Model (v3)} = \langle \mathcal{S}, \mathcal{W}, \mathcal{L}, \mathcal{M}, \mathcal{C}, [\cdot] \rangle$$

A model is a tuple of the set of all possible situations \mathcal{S} the set of all possible worlds \mathcal{W} , the set of linguistic phrases \mathcal{L} , and the set of possible move types \mathcal{M} , as defined in v2. A model additionally includes the set of all coherence relations \mathcal{C} such that:

- *Explanation* and *Elaboration* (the important relations for this chapter) are included in \mathcal{C} and may hold between moves from \mathcal{M}
 - Other potential relations in \mathcal{C} include *Result*, *Contrast*, *Violated Expectation*, and *Question-Answer Pair*

I focus on *Explanation* and *Elaboration* because they are the important relations for my purposes (I will discuss why this is shortly), and because the other coherence relations are too numerous to efficiently detail in this formalism, though I have listed some natural relations that would be included in \mathcal{C} in a more complete treatment that went beyond rich demonstrations.¹³

Coherence theory has traditionally been utilized only to understand the structure of discourses whose parts are purely linguistic, but it can naturally be extended. Put simply, my view is this: discourse is constituted by moves, and those moves need not be of the same sort. Some will be explicitly linguistic (utterances), and, as argued above, some will not (gestures). Moves are coherently related to one another on the basis of both their formal features and content. The operative coherence relation between two moves will have its stereotypical semantic effects on the discourse, for example *Result* entailing that one event preceded and caused another. In essence, this approach assigns no special status to gestures that richly demonstrate. The *richness* of a rich demonstration is accounted for by the mechanisms of discourse coherence.

My approach follows other theorists (Hunter 2018; Lascarides and Stone 2009a,b; Stone and Stojnić 2015) in conceiving of gestures as the kinds of things that can stand in coherence relations to other discourse moves. They consider iconic gestures to be full discourse units that attach to surrounding discourse via *coherence relations*—roughly, conceptual relations that impose semantic and structural constraints on discourse interpretation. That is, they analyze gestures as independent semantic entities that tie into discourse in the same way that different speech acts do. Gestures are capable of relating to speech (and to other gestures) in a variety of ways that explain both how gestures become associated with their content and how that content is integrated into discourse. I extend these analyses by considering how a non-iconic gesture can coherently relate to other parts of discourse.

¹³Additionally, *Explanation* and *Elaboration* are defined to relate elements of the power set of \mathcal{M} . This is to account for cases in which e.g. explanations are given of entire existing discourse structures, in the sense of Asher and Lascarides (2003).

In a discourse consisting in two moves m_1, m_2 , the discourse will change so as to satisfy their contents $\llbracket m_1 \rrbracket, \llbracket m_2 \rrbracket$ as well as the informational requirements set by the coherence relation that holds between the moves $R(m_1, m_2)$. The determination of R may be sensitive to the formal features of both moves as well as to general pragmatic reasoning (these are reflected in a discourse-structuring function S , which I will discuss shortly). In (26), then, the discourse is affected by the content of the initial assertion about the speaker's parents and by how it coherently relates to the demonstration, via *Explanation*. That relation roughly requires that all CS worlds be such that the explaining situation causes the explained situation. Technically, causation is too strict to be the semantic value of the *Explanation* relation, since one situation can explain another without causing it. But the aim is not to give a conceptual analysis of explanation, but rather to show how the coherence relation's presence alters the discourse. In what follows, a fuller treatment could replace causation without issue. I formally treat semantic coherence relations—those that impose informational constraints on discourses—as updates to conversation states. The definitions for the updates for *Explanation* and *Elaboration* are defined below:

Main Situation: $\text{sit}(x)$:

- If $x \in \mathcal{L}$, then $\text{sit}(x) =$ the $s \in \mathcal{S}$ such that:
 - $\forall w \in \llbracket x \rrbracket^g : s \leq w$, and
 - $\forall s' \in \mathcal{S} : \text{if } \forall w \in \llbracket x \rrbracket^g : s' \leq w, \text{ then } s' \leq s$
- Otherwise $\text{sit}(x) = \llbracket x \rrbracket^g$

Explanation Update: $K_n[\text{Exp}(m_1(x), m_2(y))] = K_{n+1} :$

- $CS_{n+1} = CS_n \cap \text{cause}(\text{sit}(x), \text{sit}(y))$

Elaboration Update: $K[\text{Elab}(m_1(x), m_2(y))] = K_{n+1} :$

- $CS_{n+1} = CS_n \cap \{w : \text{sit}(x) \leq \text{sit}(y) \leq w\}$

The main situation function $\text{sit}(\cdot)$ extracts what is intuitively the main situation from the argument to a discourse move. If the argument is a linguistic segment (detailed by the first condition, if $x \in \mathcal{L}$), as it will be when the move is an assertion, then the main situation is the maximal situation common to all worlds in the segment’s content. That is, the main situation of a sentence is whatever the “biggest” situation is that is a piece of every world in the sentence’s content.¹⁴ If the argument is not a linguistic segment (the “otherwise” condition in the definition), then it must be a deictic gesture. The main situation then is just the one determined by the conversation’s assignment function for the gesture.

Explanation requires that the main situations of the two moves stand in a causal relation, and so intersects the *CS* with the proposition that the one main situation caused the other (in the loose sense of “cause”). *Elaboration* requires that the main situations stand in a part-whole relation, and so intersects the *CS* with the worlds in which that relation holds. To incorporate coherence relations into discourse, I add a structuring function to conversation states:

Conversation State (v2) $K = \langle M, S, CS, c, g \rangle$

A conversation state is a tuple of an ordered list of moves made M , the Context Set CS , the context-situation c , and assignment function g as defined in v1. K also includes a discourse structuring function S such that

- S is a recursive partial function that structures the discourse. It typically connects moves to other moves with coherence relations, but can also connect existing coherence structures previously imposed by S (for the cases of complex structure mentioned above): $S : (M \times M) \cup (M \times S) \rightarrow \mathcal{C}$

¹⁴This is, of course, overly simple for an ideal definition of what the main situation of a sentence is. Modals, conditionals, implicatures, and more introduce difficult complications. I am inclined to think that these difficulties are related to those surrounding discourse topic discussed by Asher (2004). But the simple definition provided here should suffice for the purposes of this chapter, since the sentences are quite simple and the focus is on how *gestures* interact with situations from speech and surrounding context.

To unpack this definition: coherence theorists study the grammatical and rational factors of discourse that influence which coherence relations hold between which moves. Those factors appear in this system as imposed by S —it structures the discourse. The paradigm impact of S , and its simplest impact, will be to relate two moves by some coherence relation such as *Explanation* or *Contrast*. For example, an element of S might be the tuple $\langle\langle A(l_1), A(l_2)\rangle, Exp\rangle$, which says that the assertions of l_1 and l_2 are related by *Explanation*. S can even relate moves to to an existing complex relation (say an *Elaboration* between two moves) via e.g. *Result* (Asher and Gillies 2003). It may be, for example, that an entire narrative (a series of moves related by *Narration*) explains some fact. In that case, an element of S may be: $\langle\langle A(l_1), \mathbf{N}\rangle, Exp\rangle$, where \mathbf{N} is itself another member of S : $\langle\langle A(l_2), A(l_3)\rangle, Nar\rangle$.

S is a partial function, so not every move-move or move-structure pair will be assigned a coherence relation—after all, a move made thirty minutes later shouldn’t necessarily directly relate to the first assertion made.¹⁵

Coherence relations brought into discourse by S are not themselves moves in \mathcal{M} . Instead, moves are modified to force a two-stage update to a discourse: first with the standard move-effect, then with the effect of the new coherence relation.

Discourse Update (v2)

A move $m(x)$ updates a conversation in two stages. $K_n[m(x)] = K_{n+1}$:

First, as in v1, it updates the move list and shrinks the CS to include only those worlds that contain or satisfy the move’s content. In addition, the first stage coherently connects the move to prior discourse, unless the move is discourse initial.

- $M_{n.5} = M_1 + m(x)^{K_n}$

¹⁵Even though S is a partial function, it is a consequence of the definitions in Discourse Update v2 that every move (except discourse-initial ones) be coherently related to something else in the discourse: $\forall m \in M : \exists \beta \in M \cup S, \exists \gamma \in \mathcal{C}$ such that $\langle\langle m, \beta\rangle, \gamma\rangle \in S$ unless $|M| = 1$.

- $CS_{n.5} = CS_n \cap \{w : \llbracket x \rrbracket^{g_{n.5}} \leq w\}$
- If $|M_{n.5}| > 1$, then
 - $S_{n.5} = S_1 \cup \{\langle\langle m(x), \beta \rangle, \gamma \rangle\}$
for some $\beta \in M_n \cup S_n$ and some $\gamma \in \mathcal{C}$
 - otherwise $S_{n.5} = S_n$

Second, the discourse’s informational content is updated according to the newly established coherence relation in $S_{n.5}$:

- $K_{n+1} = K_{1.5}[\gamma(m(x)^{K_n}, \beta)]$

Each move still first has its characteristic effect on the discourse: to update the record of moves and update what is mutually known. Now each move additionally relates to some prior move or coherence structure β by some coherence relation γ thanks to the structuring function S (unless the move is discourse initial). This intermediate conversation state $K_{1.5}$ is then updated with the semantic effect of whatever the new coherence relation between the move and β is: $\gamma(m, \beta)$.

2.2.3 Worked Cases

The easiest way to see how these pieces come together is through examples. I begin with this chapter’s core case, then move to other selected cases from Chapter 1. In what follows, I use shorthand notations for propositions ‘**predicate**(*argument*)’, and notate situations using boldface ‘**sit**’.

2.2.3.1 Tattoo

The conversation begins at the discourse-initial K_0 , without any moves or special narrowing of the CS . The assertion updates the conversation to K_1 :

$$K_1 = K_0[A(26)]$$

M_1	=	$\langle A(26)^{K_0} \rangle$
S_1	=	\emptyset
CS_1	\subseteq	$\llbracket (26) \rrbracket^{g_1} = \mathbf{furious}(parents)$

As always, the assertion is recorded among the list of moves made and shrinks the CS accordingly: it becomes mutually accepted that the speaker's parents will be furious. S does no structuring work because the assertion is discourse-initial (and so the second stage of the update is ineffectual). For this reason, $K_1 = K_{0.5}$, and I omit the representation of $K_{0.5}$ (because K_1 is the final output state). The rich demonstration introduces a substantive two-stage update, whose intermediate stage is $K_{1.5}$:

$$K_{1.5} = K_1[D(\delta)]$$

$M_{1.5}$	=	$\langle A(26)^{K_0}, D(\delta)^{K_1} \rangle$
$S_{1.5}$	=	$\{\langle \langle D(\delta)^{K_1}, A(26)^{K_0} \rangle, Exp \rangle\}$
$CS_{1.5}$	\subseteq	$\llbracket (26) \rrbracket^{g_1} = \mathbf{furious}(parents),$ $\{w : \llbracket \delta \rrbracket^{g_{1.5}} = \mathbf{tat} \leq w\}$

The demonstration is added to the list of moves, and the CS shrinks so that the indexically-fixed tattoo-situation (part of the conversation state's context: $\mathbf{tat} \leq c_{K_1}$) is part of all live worlds—all interlocutors are now coordinated on the existence of the tattoo. The structuring function sets *Explanation* as the connecting relation, according to various factors including pragmatic reasonableness as well as the formal features of the discourse. The second-stage update results in K_2 .

$$K_2 = K_{1.5}[Exp(D(\delta)^{K_1}, A(26))]$$

$M_{2=1.5}$	$= \langle A(26)^{K_0}, D(\delta)^{K_1} \rangle$
$S_{2=1.5}$	$= \{ \langle \langle D(\delta)^{K_1}, A(26)^{K_0} \rangle, Exp \rangle \}$
CS_2	$\subseteq \{ w : \llbracket \delta \rrbracket^{g_{1.5}} = \mathbf{tat} \leq w \},$ $\text{cause}(\mathbf{tat}, \mathbf{furious})$

The context set is the only part of the conversation state that is affected by the second-stage update from $K_{1.5}$ to K_2 . This requires, in accordance with the semantics of *Explanation*, that all live worlds be ones in which the main situation of the indicated scene ($\text{sit}(\delta) = \mathbf{tat}$) causes the main situation of the asserted linguistic segment ($\text{sit}(26) = \mathbf{furious}$). This end result is a conversation state that records both discourse moves, structures them according to the *Explanation* relation, and sets all open possibilities to be ones compatible with the content of the sentence, the indicated scene, and the informational impact of the coherence relation. The rich demonstration, then, succeeds in functioning to communicate complex content thanks to the mechanisms underlying discourse structure and update. All it needs to supply is the situation indexically retrieved from the conversation's context.

The semantics allows for coherence relations not just between moves, but between moves and existing discourse sub-structures. To see this, imagine the hearer of (26) replies with (32):

(32) But they have tattoos themselves!

In this case, they convey *Violated Expectation*. But they are not surprised that the speaker has a tattoo or merely that the parents will be upset. Intuitively, they are surprised that the parents will be upset *because* of the tattoo. So the fact that they have tattoos contrasts with the *Explanation* structure resulting from the rich demonstration. The first stage update caused by the response returns $K_{2.5}$:

$$K_{2.5} = K_2[A(32)]$$

$M_{2.5}$	$=$	$\langle A(26)^{K_0}, D(\delta)^{K_1}, A(32)^{K_2} \rangle$
$S_{2.5}$	$=$	$\{\langle \langle D(\delta)^{K_1}, A(26)^{K_0} \rangle, Exp \rangle, \langle \langle A(32)^{K_2}, \langle \langle D(\delta)^{K_1}, A(26)^{K_0} \rangle, Exp \rangle \rangle, Vio \rangle\}$
$CS_{2.5}$	\subseteq	$\llbracket (26) \rrbracket^{g_1} = \text{furious}(\text{parents}),$ $\{w : \llbracket \delta \rrbracket^{g_{1.5}} = \mathbf{tat} \leq w\},$ $\text{cause}(\mathbf{tat}, \mathbf{furious}),$ $\llbracket (32) \rrbracket^{g_{2.5}} = \text{have-tattoos}(\text{parents})$

The first-round update caused by the reply in (32) adds to the list of moves, is assigned the *Violated Expectation* structural relation to the existing *Explanation* relation from K_2 , and shrinks the context set to ensure coordination on the fact that the original speaker's parents have tattoos. I have not given the update semantics for *Violated Expectation*, but for simplicity we can think of it as imposing a new proposition on the *CS* that gives information about the speaker's mental state—they're surprised about the information imposed by the *Explanation* relation. Where ' h ' is a constant for the speaker of (32) (and hearer of (26)):

$$K_3 = K_{2.5}[Vio(A(32)^{K_2}, \langle \langle D(\delta)^{K_1}, A(26)^{K_0} \rangle, Exp \rangle)]$$

$M_{3=2.5}$	$=$	$\langle A(26)^{K_0}, D(\delta)^{K_1}, A(32)^{K_2} \rangle$
$S_{3=2.5}$	$=$	$\{\langle \langle D(\delta)^{K_1}, A(26)^{K_0} \rangle, Exp \rangle, \langle \langle A(32)^{K_2}, \langle \langle D(\delta)^{K_1}, A(26)^{K_0} \rangle, Exp \rangle \rangle, Vio \rangle\}$
$CS_{3=2.5}$	\subseteq	$\llbracket (26) \rrbracket^{g_1} = \text{furious}(\text{parents}),$ $\{w : \llbracket \delta \rrbracket^{g_{1.5}} = \mathbf{tat} \leq w\},$ $\text{cause}(\mathbf{tat}, \mathbf{furious}),$ $\llbracket (32) \rrbracket^{g_{2.5}} = \text{have-tattoos}(\text{parents}),$ $\text{surprised}(h, \text{cause}(\mathbf{tat}, \mathbf{furious}))$

The final result is a discourse that has evolved through two assertions and a rich demon-

stration, where the first speaker communicated that their parents will be upset by their new tattoo, and the hearer communicated surprise about that fact.

2.2.3.2 Scratch

As a second case, I examine (9) from Hunter et al. (2018), repeated here:

- (9) a. $\frac{\text{NOD: scratch}}{\text{I moved the table into the living room this morning.}}$
 b. I had to buy some new paint.

Though it is co-speech, I treat the gesture as structurally discourse-initial.¹⁶ The demonstration updates the list of moves and shrinks the *CS* to include only worlds in which **scratch** obtains.

$$K_1 = K_0[D(\delta)]$$

M_1	=	$\langle D(\delta)^{K_0} \rangle$
S_1	=	\emptyset
CS_2	\subseteq	$\{w : \llbracket \delta \rrbracket^{g_1} = \mathbf{scratch} \leq w\}$

This demonstration is in a sense slightly less rich than are those in the other cases. It still counts as a discourse move that updates the conversation state with new information, but because it is discourse-initial, it cannot coherently relate to anything prior.¹⁷ Nevertheless, as is clear, it coherently relates to moves that come later. The speaker’s assertion is added to *M*, its content is added to the *CS*, and and the discourse-structuring function *S* imposes

¹⁶As I discuss in the next section, the specific effects of a rich demonstration’s timing make for a clear next step for research.

¹⁷This is also why for discourse-initial moves in each of these cases, I only represent a one stage update, even though technically two stages take place. It is just that the second stage returns a *K* identical to the intermediate *K* resulting from the first stage, since there are no constraints imposed by any coherence structure.

Explanation. That the *Explanation* reading is preferred is a matter of the grammatical and pragmatic features of the assertion. Where ‘*j*’ is a constant for the speaker:

$$K_{1.5} = K_1[A(9a)]$$

$M_{1.5}$	=	$\langle D(\delta)^{K_0}, A(9a)^{K_1} \rangle$
$S_{1.5}$	=	$\{\langle \langle A(9a)^{K_1}, D(\delta)^{K_0} \rangle, Exp \rangle\}$
$CS_{1.5}$	\subseteq	$\{w : \llbracket \delta \rrbracket^{g_1} = \mathbf{scratch} \leq w\},$ $\llbracket (9a) \rrbracket^{g_{1.5}} = \mathbf{move}(j, table)$

The update’s second stage forces the causal semantics of *Explanation* to be reflected in the *CS*—the moving situation must cause the scratch situation:

$$K_2 = K_{1.5}[Exp(A(9a)^{K_1}, D(\delta)^{K_0})]$$

$M_{2=1.5}$	=	$\langle D(\delta)^{K_0}, A(9a)^{K_1} \rangle$
$S_{2=1.5}$	=	$\{\langle \langle A(9a)^{K_1}, D(\delta)^{K_0} \rangle, Exp \rangle\}$
CS_2	\subseteq	$\{w : \llbracket \delta \rrbracket^{g_1} = \mathbf{scratch} \leq w\},$ $\llbracket (9a) \rrbracket^{g_{1.5}} = \mathbf{move}(j, table),$ $\mathbf{cause}(\mathbf{move}, \mathbf{scratch})$

Unlike with (26), in this case it is an assertion that explains a rich demonstration. As desired, this requires that the main situation making up the worlds in the proposition causes the demonstrated situation, the scratch on the wall. The followup assertion shrinks the *CS* and relates to the whole *Explanation* structure by *Result*, reflected in the structuring function *S*:

$$K_{2.5} = K_2[A(9a)]$$

$M_{2.5}$	$=$	$\langle D(\delta)^{K_0}, A(9a)^{K_1}, A(9a)^{K_2} \rangle$
$S_{2.5}$	$=$	$\{ \langle \langle A(9a)^{K_1}, D(\delta)^{K_0} \rangle, Exp \rangle, \langle \langle A(9a)^{K_2}, \langle \langle A(9a)^{K_1}, D(\delta)^{K_0} \rangle, Exp \rangle \rangle Result \} \}$
$CS_{2.5}$	\subseteq	$\{ w : \llbracket \delta \rrbracket^{g_1} = \mathbf{scratch} \leq w \},$ $\llbracket (9a) \rrbracket^{g_{1.5}} = \mathbf{move}(j, table),$ $\mathbf{cause}(\mathbf{move}, \mathbf{scratch}),$ $\mathbf{buy}(j, paint)$

This means that the asserted proposition $\mathbf{buy}(j, paint)$ relates via *Result* to the *Explanation* structure previously established thanks to the assertion. This outcome is desirable: as a result of moving the table and its causing the scratch, the speaker had to buy new paint. The update's second stage reflects the causal semantics of *Result* in the *CS*:

$$K_3 = K_2[A(9a)]$$

$M_{3=2.5}$	$=$	$\langle D(\delta)^{K_0}, A(9a)^{K_1}, A(9a)^{K_2} \rangle$
$S_{3=2.5}$	$=$	$\{ \langle \langle A(9a)^{K_1}, D(\delta)^{K_0} \rangle, Exp \rangle, \langle \langle A(9a)^{K_2}, \langle \langle A(9a)^{K_1}, D(\delta)^{K_0} \rangle, Exp \rangle \rangle Result \} \}$
CS_3	\subseteq	$\{ w : \llbracket \delta \rrbracket^{g_1} = \mathbf{scratch} \leq w \},$ $\llbracket (9a) \rrbracket^{g_{1.5}} = \mathbf{move}(j, table),$ $\mathbf{cause}(\mathbf{move}, \mathbf{scratch}),$ $\mathbf{buy}(j, paint),$ $\mathbf{cause}(\mathbf{cause}(\mathbf{move}, \mathbf{scratch}), \mathbf{buy-paint})$

The final result is a conversation state that reflects the information given explicitly in both assertions and directly indicated by the gesture, as well as connects everything within a complex causal chain.

2.3 Future Research

I have argued that rich demonstration is a phenomenon that should be characterized as a full discourse move, and I have provided a high-level proposal for understanding how this can be so. In this section, I provide preliminary data that speak to rich demonstration's more specific empirical profile, and so go beyond the stated aim of this chapter. Despite the need for future research regarding exactly how the data should be understood, I take the observations in this section to provide additional indirect support for the perspective I have argued for.

In the previous chapter I offered data that indicate that there are interpretive constraints on what kinds of contents rich demonstrations can express. In those cases, a particular interpretive scheme *should be* pragmatically available, but it can only be felicitously realized when the rich demonstration leads the assertion. I argued that this is evidence that rich demonstrations are sensitive to the *coordinating/subordinating* structure of coherence relations. In the formal system put forth in this chapter, this hypothesis could be understood as a requirement on the discourse-structuring function S :

Rich Demonstration Constraint on Structure

For any rich demonstration D indicating a situation δ in a conversation state K_n and any element in $S_{K_{>n}} \langle \langle D(\delta)^K, \beta \rangle, \gamma \rangle$ such that $\beta < D(s)$ in $M : \gamma$ must be a Subordinating relation from \mathcal{C} .

More simply: for any coherence-structure in S with a rich demonstration as the move in the “more recent” position, the coherence relation γ must be subordinating.

But additional data would need to be accommodated by this hypothesis. For example, though (33) is questionable, adding the explicit contrast-marker ‘but’ in (34) greatly improves its acceptability, while adding ‘and’ in (36), which forces coordinating relations (Txurruka 2003), does not improve the acceptability of (35):

(33) *Contrast?*

??I always heard that people in New York walk fast

OPEN HAND: slow

(34) I always heard that people in New York walk fast, but

OPEN HAND: slow

(35) *Result?*

??Leila threw a football

OPEN HAND: vase

(36) *Result*

??Leila threw a football, and

OPEN HAND: vase

It may be that there are delicate mechanisms at work, perhaps having to do with the fact that ‘but’ is tied to a particular coherence relation whereas ‘and’ only to a family of them. Such intricacies are to be expected from the perspective of a coherence-theoretic framework.

An alternative hypothesis would explain these data not in terms of the properties of rich demonstration *qua* discourse move, but rather in terms of larger discourse-structuring mechanisms present in the cases. For example, it may be that the patterns of acceptability are explained by the Question Under Discussion (QUD, Beaver et al. 2017; Roberts 2012a). Evidence for this approach over the previous one is provided by the fact that parallel observations can be made of cases involving linguistic counterparts to rich demonstrations:¹⁸

(37) *Contrast?*

??I always heard that people in New York walk fast. That person’s walking slowly.

(38) *Contrast*

That person’s waling slowly. I always heard that people in New York walk

¹⁸I’m grateful to an anonymous reviewer for emphasizing this point and initially articulating a QUD-based hypothesis.

fast.

This pair seems to follow the same pattern as do the cases of rich demonstration, despite lacking one altogether (though it's highly likely that intonational cues could influence their acceptability in a number of ways). Given that the pattern arises without the presence of a rich demonstration, a hypothesis might be that the implicit QUD (perhaps *do New Yorkers actually walk fast?*) more naturally lends itself to one ordering over another. The easiest way to integrate this approach into the formal analysis provided here would involve (i) modifying the definitions of the interpretation function $\llbracket \cdot \rrbracket$ to capture the content of questions, (ii) adding an element, Q , to the conversation state K which records the discourse's QUD structure, and (iii) modifying the definition of discourse update to capture interactions a move has with Q_K .

As with the previous hypothesis, this QUD-based one would need to account for complexifying data. For example, it seems that the linguistic counterparts of the second pair of cases do not follow the same pattern:

(39) *Result*

Leila threw a football. She broke that vase.

(40) *Explanation*

Leila broke that vase. She threw a football.

In these linguistic versions, even the *Result* reading is acceptable. This would suggest that there is something interfering with the parallel between the gestural and linguistic versions, even though the same QUD is presumably operative in all of the cases. Further research into rich demonstration and into the semantic/pragmatic mechanisms of discourse interpretation more generally should aim to explain these subtleties in the empirical data. In particular, the data surrounding these hypotheses suggest difficult but promising directions for research

into the relation between discourse coherence and Questions Under Discussion (cf. Hunter and Abrusán 2017).

In addition to the discourse-structuring constraints rich demonstration has with respect to prior and posterior speech, future research may investigate specific interpretive effects of a gesture's timing with respect to its co-occurring speech. For example, it is plausible that the following variants of the tattoo case subtly differ in acceptability and/or specific interpretation:

(26) My parents are going to be ^{ARM TURN: tat} furious

(41) ^{ARM TURN: tat} My parents are going to be furious.

(42) My parents are ^{ARM TURN: tat} going to be furious.

(43) My parents are going ^{ARM TURN: tat} to be furious.

The timing of the deictic gesture may function analogously to prosodic emphasis, so that perhaps one backgrounded possible reason for the parents to be upset is made more salient—the timing may for example bias an interpretation that places the explanatory work on the tattoo's depictive content, or alternatively on the parental standards at work.

In addition, the duration of the deictic gesture could affect the rich demonstration's interpretation or even acceptability. A long drawn-out gesture such as the one in (41) may bring a sense of anticipation and revelation, whereas a short quick gesture such as the one in (43) may convey urgency or surprise. Considerations along these lines suggest that there are both semantic and pragmatic effects of a rich demonstration's timing with respect to speech.

This is unsurprising given recent work on the timing of iconic gestures (Ebert et al. 2020; Esipova 2019; Hunter 2018; Schlenker 2018). However given that rich demonstrations lack an iconic element, it is likely that the effects of timing should be understood using different resources. Iconic gestures (roughly speaking) function thanks to their surface properties such as movement, hand shape, duration, and importantly timing with respect to speech. Because a deictic gesture’s surface properties seem less relevant in marking it as a rich demonstration, the theoretical resources needed to capture the effects of timing will likely diverge from those needed in the case of iconic gestures. Instead, analysis of a rich demonstration’s timing seems to require explanations that more closely connect to how deictic reference is resolved in the first place (Lücking et al. 2015).

Rich demonstrations can furthermore engage in far more complex discourse structures than the system here can model. Modal expressions, for example, can combine with rich demonstrations to form complex meanings. (11), the modal anaphora case from Chapter 1, is a simple example of this.

I take rich demonstrations to be only the beginning of research toward new roles for deixis in conversation. Deixis may have discourse functions that go beyond its traditional referential one, but that are more subtle than the fully communicative one investigated here. Gestures that integrate aspects of iconicity and deixis (Ebert et al. 2020), for example, may richly demonstrate in more complex ways.

2.4 Conclusion

I have argued that rich demonstrations offer insight into the role deixis can play in discourse. Contrary to what is traditionally assumed, deixis is not merely a mechanism for determining or assisting the content of a referential linguistic expression. Instead, deixis can serve a fully communicative function by contributing complete informational content to a discourse. Rich demonstrations offer intuitive cases of this, but how they accomplish it is a difficult

problem. I have offered a high-level discourse-update semantics as a solution. It offers a way for deictic gestures to indicate situations pulled from a situation's context and a way to use them to generate complex content. The mechanisms driving that generation are those that underlie discourse structure; they are the conventions for establishing conceptual connections between discourse moves. The final picture is one that allows an attention-directing gesture to integrate seamlessly into conversation as a full contribution in its own right. I presented preliminary data suggesting that the formal features of rich demonstration are complex and make for avenues for future research. In particular, it seems that considerations of information structure and the deictic gesture's formal features (especially timing) have strong influence. I take the analysis presented here, and the analyses it builds on, to be the start of further research into both rich demonstration as a particular phenomenon and into the varied functions of deixis generally.

CHAPTER 3

Conversational Salience and Mutual Attention

It is often useful in theorizing about the semantics and pragmatics of language to appeal to a notion of *salience*. Consider (44) and (45), uttered in the following context:

Theme Park

Jorge and Elena are spending the day at a busy theme park. There are lots of sights, sounds, and people in every direction. While walking around, their conversation is interrupted by a piercing scream from off in the distance. Elena glances in the direction of the scream, then looks back at Jorge and says:

(44) They're sure having a good time!

Jorge then motions in the direction of a nearby baby who has been crying for some time and responds:

(45) *They* sure aren't!

Why is it that 'they' refers to the screaming person in (44) but to the crying baby in (45)? According to one family of views, *salience-based* views, the reference-determining factor is that the screaming person is salient at the utterance-time of (44) while the crying baby is salient at the utterance-time of (45).¹ On such views, salience is a feature of the

¹Some examples of views that I would characterize as salience-based: Gauker (2019), Kaplan (1989b), Mount (2008), Stojnić et al. (2013, 2017), and Wettstein (1984).

semantics or metasemantics of demonstrative pronouns—reference is determined by facts about salience (at least in part). According to an opposing family of views, *intention-based* views, the reference-determining factor is the referential intention of the speaker in each case: Elena intends to refer to the screamer, Jorge intends to refer to the baby.² On these views, salience is a feature of the pragmatics of demonstrative pronouns—salience does not determine reference, but it is a rational guide to referential intentions, which do determine reference (at least in part).

While salience-based and intention-based theories offer competing explanations of reference, they all satisfy a more-or-less obvious intuition: salience plays a role in the interpretation of demonstrative pronouns. Some take that role to be semantic, some take it to be pragmatic. But it would be a mistake to ignore salience entirely.³

What does it mean to say that something is salient in this sense? That is, what is that thing whose interpretive role the salience-based and intention-based theorists disagree about? My aim in this chapter is to argue that this notion of salience, which I'll call *conversational salience*, is best understood in terms of a psychological notion: *mutual attention*. More specifically, I develop an account of conversational salience according to which something is conversationally salient if and only if the conversation's participants have mutually attended to that thing during their conversation. This does not mean that conversational salience reduces to mutual attention. Instead, conversational salience is achieved in virtue of the mutual attention of a conversation's participants.

This position is opposed to views on which conversational salience is a notion *sui generis* to linguistic theory. It is also opposed to views on which conversational salience is best

²Some intention-based views: Bach (1992), Kaplan (1989a), and King (2014). See Michaelson and Nowak (2022) for a recent overview of salience and intention-based views, and for a critical discussion of the former.

³The gloss on views about reference I've provided is quite rough. Reimer (1991, 1992), for example, offers an attempted middle ground. But the nuances of actual views are not important for my purposes. The point is that no matter one's views about what determines reference, the story one tells about interpretation generally must say something about salience.

understood in terms of different psychological mechanisms (in particular, different forms of joint attention).

I will begin by providing rough overviews of some ways in which salience has proven to be a useful notion for linguistic theory in Sec.3.1. I'll push for the prospect of analyzing conversational salience in terms of psychological processes and present two desiderata for doing so in Sec.3.2 In Sec.3.3 I'll argue that mutual attention is the right candidate. The bulk of my argumentation will involve rejecting weaker psychological states as ways of analyzing salience. I will defend mutual attention as the correct way of analyzing salience in Sec.3.4. The upshot of will be an elucidation of salience as the object of theoretical debate in philosophy and linguistics. In understanding the notion in particular psychological terms, the theses and consequences of salience-based and intention-based theories of reference become clearer. For salience-based theories, my position offers deeper insight into what underlies reference. For intention-based theories, my position offers deeper insight into the mechanisms of intention-recognition.

3.1 Why Model Salience

The cases in **Theme Park** involve demonstrative uses of the pronoun 'they', and many of the clearest cases that illustrate the importance of salience in interpretation are demonstrative in the sense of Kaplan (1989b). One early such example, which is standardly taken to make vivid the importance of salience, comes from Robert Stalnaker:

[...] any obviously observable change in the physical surroundings of the conversation will change the presumed common knowledge. If a goat walked into the room, it would normally be pre-supposed, from that point, that there was a goat in the room. And the fact that this was presupposed might be exploited in conversation, as when someone asks *How did that thing get in here?*, assuming that others will know what he is talking about. (Stalnaker 1978, p. 86, emphasis

original)

Some theory-neutral intuitions: “that thing” refers to the goat, the speech act succeeds, and these facts obtain or are known to obtain because of how salient the goat is. For salience-based theories, the success of both reference and communication is owed to the goat’s salience. For intention-based theories, only the communicative success is.

But salience plays a role in the interpretation of referring expressions more generally, not just demonstratives. David Lewis forwarded an early defense of the view that definite descriptions are semantically sensitive to conversational salience: “The proper treatment of descriptions must be more like this: ‘the *F*’ denotes *x* if and only if *x* is the most salient *F* in the domain of discourse, according to some contextually determined salience ranking” (Lewis 1979, p.348). Lewis took salience to be a feature of context that is sensitive to both linguistic and extra-linguistic happenings:

There are various ways for something to gain salience. Some have to do with the course of conversation, others do not. Imagine yourself with me as I write these words. In the room is a cat, Bruce, who has been making himself very salient by dashing madly about. He is the only cat in the room, or in sight, or in earshot. I start to speak to you:

The cat is in the carton. The cat will never meet our other cat, because our other cat lives in New Zealand. Our New Zealand cat lives with the Cresswells. And there he’ll stay, because Miriam would be sad if the cat went away.

At first, “the cat” denotes Bruce, he being the most salient cat for reasons having nothing to do with the course of conversation. If I want to talk about Albert, our New Zealand cat, I have to say “our other cat” or “our New Zealand cat”. But as I talk more and more about Albert, and not any more about Bruce, I

raise Albert’s salience by conversational means. Finally, in the last sentence of my monologue, I am in a position to say “the cat” and thereby denote not Bruce but rather the newly-most-salient cat Albert. (Lewis 1979, pp. 348–349)

Lewis conceived of conversation as a process that continually updates an abstract information state that is shared by all conversational participants—a *conversational score*, or for some, a conversational *record* (Thomason 1990b). Some of the information recorded on the conversational score is a ranking of salient people, places, and things, which can change according to what is said and what happens in the conversational vicinity.

Lewis’s suggestion has proved fruitful for formally modeling conversation. Centering Theory (Grosz and Sidner 1986; Grosz et al. 1995) studies grammatical influences on salience, for example. This approach represents potential discourse referents as a stack, and aims to show that there are empirical generalizations to be made about how the syntactic position of a referential NP (or in the case of Bittner 2009, 2011, implicit references to times and possibilities) adds elements to or re-orders the elements of the stack. The thought is that the stack-structure usefully models how interlocutors interpret later anaphoric and demonstrative NPs.

Explicitly including a stack of salient referents in one’s discourse model makes sense regardless of whether one is a salience-based or intention-based theorist about reference. A recent implementation from the salience-based camp comes from Stojnić et al. (2013, 2017), who borrow but re-conceptualize formal tools from Centering Theory. They argue that there is no distinction between demonstrative and anaphoric uses of pronouns. To make sense of this, they take Lewis’s view a step further: not only can language impact salience, but so too can coherence relations—implicit conceptual connections between speech acts. The act of making an assertion so as to provide a *Narration*, for example, will conventionally raise the sentence’s subject to salience. Making an assertion as an *Explanation*, by contrast, will raise the sentence’s object to salience. This predicts, they argue, the correct referential interpretations of pronouns in assertions that follow.

But one need not think salience is relevant to semantic rules in order to include it in the representation of a conversational score. Craige Roberts, in expanding on her research program (Roberts 2012a,b), builds a system that formally represents a salience ordering of discourse-referents, but explicitly notes: “Salience facilitates the Retrieval of those aspects of the intended meaning of an utterance which are not explicitly given by the conventional content of the utterance” (Roberts 2011, p. 14). On this and other intention-based views, the role of salience in interpretation is to enable the hearer to recognize what the (intention-determined) meaning of a speaker’s utterance is.

I’ll roughly sketch the two styles of explanation with respect to (46) and (47) from Heim (1982), who attributes them to Barbara Partee.

- (46) a. I dropped ten marbles and found all of them, except for one.
b. It’s probably under the sofa.

- (47) a. I dropped ten marbles and found only nine of them.
b. #It’s probably under the sofa.

These cases are interesting because (46a) and (47a) are truth-conditionally equivalent, but their morpho-syntactic differences mean that the same followup is only licensed in the former case. The salience-based and intention-based theorists offer competing explanations for why the followup containing ‘it’ is felicitous in (46b) but not in (47b). For both approaches, though, it’s useful to conceive of the discourse record as tracking salient potential referents. Representing the stack as a tuple where earlier elements are more salient, we can represent one effect of (46a) as enforcing the structure in (48), with a missing marble as most salient, and the whole collection of ten as next-most salient. The structure enforced by (47a), however is different. It encodes an ordering where the collection of nine marbles that were found is more salient than the entire collection of ten. But, because the the lone missing marble isn’t explicitly mentioned, it either isn’t represented as salient at all (49a) or is represented as

least-salient (49b).

(48) *Salience-Structure of (46a):*

⟨missing-one, all-ten⟩

(49) *Salience-Structure of (47a):*

a. ⟨found-nine, all-ten⟩

b. ⟨found-nine, all-ten, missing-one⟩

This kind of representation of salience is useful for both salience-based and intention-based theories. For a salience-based theory, the extension of ‘it’ is determined by the contextual salience-ranking according to something like the rule in (50). For an intention-based theory, the extension is determined by the attitudes of the speaker according to something like the rule in (51):

(50) $[[it]]^c =$ the most salient object in c

(51) $[[it]]^c =$ the object intended by the speaker in c

For the salience-based denotation in (50), the explanation of the felicity in (46b) is that the most salient thing is *missing-one*, whereas (47b) is infelicitous because ‘it’ refers to *found-nine*, which conflicts with the singular requirement of ‘it’.⁴

Intention-based theories, utilizing a denotation like the one in (51), could generally offer one of two explanations. On one, ‘it’ in both (46b) and (47b) refers to *missing-one*. The problem is that that information is rationally inaccessible to the hearer in (47b) because of the

⁴This is only a sketch. More worked out versions would fill out the singular/plural object distinction more fully. Depending on the details, the salience-based explanation may differ. For example, the discourse may separately track singular and plural drefs, and take the denotation of ‘it’ to be the most salient singular dref. In such a case, the explanation of the infelicity of (47b) would be that ‘it’ lacks a referent entirely, since no singular object is salient.

salience structure imposed by (47a). That is, the rational thing for the speaker to do would be to use a pronoun to refer to whatever is currently most salient, and this principle guides the hearer’s interpretation of the speaker’s assertion. Given the shared shared representation of salience in (49) forced by (47a), this principle would lead the hearer to conclude that the hearer intended to refer to the nine missing marbles, which clearly cannot be right. So while there is nothing semantically wrong with (47b), it’s pragmatically uncooperative to assert it.

The other explanation the intention-based theorist could give would focus on the rational capacity to assert (47b) in the first place. Inspired by Grice (1957), the theorist might hold that to intend to refer to something is (among other things) to intend to be recognized as referring to that thing. Because of the absence of *missing-one* from the salience structure resulting from (49a) (or, alternatively, its low-ranking in (49b)), one could not be recognized as referring to *missing-one*. This inability for the intention to be recognized is no different from the explanation above. The difference is that on this explanation the inability to be recognized as referring means that the speaker cannot coherently intend to refer to *missing-one* using ‘it’ in the first place. So (47b) is infelicitous not just because it is an uncooperative move, which it surely is, but because ‘it’ fails to refer—the right kind of good-faith communicative intention could not be formed.⁵

This is a gloss, but the point has been to show that regardless of one’s philosophical stance on reference, there is no escaping the need to appeal to salience in theorizing.

3.2 Conversational Salience, Psychological Attention

Clearly, the notion of *salience* that is relevant to linguistic theory is somewhat specialized. To officially distinguish it from a more colloquial use, I’ll call the intended notion *conversational salience* (though I’ll still occasionally shorten it to *salience*). The salience-based family of

⁵A similar point is made by Kaplan (1989a).

theories and the intention-based family have competing stories about how something's status as conversationally salient factors into interpretation. But what is it to be conversationally salient in the first place? The urge to connect this linguistic notion to something psychological, namely *attention*, is powerful. In fact the terms *salience* and *attention* are often used in the same breathe in philosophical, linguistic, and psychological literatures.⁶ But this is typically only for grammatical efficiency; to be x salient to y is for y to attend to x . I'll reserve *conversational salience* for the linguistic notion and *attention* for the psychological one. The overall goal of this chapter is to argue that the two notions are in fact substantively connected. In this section, I will focus on establishing two desiderata that a psychological account of salience should meet. First, it should account for the sense in which conversational salience is a *shared* information state. Second, it should resist reduction to perceptual attention.

Mount (2008) argues for a salience-based view of demonstratives that, unlike others, explicitly characterizes salience as something psychological:

By saying that something is salient, I mean that it is the focus of perceptual or cognitive attention. Salience, on this view, is not some objective feature that can be determined independently of the mental states of conversational participants; it is essentially a mind-dependent matter. An object is mutually recognized as maximally salient by conversational participants when all interlocutors have focused their attention on it, and are aware that they have all focused their attention on it. Thus my claim is that a demonstrative refers to the object mutually recognized as maximally salient. When there is no such object, the demonstrative does not. Demonstrations themselves have no special place in this theory, except insofar as demonstrating is an especially reliable way of directing attention to a certain thing. Speaker intentions have no special authority either,

⁶For example, Grosz and Sidner (1986) and Stojnić et al. (2017) almost exclusively use *attention* over *salience*, but do not mean it in any psychologically loaded sense.

although the object a speaker has in mind will often be the object maximally salient to conversational participants. (pp. 154–155)

I'll argue that this is not the right way to flesh out conversational salience (or at least that one way of interpreting it is wrong and another is right). But for now I want to highlight the following point. Conversational salience, if it is to be understood in psychological terms, must be determined by the psychological states of all interlocutors. That is, for something to count as conversationally salient, it is not enough that one individual be in the appropriate psychological state. For x to be conversationally salient to us in our conversation, it is not enough that only one of us have the right attentional focus directed toward x . It may be, according to some theories, sufficient for achieving referential success that only one person (presumably the speaker) be in some attentional state. But such examples are at best cases of pragmatic failure and at worst cases of complete semantic failure. Say you and I are talking on the phone about dogs and we haven't mentioned any women recently. I casually say "She told me that her puppy doesn't like dry food". It is dubious that my attempt to refer to somebody using "she" is genuinely successful, but even if it is, it is certainly still a failure of coordination. There is no conversationally salient woman. That is, there is no woman who we are *both* attending to in the right way in our conversation, even if *my* cognitive attention is focused on her in the right way. This is the first desideratum for an account of conversational salience: it must be a function of the psychological states of all conversational participants.

The second point that Mount appropriately captures is that the relevant psychological states need not be perceptual. Many paradigmatic cases of conversational salience involve directions of perceptual attention—both cases in **Theme Park** (the sound calling the attention of Lalo and Noemí, and Lalo's directing Noemí's gaze via a pointing gesture), for example. But, as is illustrated in Lewis's example involving his cats, it is reasonable to think of conversational salience as encompassing objects outside of perception as well ("our other cat lives in New Zealand" makes that cat more conversationally salient than the cat in the

room). Objects of *cognitive attention* must be eligible as well. Following Wu (2014, ch. 3), it is helpful to characterize cognitive attention as *attention in thought*. This is the second desideratum for a psychological account of salience: it must countenance cognitive attention.

Given these two desiderata as starting points, one might be skeptical that a psychological account of conversational salience can be given, with a worry to accompany each one. First, one might worry whether there could be a robust scientific account of joint psychological states, as would be required by the first desideratum. Psychology is the study of *an agent's* psychological states; groups don't have psychological states. Second, one might worry that cognitive attention is too sophisticated a mental state to be scientifically tractable. *What one is thinking about/focusing on* is too complex or abstract a mental state to be theoretically useful, the worry goes.

Both of these worries are raised by Michaelson and Nowak (2022), albeit with a slightly different target. They criticize salience-based views of demonstrative reference, not accounts of conversational salience per se. But their worries apply to the present project. My aim in the remainder of the chapter will be to show that these can be overcome: there is a productive sense in which an object is conversationally salient in virtue of the joint cognitive attention of the conversation's participants.

3.3 Salience and Varieties of Joint Attention

The first desideratum for an attentional theory of conversational salience is that salience must be a function of the psychological states of all conversational participants. This means that conversational salience depends on a state of **joint attention** among interlocutors, as opposed to the **individual attention** of just one of them.

Joint attention is a class of attentional states of varying complexity. In this section I discuss four varieties of joint attention of increasing complexity and argue that only one, **mutual attention** (the most complex), corresponds to conversational salience. The collec-

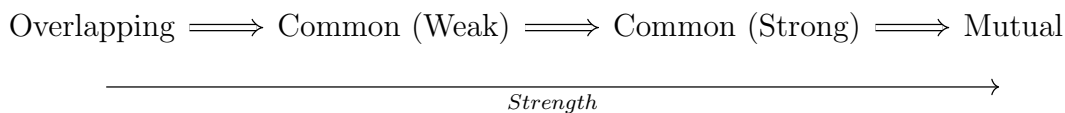


Figure 3.1: Levels of Joint Attention

tion of attentional states I discuss can be seen in Figure 3.1 and roughly follows the typology proposed by Siposova and Carpenter (2019). Stronger kinds of joint attention build on the weaker ones. When two people are in a state of strong common attention, for example, this entails that they have also achieved weak common attention and overlapping attention.

The bulk of the arguments in this section are negative. I will argue that the first three types of joint attention are insufficient for conversational salience. For each, I present a case in which two conversational participants achieve some level of joint attention toward something, but nevertheless that thing fails to be salient in their conversation. This failure is demonstrated by the infelicity of a sentence with a pronoun. Importantly, my arguments make no assumptions about what the infelicity is owed to. One might take these cases to be complete failures of reference, or more conservatively take them to be cases of technically successful reference but failures of coordination. But whether the failure is of reference or of coordination, the resulting infelicity is ultimately due to the intended referent’s lack of conversational salience, despite its achieving some level of joint attention. No such counterexample, I will claim, can be given for mutual attention, the fourth and strongest kind of joint attention.

3.3.1 Overlapping Attention

The natural starting point is the weakest form of joint attention:

Overlapping Attention

The attentional states of all interlocutors are directed at some particular thing:

A and *B* each individually attend to *X*.

When two subjects, *A* and *B*, have (merely) overlapping attention on *X*, they achieve joint attention in virtue of the co-reference of their individual attentional states. The *jointness* is achieved thanks to what the object is that they each attend to, not because of anything that internally distinguishes between, say, *A*'s state of individually attending to *X* and *A*'s state of jointly attending to *X* with *B*. The jointness of their attentional states can therefore be thought of as *accidental*.

It is easy to see why overlapping attention is too weak to be an account of conversational salience. The hypothesis that something is conversationally salient in virtue of its having been the object of overlapping attention is subject to counterexamples like the following:

- (52) *Lalo and Noemí are judging a chili cook-off. Both of them have just finished sampling the first entry, Chili A. They each are focusing on A's flavor profile. But they were out of each other's view, and neither knows which chili, if any, the other has tried. Lalo notices Noemí and says to her:*

#It's not spicy enough.

Clearly, Lalo's attempt to use the pronoun 'it' to refer to Chili A is infelicitous. For all Lalo knows, Noemí has started with a different Chili, has already tasted multiple entries, or has not started at all. The same can be said of what Noemí knows about Lalo. The problem, intuitively, is that even though it is in fact true that Lalo and Noemí have achieved overlapping attention, neither is in a position to know that this is the case. From each perspective, all that is known is what they themselves are attending to. Because of this, neither is in a position to coordinate with the other in the appropriate way—internally they are in the same position as they would be if they were not in a state of joint attention at all. Chili A is not conversationally salient in (52), and it seems that this is because of what the interlocutors do not know.

3.3.2 Weak Common Attention

Overlapping attention is not a serious contender for an explication of what it is for something to be conversationally salient. After all, one may have overlapping attention with another and not even know that the other exists, since all that is required is co-referring psychological states. Instead, one might take a simple form of **Common Attention** to be a more viable possibility:

Weak Common Attention

It is known to all interlocutors that their attentional states are directed at some particular thing: *A* and *B* each know that their attention overlaps on *X*.

A and *B* achieve weak common attention when they have overlapping attentional states and they each know this to be the case. What made overlapping attention a form of *joint* attention is now recognized, and this recognition can play a role in interpretation (be it semantic or pragmatic). This is therefore a more promising candidate for a joint psychological state that underwrites conversational salience. But strengthening the joint attention achieved by Lalo and Noemí from overlapping to weak common attention is still insufficient for making Chili A conversationally salient:

- (53) *Both Lalo and Noemí have tried Chili A. They have each been told by a third party which chili the other has tried. But neither knows that the other has been told. Lalo notices Noemí and says:*

#It's not spicy enough.

Just as in (52), Lalo's use of the pronoun 'it' is pragmatically uncooperative at best and semantically incomplete at worst. As a matter of linguistic theory, the explanation is identical: Chili A is not conversationally salient. Both Lalo and Noemí are aware that they are jointly

attending to Chili A, but the reason why this weak common attention does not guarantee salience is that neither person knows that the other is aware of their joint attention. From each perspective, the other might not have any awareness of their joint attention. Unlike the issue with overlapping attention, which involved first-personal ignorance, the issue with weak common attention is second-personal ignorance: neither Lalo nor Noemí knows what the other knows. Because of this, full coordination on intended reference is impossible.

In labeling (53) and other cases with ‘#’ and deeming them infelicitous, I am not claiming that they would be wholly unintelligible in context. Some report that (53) is acceptable when the pronoun is interpreted as cataphoric; ‘it’ will pick up its referent from what Lalo says or does next. On such a reading, Lalo could (and indeed should) followup in some way that makes clear which chili he intends to talk about. Such a reading may be available for other cases in this section as well. I have claimed that these cases are infelicitous (and so demonstrate that some level of joint attention does not guarantee conversational salience), but strictly this is only through the lens of anaphoric interpretations of pronouns. I agree that cataphoric readings of these cases are possible.⁷ However even if Lalo’s use of ‘it’ is felicitous because it is cataphoric, the cataphora exploits the fact that the intended referent is not conversationally salient at the time of the pronoun’s utterance. On a cataphoric reading, both Lalo and Noemí know that there is no salient chili (yet) for ‘it’ to refer to. So the dialogue might naturally unfold as (54):

- (54) L: It’s not spicy enough.
N: Which one?
L: Chili A.

In this case, Noemí must initiate a conversational repair strategy. Intuitively, uses of cat-

⁷Whether cataphora is possible in these contexts seems to depend on how one imagines subtleties that I have left undetailed in the descriptions of the contexts, such as intonational cues, personality traits of Lalo and Noemí, and how one should conduct oneself when judging a chili cook-off.

aphora like Lalo’s are deployed with the knowledge that they will necessitate such a repair. The point is that, even on a charitable reading of Lalo’s assertion, the intended referent cannot be understood as conversationally salient at the time of utterance, despite the fact that some level of joint attention (in this case, weak common attention) has been achieved at that time.

3.3.3 Strong Common Attention

The intuitive problems with overlapping attention and weak common attention involve lacking information. In a state of merely overlapping attention, interlocutors lack information about what the other is attending to. In a state of merely weak common attention, interlocutors have that information but lack information about what the other knows about what they themselves are attending to. These might be considered the same problem—there is a fact about which level of joint attention has been achieved, but that fact is itself not recognized in a strong enough way. On this diagnosis, what prevents the right chili from being conversationally salient is the kind of knowledge Lalo and Noemí have (or rather, fail to have). **Strong Common Attention** is the next strongest kind of joint attention, and it remedies this problem:

Strong Common Attention

It is known to all interlocutors that they know that their attentional states are directed at some particular thing: *A* and *B* each know that they’re commonly attending to *X*.

When *A* and *B* achieve strong common attention, the fact that they are jointly attending to something becomes *common ground* in the Stalnakerian sense (1978; 2002; 2014)—they have *mutual knowledge* about their attentional states. That this knowledge is mutual means that it is iterative. Not only does each know what the other is attending to, but each also knows that the other has that knowledge, that very fact is also known to each, as is *that*

fact, and so on. Note that the definition does not specify that the common attention that is mutually known is weak. This is because of the nature of iterated knowledge. Strong common attention involves mutual knowledge of weak common attention, and this in turn entails mutual knowledge of strong common attention. There is no stronger kind of joint knowledge to have, and in this case it bottoms out in a fact about what each is attending to.

It is for this reason that one might take strong common attention to be the right kind of joint attention for explicating conversational salience, especially if coming from the Stalnakerian tradition. Strong common attention simply cannot suffer from the same kind of ignorance that disqualified overlapping and weak common attention: ignorance that prohibits complete coordination regarding what kind of joint attention has been achieved. At each of the weaker levels, uncertainty about which level of joint attention has been reached persists. But the iterated nature of mutual knowledge ensures that strong common attention achieves full recognition: two agents cannot strongly attend to something without mutually recognizing that they are doing so. It would seem then that any situation in which Lalo and Noemí achieve strong common attention would be one in which they have all that they need to coordinate on the intended referent of ‘it’. After all, if they each know that the other knows that the other knows (and so on) that they are each attending to Chili A, then there is plenty of information at hand to use in pragmatic reasoning.

I do not think that strong common attention suffices for establishing something as conversationally salient. To show why, I must modify the case.

(55) *Both Lalo and Noemí have tasted all the chili (A, B, and C) and are now deliberating. Noemí says:*

Chili B had a really unique taste. Its ingredients were so unusual.

Suddenly a passionate cook in the distance holds up Chili A and yells several times “My chili is best!”. The cook’s yelling is loud and distracting, and so demands to focus of all who are around. Nevertheless, Lalo and Noemí do not outwardly respond to the outburst in any way—they do not say anything, do not startle, and do not glance in the cook’s direction. Attempting to refer to the cook’s chili, Lalo says:

#It’s_A not spicy enough.

Lalo and Noemí in this case behave as true professionals, and so do not react to the interruption as others would. Though they have maintained their composure, it is understandable that their internal focus is now on the cook’s Chili A. But Lalo and Noemí’s achieving strong common attention is not enough to establish Chili A as conversationally salient. My argument depends on four claims, each of which is independently plausible.

First, in context, Lalo’s use of ‘it’ more naturally refers to Chili B. Prior to the cook’s outburst, the conversation has focused on Chili B. Despite how distracting the interruption is, Lalo and Noemí maintain eye contact, do not smirk, or otherwise outwardly indicate that the outburst has happened. Given the physical description of the discourse and its participants, the conversation proceeds in exactly the same way as it would had the outburst never occurred. Lalo’s claim using the pronoun would naturally then be interpreted as continuing discussion about Chili B, and would not indulge the distraction regarding Chili A.⁸

This observation makes clear why the notion at hand is a term of art in linguistic theory, and why I have been careful to call it *conversational salience*. There is a perfectly good use of

⁸Stojnić et al. (2017, p.528) make a similar point, and this case is inspired by one of theirs. They imagine a conversation that ignores Bill jumping up and down, and claim that a use of ‘he’ would by default be interpreted anaphorically and not deictically. I agree, but provide a slightly different explanation to theirs. Stojnić et al. use the case to argue that attentional states are not sufficient for establishing intended referents (*signals of attentional states are*). As I will argue shortly, signals of attentional states enable more complex attentional states. This is what enables a kind of joint attention to be sufficient for establishing conversational salience.

“salience” on which the cook’s outburst makes Chili A salient. But when their conversation is unfazed by the cook’s yelling, Lalo and Noemí’s failure to react prevents Chili A from becoming conversationally salient.

The second claim is that the cognitive attentional states of both Lalo and Noemí shift to Chili A thanks to the cook’s yelling and holding it in the air. Often one’s outward behavior is a guide to where their cognitive attention is focused, but it is of course imperfect. Our minds can wander, we do not react to every change in attention, and we can suppress external reactions even when they are natural. This is how Lalo and Noemí behave; they resist acknowledging the outburst. But this does not mean that their attention internally does not shift to Chili A when its cook holds it up and yells. Rather, one can easily imagine the individual cognitive attention of both Lalo and Noemí being drawn in the middle of their conversation. Being confronted with the cook’s outburst, it would be odd for the judges to not entertain the claim that Chili A is best, for its flavor profile to not be called to mind, and for them to not be reminded of their previous comments about Chili A.

The third claim for my argument is that this previous claim (claim two) would be obvious to Lalo and Noemí in context. That is, given the nature of the outburst, not only is it very natural for one’s cognitive attention to be drawn, but everybody around would know that everybody’s attention has been so drawn. The following would be implausible in (55): thanks to the cook’s outburst, Lalo and Noemí each have their attention drawn to Chili A, but they each are uncertain as to whether the other noticed the outburst or whether the other attention was also drawn. This is not to say that this kind of pattern is impossible. While you and I are talking, I might hear a distant sound, but be uncertain that you heard it. But on the natural understanding of (55), the cook’s outburst is so disruptive and obvious, that neither conversationalist could be uncertain about whether the other realized that it took place.

The cook’s outburst is what Stalnaker (1978, 2014) calls a *manifest event*—an event that is so manifestly obvious that observers cannot help but mutually believe that it has taken

place. I will discuss manifest events in more detail in the next section, but for now the point is merely that they are standardly recognized in linguistic theorizing.

The status of the outburst as a manifest event is what justifies the claim that Lalo and Noemí each know that their attention has shifted to Chili A. It also, for the same reasons, justifies the fourth and final claim for my argument: the knowledge of the attentional shift is itself known to Lalo and Noemí. Because the event so manifestly obvious, not only would all in the vicinity know that everybody's attention had been drawn, but all would know that everybody would know about the shift in collective attention. This is what separates manifest events from events that are merely commonly known. A manifest event is plainly obvious, and its taking place automatically brings mutual knowledge of its having taken place.

Neither the third nor fourth claim requires that Lalo and Noemí consciously entertain the thought that the other knows about and is attending to the outburst. Mutual knowledge of a proposition (in this case, the proposition that *Lalo and Noemí are commonly attending to Chili A*) entails mutual knowledge of other propositions (e.g. Noemí's knowing that *Lalo knows that Noemí knows that Lalo and Noemí are commonly attending to Chili A*). But, following Stalnaker, this does not require a conscious act of accepting each proposition. Instead, it is enough that a single kind of propositional attitude is held by each conversational participant.

Presented more concisely, the four claims about (55) for my argument are:

- (i) Lalo's use of 'it' is naturally taken to refer to Chili B. His attempt to refer to Chili A is infelicitous.
- (ii) Lalo and Noemí, as a result the outburst, each cognitively attend to Chili A.
- (iii) Lalo and Noemí each know (ii).
- (iv) Lalo and Noemí each know (iii).

Given the test I have been using for conversational salience (referential success), Claim (i) entails that Chili A is not conversationally salient at the time of Lalo’s utterance. Claim (ii) means that Lalo and Noemí achieve overlapping attention. The conjunction of Claims (ii-iii) entails that Lalo and Noemí achieved weak common attention on Chili—their overlapping attention is known to each of them. Those conjoined with Claim (iv) entail that Lalo and Noemí’s common attention is in fact *strong*—their weak common attention is known, which is enough to generate iterative common knowledge about their joint attention. So despite their strong common attention, Chili A fails to be salient in their conversation. Therefore, strong common attention is not sufficient for conversational salience.

3.3.4 Mutual Attention

Despite the insufficiency of even strong common attention, I will argue that there is a kind of joint attention that does establish something as conversationally salient. To do so, I will begin here by showing that the next level of joint attention, **mutual attention**, is not subject to the same kind of counterexamples that make trouble for the weaker levels. The remainder of the chapter will then focus on (a) explaining why mutual attention is sufficient while strong common attention is not, and (b) exploring the ways in which mutual attention factors into conversation.

Mutual Attention

It is known to all interlocutors that they know that their attentional states are directed at some particular thing, and the attentional state of each depends on the signals of the attentional states of all others: *A* and *B* attend to *X* together.

To mutually attend to something with another is to, in part, achieve strong common attention with them. But what separates mutual attention from strong common attention is that the former requires that the individual attentional states depend on the (signals of) the attentional states of the others. This does not mean that for *A* and *B* to mutually attend to

X , the fact that A attends to X causally depends on B 's attending to X . Rather, it means that the way in which A attends to X causally depends on the way in which B attends to X , and vice-versa. That is, mutual attention does not require that A would not have attended to X without B 's having attended to it. After all, X may be a very loud noise. Instead, the fact that A and B mutually attend to X means that the kind of attention they each engage in depends on the kind of attention the other engages in.

By “kind of attention” or “way in which one attends”, I do not mean which level of joint attention one engages in with another. Otherwise, merely achieving overlapping attention with another would then trivially produce a state of mutual attention (since A 's overlapping attention depends on B 's attentional state). The “way of attending” that is relevant for mutual attention involves what Siposova and Carpenter (2019) characterize as the *perspective* of one's attentional state.

To jointly attend to something with another through a *third-person* perspective is to engage with their attentional state indirectly. The prototypical way to do this is through inference. In (55), the knowledge that each judge has about what the other knows and is attending to has nothing to do with how the other is acting. Instead, Lalo and Noemí each know that the other has noticed the cook's outburst because it would be impossible to not. Their engagement with the other about what happened is no different from their engagement with some third party nearby who is not taking part in the conversation.

On the other hand, what characterizes mutual attention is engagement through a *second-person* perspective. This kind of engagement is direct— A 's engagement with B 's attentional state is owed to A 's observing B 's attentional signals and vice-versa. The degree of detachment that is possible with strong common attention is not possible with mutual attention. Whereas the jointness of strong common attention is owed to layers of knowledge about separate attentional states, the jointness of mutual attention is owed to an act of attending together.

A second-person perspective is reciprocal in that the engagement must be two-way (or

more, in cases of multi-party mutual attention). This requires what Gómez (2005) characterizes as *attention contact*: attention of another’s attentional state. Attention contact can be established in a number of ways, including eye contact, touch, and speech. This is what is absent from (55); Lalo and Noemí fail to make attention contact regarding their (strong common) attention on Chili A.⁹ Without this attention contact, a second-person perspective is impossible. Without a second-person perspective, the judge’s attentional states are not interdependent.

For one’s attentional state to take a second-person perspective is not merely for it to be ushered by attention contact. A second-person perspective brings with it a phenomenal difference as well. The difference between how a second-person perspective feels in contrast with how a third-person perspective feels is described as a difference in the lens through which one perceives another party. To take a third-person perspective is to perceive another as *he/she/they*, whereas taking a second-perspective involves representing them as *you* (Reddy 1996; Reddy and Morris 2004).

The reason why strong common attention is insufficient for conversational salience, and the reason why mutual attention is necessary, becomes vivid when (55) is modified to allow for attention contact:

(56) *Both Lalo and Noemí have tasted all the chili (A, B, and C) and are now deliberating. Noemí says:*

Chili B had a really unique taste. Its ingredients were so unusual.

Suddenly a passionate cook in the distance holds up Chili A and yells several times “My chili is best!”. The cook’s yelling is loud and distracting, and so de-

⁹Furthermore, Lalo and Noemí do make attention contact regarding their attention on Chili C, given the conversation prior to the outburst.

mands to focus of all who are around. Startled, Lalo and Noemí glance in the cook's direction. Lalo says:

It's_A not spicy enough.

Lalo's referential intention intuitively succeeds. This is thanks to the attention contact that is achieved by way of looking at the cook and Chili A. The judge's externally signalling where their individual attention is focused allows for the other to directly engage second-personally with them. Lalo and Noemí can then attend together to the outburst and Chili A. This way of attending for each depends on the other's attentional state, and so the judges achieve mutual attention. The fact that the contextual change between (55) and (56) both enables mutual attention and referential success is evidence that mutual attention is the kind of joint psychological state that is relevant for conversational salience.

This does not entail that mutual attention is necessary for something to be conversationally salient. At most, this is an argument that it is sufficient. In the next section, I defend the position that mutual attention is in fact necessary. To do so, I focus on comparing it to its natural alternative: mutual knowledge. Following this, I explore in more detail what it means to say that something is conversationally salient if and only if it has been mutually attended to during the conversation.

3.4 Knowledge of Attention

In the previous section I took as a starting point the intuition that conversational salience is a property that holds in virtue of the interlocutors' attentional states. I argued that all but one kind of joint attention fail to guarantee conversational salience. One might take issue with this starting point and argue that attention is the wrong kind of psychological state for underlying conversational salience. Instead, one might claim, mutual knowledge is the

right candidate. The fact that there are knowledge requirements on states of joint attention might even indicate that knowledge is what's important.

Views that ground communication in mutual knowledge have a strong tradition. Grice (1957) articulated a theory of non-natural meaning on which utterances have content in virtue of being produced with the right kind of communicative intention. One must intend, among other things, that their utterance be recognized as the product of the communicative intention that in fact produced it. As Stalnaker (2014) emphasizes, such complex intentions can only be performed when the speaker and hearer have sufficient *common ground* between them.

The upshot for our purposes is that common interest and common knowledge are necessary for the possibility of communication. Only against a relatively rich background of common belief is it possible to get people to recognize the very specific intentions that must be recognized for successful acts of meaning, and only where there are mutually recognized common interests will the recognition of the intentions be effective in changing beliefs. (Stalnaker 2014, p.42)

Stalnaker conceives of the common ground—the set of propositions mutually accepted by two or more agents—as the state that enables communication in the first place as well as the state that communication functions to act on.

Whereas Stalnaker's later work (e.g. 2002; 2014) is careful to define the common ground in terms of a complex family of propositional attitudes (all under the umbrella of *acceptance*), the Gricean arguments regarding the grounding of communication directly involve mutual knowledge. Arguments along these lines have been influentially argued for by Clark (1996), Clark and Brennan (1991), Clark and Marshall (1981), and Schiffer (1973).

This perspective does not directly conflict with the position I have been arguing for. But it naturally lends itself to a collection of views that do. Believing that mutual knowledge is a prerequisite for the possibility of communication, one might argue that conversational

saliency reduces to facts about what is mutually known/accepted (like facts about conversation generally). On such a position, attentional states are important (perhaps causally), but they do not ground any aspect of communication. It is not attention that is key, it is what we know together about what we are attending to.

A knowledge-first view of saliency could take two forms, each each of which conflicting with my position in a slightly different way. They are:

- a) Mutual attention is not necessary, strong common attention is. The purported counter-example fails.
- b) Mutual attention is necessary, but it is just a special kind of mutual knowledge.

According to each version, what ultimately underlies conversational saliency is a kind of mutual knowledge, and to focus on joint attention is to miss what is crucial in coordinating with another through conversation. I will argue that each version of the knowledge-first view fails to be as adequate as the attention-first view.

3.4.1 Saliency in the Common Ground

The Stalnakerian common ground is modelled as the set of propositions (sets of possible worlds) that are the objects of certain propositional attitudes of all conversational participants: $CG = \{P : \text{all interlocutors mutually accept } P\}$. Its standard use is in defining the *context set*, which represents the possibility space taken to be live in the conversation: $CS = \bigcap CG$. This is useful because it provides an abstract and technical way of understanding the speech act of assertion. A successful assertion of Q adds the proposition to the common ground, $CG \cup \{Q\}$ and thereby shrinks the possibility space that is the context set to include only worlds in which Q is true, $CS \cap Q$.

As simple as it is, this traditional gloss on the theoretical role of the common ground makes it easy to forget that the common ground's core value (and indeed the reason why

it is said to explain assertion in the way that it does) is in representing the totality of how conversationalists coordinate. Not only is general background knowledge (such as facts about what year it is or which mutual friend recently got married) included, but so is “foundational” knowledge: all the participants speak the same language, “cat” means *cat*, assertions purport to convey knowledge, the current topic is *whether golf is boring*, for example (see Stalnaker 2014, Sec. 2.5).

Taking seriously Stalnaker’s suggestion that foundational knowledge about language, communication, and the state of the conversation be represented in the common ground along with more general background knowledge, one might hypothesize that conversational salience is a phenomenon that must reduce to facts about what is accepted as common ground. Even if attention qua psychological state is relevant, ultimately what should matter is whether the common ground contains some proposition (or multiple propositions) about attention. In the cases above, such a proposition might be *You and I are both attending to Chili A*.

Such a position would need to make sense of the intuitions regarding (55), which I claimed demonstrated the insufficiency of strong common attention for establishing something as conversationally salient. One might deny the judgment about the case and claim that Lalo’s assertion is perfectly felicitous. Alternatively, one might offer a different explanation of the assertion’s infelicity. I am skeptical that either path would be convincing. Nevertheless, there are cases that *prima facie* favor this knowledge-first position (on which strong common attention is sufficient). One might offer the following case, inspired by a case from Stalnaker:

(57) *Victoria and Rigo are having a conversation indoors, when suddenly the lights go out. It is so dark that nothing is visible. Victoria says:*

Ugh, we do *not* need this right now.

Without question, Victoria's use of 'this' refers to the event of the lights going out. A Stalnakerian might argue that this cannot be explained by my view. After all, in the dark no eye contact, glances, or the like can be seen to establish attention contact. This means that the kind of joint attention achieved by Victoria and Rigo cannot be mutual attention—one's attention on the lights-out event doesn't depend on the attentional state of the other. In contrast, it seems that they do achieve strong common attention. Their attention overlaps and the event was so manifestly obvious that each can infer an iterated series of propositions regarding their common attention.

This objection's description of which level of joint attention is achievable in (57) is correct. In fact, Siposova and Carpenter (2019) use an example of lights going out on an airplane to contrast common and mutual attention:

Example 1 [of common attention]: Suddenly, the lights go off in the plane and Allison and Rob find themselves in darkness. Allison is attending to the fact that it is dark, to the fact that Rob must also be perceiving the darkness, and to the fact that each knows that they each know this, and Rob is engaging in the same attentional processes. (p.262)

Example 1 [of mutual attention]: When the lights go off in the plane, Allison and Rob grasp each other's hand reflexively. (p.263)

It should be noted first that an example like (57), in which reference using a demonstrative is possible in the absence mutual attention on the intended referent, strictly does not conflict with the view I have been arguing for. I have been arguing that something is conversationally salient if and only if it has been mutually attended to during the conversation. But I have not claimed that conversational salience is always necessary for successful reference. I have used referential failure as evidence for the absence of salience, since salience tends to guarantee coordination on intended reference. But even the strongest version of that view only requires that conversational salience be sufficient for referential success, not necessary.

But this is a minor point, and a larger explanation is called for. From the perspective of an attention-first view of salience, what is to be said of (57)?

I have so far focused on how attention impacts conversation and said little about the other way around. But cases like (57) make for opportunities to discuss how speech acts guide attention. On my view, Victoria and Rigo achieve strong common attention as a result of the lights-out event, but Victoria's speech act establishes mutual attention. So (57) is a rare case in which demonstrative reference succeeds without conversational salience (because of the lack of mutual attention before the time of utterance). Nevertheless, mutual attention is not uninvolved.

What utility is there in theorizing that mutual attention is achieved thanks to the speech act? I will go into more detail in the next section about attention management in conversation, but note the following data. Victoria's use of the demonstrative is clearly felicitous, whereas it would be unacceptable to use the pronoun 'it':

(58) #Ugh, we do *not* need it right now.

Yet it would be perfectly natural for Rigo to respond to (57) with a sentence containing the pronoun:

(59) *V*: Ugh, we do *not* need this right now.

R: Oh come on, it isn't that bad.

There is of course a linguistic explanation to be offered for why Rigo's use of 'it' is acceptable while Victoria's is not; the anaphor 'it' requires a linguistic antecedent, whereas the demonstrative 'this' does not.¹⁰ In (59) Victoria provides the antecedent (using the demonstrative), but no such antecedent exists in (58), and so it is infelicitous.

¹⁰This is a generalization. There are demonstrative uses of 'it'.

But understanding conversational salience through the lens of mutual attention helps to provide a more abstract explanation. An anaphor is sensitive (semantically or pragmatically, depending on your theory) to conversational salience, whereas demonstratives are less so. This is not arbitrary—demonstratives *demonstrate*. They can function to coordinate agents in a state of mutual attention. This establishes salience, which can then be exploited by anaphors like ‘it’.

This is the kind of explanation that the knowledge-first view of salience fails to provide. If strong common attention suffices for conversational salience, then it makes sense why ‘this’ succeeds in referring, but (a) conversational salience is less linguistically explanatory (since anaphors are sensitive to *what has been mentioned* as opposed to *what is salient*), and (b) the in principle difference between demonstratives and anaphors is lost.

Recall that in introducing conversational salience, a simple example of its use in linguistic analysis involved available anaphoric reference to the lone missing marble in (46) and (47). Put simply, the point here is that the knowledge-first view of salience that relies on strong common attention weakens the role of conversational salience so that it no longer distinguishes between (46) and (47), as the missing marble would be the object of strong common attention in both cases.

3.4.2 A Privileged Common Ground

One need not endorse strong common attention as an explication of salience to have a knowledge-first view. One might agree with my reported judgments on all of the above cases and even generally agree with my arguments in favor of mutual attention, but maintain that mine is actually a knowledge-first view in disguise. To achieve mutual attention with another is to coordinate your mental states in a particular way. That way involves making attention contact and jointly attending through a second-person perspective. But the end result, one might argue, is simply a state of mutual knowledge.

It is easy to see why this kind of position is available with respect to strong common attention. Its definition straightforwardly involves mutual knowledge about a lower level of joint attention (weak common attention, which straightforwardly involves knowledge about overlapping attention). This is what makes the position discussed above a knowledge-first position.

The thought here, though, is that the requirements on mutual attention that separate it from strong common attention are merely requirements on the *source* of the relevant propositional knowledge. So while the mutual knowledge that is relevant for establishing conversational salience cannot come from anywhere (specifically, it cannot be inferred indirectly), it is still mutual knowledge that in fact establishes something as salient in a conversation. When Chili A is successfully established as salient in (56), thereby enabling Lalo's successful reference, the proposition *Lalo and Noemí are both attending to Chili A* is mutually known to Lalo and Noemí in the right way and from the right evidence (attention contact).

In effect, this position defines a privileged common ground, a subset of the Stalnakerian common ground that contains all and only those propositions mutually known thanks to attention contact. So, on this view, while my arguments may provide insight into the specific ways in which mutual knowledge can impact conversation, they do not actually succeed in explicating conversational salience in terms of psychological attention (only knowledge of attention).

This is an internally coherent view, but it misunderstands what is at the core of mutual attention. To mutually attend with another is not to merely mutually know about each other's attentional states thanks to the outward cues each provides. Instead, one's attentional state must depend on the other's. The attention contact necessary for this kind of dependence involves attending both to the object of mutual attention but also attending to the other's act of attention (Gómez 1996). Insofar as one takes attention to yield perceptual knowledge (see Peacocke 2005 for a discussion of issues) then my view might entail that mutual knowledge from a particular source is required for conversational salience. But

that knowledge is not itself constitutive of conversational salience. It is a byproduct of the psychological, attentional, state that is.

These attentional states, unlike mutual knowledge, are not “meta-representations or second-order representations (i.e. representations of other subjects holding internal representations separable from their behaviour), but practical, first-order representations” achievable by infants and apes (Gómez 2005, p.66).

It is this idea that underlies the second-person quality of mutual attention. Lower levels of joint attention are joint only inasmuch as they involve particular knowledge. But to achieve mutual attention, one attends with another.

The upshot is this. It is possible, but not obvious, that mutual knowledge of a state of joint attention may be present in all cases of mutual attention, and so present in all cases of conversational salience. But even if that is the case, it would still be wrong to say that something achieves conversational salience thanks to the mutual knowledge. Instead, both mutual knowledge and conversational salience are achieved thanks to the state of mutual attention.

So to flesh out what the psychological notion is in virtue of which something achieves the linguistic status of *conversationally salient*, one must appeal to attention, not knowledge. Specifically, mutual attention is the key notion.

3.5 Conversation as Attention Management

My aim has been to argue for an understanding of conversational salience in terms of mutual attention. To do so, I have focused on cases in which the contexts fix the attentional states of the interlocutors and proceeded to argue on the basis of acceptability judgments.

Facts about joint attention, then, impact facts about which conversational moves are felicitous. It should therefore not be surprising that much of conversation functions to

manage the attentional states of the participants. That is, given that attention can guide conversational moves, it makes sense that one way for conversation to proceed is by managing attention—so that future moves are guided appropriately.

Deictic gestures make for a particularly simple example of this. Gestures function to guide attention. More specifically, they function to coordinate conversational participants in a state of mutual attention. This is so that something can be strategically made conversationally salient, depending on the communicative aims of the one performing the gesture. Sometimes the aim is simple: to fix the reference of a demonstrative linguistic expression such as *that*.

Rich demonstration, the phenomenon discussed throughout Chapters 1–2, also illustrates the communicative function of managing attention in conversation. When one richly demonstrates, they deliberately establish mutual attention with another and thereby make something conversationally salient. Thanks to specific dynamic discourse mechanisms discussed in Chapter 2, the speech act effect argued for in Chapter 1 can be seamlessly achieved.

It is not just gestures that are used to manage attention, and it is not just situations from a conversation’s physical context that are attended to. It is cognitive attention that is important for conversation, and cognitive attention can be managed linguistically (O’Madagain and Tomasello 2019). The use of referential NPs, for example, make for a straightforward form of attention contact that establish mutual attention on the referent. This in turn has consequences for anaphora, future pragmatic inferences, and more. This is why, for example, in (55) and (56), Noemí’s mentioning Chili B coordinates her and Lalo’s attention so that it is conversationally salient.

The importance of conversational salience in linguistic theory is difficult to overstate. Regardless of whether one conceives of its role in interpretation as semantic, pragmatic, or both, it is clear that something’s being salient in a conversation is important to how that conversation may proceed and that discourse moves take advantage of that fact by managing how interlocutors coordinate on what is salient. I have argued that mutual attention is the psychological notion that is key for understanding conversational salience. Given this, I take

myself to have not only provided an elucidation of a particular theoretical concept, but also to have provided insight into a core function that communicative moves have, to manage attention.

CHAPTER 4

Common Ground and the Structure of Context

The Common Ground (*CG*) is a standard tool in Linguistics and Philosophy for modeling and explaining discourse-level communicative phenomena. Thanks to the influential work of Stalnaker (1973, 1978, 2002, 2014), it remains a useful resource for capturing the effects of assertion, presupposition (failure/accommodation), the at-issue/not-at-issue content distinction, and the impact of world knowledge on conversation generally. Despite this popularity, *CG* is often treated as only one of several mechanisms that operate in discourse, incapable of serving as the root explanatory tool at the semantics/pragmatics interface. Such treatments posit extra structure in discourse, and in so doing apparently run contrary to various arguments from Stalnaker (1998, 2002, 2014), which emphasize the philosophical import of a discourse context's *unstructured* nature.

Structured treatments vary in the kinds of structure they posit. For example, dynamic semantic accounts (e.g. Groenendijk and Stokhof 1991; Heim 1982; Kamp and Reyle 1993) are distinctive in that they identify the content of a conversational move with its characteristic effect on discourse (as opposed to its informational meaning), and so require that discourses be sensitive to various kinds of change.¹ Other accounts, broadly in the tradition of Lewis (1979), are distinctive in that they conceive of discourse contexts as composed of different types of information structure (e.g. Farkas and Bruce 2010; Portner 2004, 2007; Roberts 2012–1998).²

¹See Rothschild and Yalcin (2015, 2016) for in-depth discussions of this.

²I will mostly characterize structured approaches in this latter way for simplicity.

The purpose of this chapter is to investigate the tension between structured and unstructured approaches to discourse. I will argue that an entirely unstructured theory—i.e. an entirely *CG*-based theory—is untenable. There is theoretical utility to conceiving of discourses as genuinely structured when studying high-level communicative phenomena. Given how I conceive of an entirely unstructured theory, the considerations I offer will be familiar ones in current formal semantics and pragmatics. My real focus is on a deflationary view of discourse structure—a view I take to be tacitly endorsed by many theorists—on which apparently structural aspects of discourse are reducible to the propositional attitudes of a conversation’s participants. The details of the deflationary view have not been discussed at length, largely because it is a philosophical position about the nature of formal models. I attempt to work through the details, and argue that the deflationary view requires a notion of *pseudo-structure* brought on by complex context-referential propositions. I will argue that those propositions cannot plausibly be the objects of propositional attitudes normally taken to define the *CG*.

I begin in Sec. 4.1 by describing in more detail the tension between structured and unstructured analyses, along the way emphasizing the natural appeal of structured approaches. I will then in Sec. 4.2 introduce my notion of *pseudo-structure*, which underpins the intuitive deflationary view of discourse structure mentioned above. In Sec. 4.3-4.4 I develop my argument against the deflationary view by focusing on the nature of the propositions it requires that discourse participants accept. I argue in Sec. 4.3 that, to explain apparent structure, an unstructured approach to context must appeal to a misguided order of explanation of why facts about discourse contexts are true. In Sec. 4.4 I discuss the practical and theoretical benefits of taking on genuinely structured view of context.

4.1 How Common Ground is Used

Conversation crucially involves coordination between two or more people. When we talk, our back and forth adds to the store of things on which we've coordinated. When we compare our opinions about campy movies from the 90's or when we learn about the other's favorite memories from college, we coordinate more. We coordinate on more *facts*. Some facts that previously only one of us knew become facts that both of us know. The Common Ground (*CG*) is the theorist's tool for capturing this central aspect of conversation. The *CG* is just the set of established facts in a discourse. After letting you know that I ate pizza last night, that fact about my dinner becomes part of the *CG* of our conversation. It becomes another piece of background information, joining the other things we've said to one another in the conversation. Included in that background information are many facts we have not explicitly discussed but nevertheless mutually know—that there are seven days in a week, a mutual friend is having a child soon, and it's been uncharacteristically rainy recently. More carefully: as conversation proceeds, the collection of propositions that we *mutually accept* grows. Mutual acceptance is an iterative notion. A proposition is mutually accepted by two people if and only if it is accepted by each, the fact that it is accepted by each is itself accepted by each, *that* fact is accepted by each, and so on.³

The utility of *CG* does not merely have to do with its intuitive appeal. It serves as a predictive modeling tool as well. A famous application is to the phenomenon of presupposition. If one says “Since it's raining, I'm mad that I forgot my umbrella”, while holding their umbrella on a sunny day, what they've said is conversationally marked (infelicitous) and cannot obviously given a truth value. What they say presupposes that it is raining and that they forgot their umbrella, and asserts that they're mad about the latter fact.⁴

³This of course glosses over many detailed debates. For example whether the *CG*'s iterative nature is technically viable (Lederman 2018), and whether communicative intentions need to be context-directed (Harris 2019), among others.

⁴Some, including Stalnaker, offer substantive arguments about whether it is speakers or sentences that presuppose things. Since this is a simple overview, I gloss over those debates and drift between both ways

This pattern—presuppositions being preconditions for the felicity of an utterance—is handled nicely in a *CG* framework. A sentence *S* that presupposes *P* is felicitous only if *P* is among the propositions mutually accepted as *CG*. And so the core intuitions regarding presupposition that trace back to Frege (1892) and Strawson (1950) can be preserved.

Often, though, when a presupposition has not been established, the assertion remains successful. In those cases, the presupposition is *accommodated* into the conversation. When *P* is not already in the *CG*, its accommodation can be captured by forcing the *CG* to update with *P*, which sets the stage for the assertion’s main (at-issue) content to be accepted or rejected. In a situation where you’ve asserted that it’s raining, and I followup with the sentence above, we can say that the raining presupposition is satisfied, that the presupposition that I forgot my umbrella is automatically accommodated, and that my assertion about my emotional state is slated to be added to or rejected from the *CG* as usual.

Generally speaking, these kinds of uses for and glosses of the *CG* are uncontroversially accepted. Where we see variation is in whether models of discourse posit additional mechanisms to work alongside the *CG*. If an analysis does posit such mechanisms, I count it as *structured*. Before elaborating on what this means, I can now give a gloss on what it is for an analysis to make use of an *unstructured* context:

Discourse Context

The information state (or set of information states) that is (or are) updated by conversational moves.

Unstructured Context

A Context with a single information state, the Common Ground, which contains one informational type that is updated in exactly one way.

On an unstructured picture, every conversational move has the exact same type of effect

of speaking.

on the exact same information state. At bottom, conversation is fundamentally a matter of information exchange. So every meaningful communicative act is understood as an act of exchanging information. A simple (and standard) formalism will help to anchor these ideas:

Paradigm Unstructured Context

A Discourse $D =$ the Common Ground (CG) such that:

- $CG = \{P : P \text{ is mutually accepted by all interlocutors}\}$

Views of discourse with the CG at their center frame the formal effects as consequences of how the CG is defined. These involve on naive set-theoretic operations on the CG to generate predictions, relying on the *Context Set* (CS), defined in terms of the CG . $CS = \bigcap CG$, representing all the worlds currently taken as live possibilities in the conversation. To successfully assert a sentence is to cause everybody in the conversation to mutually accept it. This means its content is added to the CG , thereby eliminating the newly incompatible worlds from the CS : $D[A(S)] = CG \cup \{\llbracket S \rrbracket\} \equiv CS \cap \llbracket S \rrbracket$. A consequence of this view is that if at two times the CG is the same ($CG_{t_1} = CG_{t_2}$), no communication whatsoever has taken place.

A natural worry one might have about this picture involves moves that don't apparently contribute any information—communicative acts that don't obviously increase our stock of shared knowledge. Non-assertive speech acts such as questions and commands are natural candidates. I will focus on the former. To put the matter simply: to ask is not to inform. Roughly, the issue is that the CG is designed for a purely assertive conversations: I inform you of something, then you inform me of something, and so on. But questions are a different type of speech act; they don't primarily serve to increase our shared knowledge. More specifically, in the tradition following Hamblin (1973), questions denote sets of propositions—their potential answers: $\llbracket Q \rrbracket = \{P_1, P_2, P_3\}$. This means that questions are formally of the same type as the CG itself, so it is difficult at first glance to imagine how they could be

captured within it. Importantly, questions that have been asked guide discourse in a tangible, truth-conditional way:

- (60) a. What ethics have you read?
b. I've only read Kant's *Groundwork*.
 \rightsquigarrow Haven't read J.S. Mill's *Utilitarianism*.

- (61) a. What Kant have you read?
b. I've only read Kant's *Groundwork*.
 $\not\rightarrow$ Haven't read J.S. Mill's *Utilitarianism*.

The same answer in (60) and (61) has different discourse effects in each case. That is, the latter should eliminate worlds as live possibilities from the *CS* that the former should not. This suggests that questions, despite being of a different informational type than assertions, are tracked in discourse. To make matters worse, there is an extensive literature on the hierarchical structure of questions in discourse, and on how both explicit and implicit questions serve to guide the flow of conversation. This research on Questions Under Discussion (QUDs). Roberts (2012–1998) aims to model various communicative strategies that speakers employ and how they are seamlessly integrated into a conversation. In order to get the answer to a question, a speaker may ask a sub-question—“What Kant have you read? Have you read the first Critique?”. Questions and sub-questions form a hierarchical structure that bears on which assertions are acceptable in a discourse. Answering an un-articulated sub-question may count as a felicitous move, but answering a different question altogether is a non-sequitur.

With questions on the table, we are now in a position to give a gloss on structured

discourse contexts:

Structured Context

A Context with more than one type of information state such that different types of move may update different states.

A very simple example of this that takes the above observations about questions into account may be as follows:

Simple Structured Discourse

A Discourse is a double $D = \langle CG, Q \rangle$ such that:

- $CG = \{P : P \text{ is mutually accepted by all interlocutors}\}$
- Q is ordered stack of questions/sub-questions:
 $Q = \langle q_1 \dots q_n \rangle$:

With this kind of discourse available, one could define different types of update for different speech acts. To ask a question $?(S)$ is to add a new element to Q . To make an assertion $A(S)$ is to add new knowledge to the conversation, and to potentially resolve questions.

Simple Structured Update

- $D_1[?(S)] = D_2 :$
 - $CG_2 = CG_1$
 - $Q_2 = \text{push}(\llbracket S \rrbracket, Q_1)$
- $D_1[A(S)] = D_2 :$
 - $CG_2 = CG_1 \cup \{\llbracket S \rrbracket\}$

- $Q_2 = \text{pop}(q, Q_1)$ for any $q \in Q_1$ s.t. $\exists P \in q : CS_2 \subseteq P^5$

Representations of discourse context as structured in this way allow for semantic content to be sensitive to various streams of information in a desirable way. For example, the denotation of “only” could be sensitive to the question at the top of Q , providing the necessary kind of contextual restriction on declarative sentences demonstrated by (60)-(61).

But there are reasons beyond the existence of non-assertive speech acts that motivate representing context as structured. Historically, patterns involving referential anaphora have pointed theorists toward representing lists of potentially available referents. Anaphora involves reference to something previously brought up in conversation. The most direct articulation of the potential problem for the *CG* approach comes from Heim (1982) and Kamp (1988), and historically tracks the general motivations for dynamic semantics. The idea is this: Different sentences with identical truth-conditional content license different anaphoric continuations. Without a structured division of information types, the contexts resulting from each will be identical, and so incorrectly predict that the same continuations are licensed. Kamp’s example of this is originally from Partee:

- (62) a. Exactly one of the ten balls is not in the bag.
 b. It is under the sofa.
- (63) a. Exactly nine of the ten balls are in the bag.
 b. #It is under the sofa.

(62a) and (63a) have identical truth-conditional content; they are true in exactly the same possible worlds, and so express the same proposition. So the updates to the *CG* are identical: $CG[(62a)] = CG[(63a)]$. Nevertheless, the same followup is only licensed by the former, and *CG* analyses cannot obviously explain why. Kamp characterizes the issue with the *CG*

⁵This condition removes questions from Q if an assertion answers them: for any q , remove it from Q if the open possibilities of the *CS* are all contained in one of q ’s possible answers.

approach as residing in Stalnaker’s view of propositions as sets of worlds. Heim similarly criticizes the possible worlds view of propositions, but connects the problem to contexts more explicitly. Her file-change semantics characterizes semantic content as a Context-Change Potential (CCP), where context has a more fine-grained structure than can be offered by the *CG*. In each case the repair involves enriching discourse contexts with more structure—discourse referents (drefs). The thought is this: In addition to tracking the truth conditional evolution of the discourse, interlocutors also track potential referents (Karttunen 1969). There are rules associated with particular linguistic expressions (like the indefinite NP “Exactly one of the ten balls” in (62a)) that introduce drefs for later anaphoric reference. If this is approximately correct, then there are systematically different kinds of possible update to a conversation’s context.

For the purposes of this chapter, there is no need to go into great technical detail of how anaphora should be analyzed. Instead, it will suffice to create a simplified picture of discourse contexts that includes the right elements, to make explicit the different information states that seem to be tracked. The general picture that emerges from the above observation is that, minimally, a discourse *D* is a tuple $\langle CG, R \rangle$ consisting of the Common Ground *CG* and an ordering of potential referents for anaphoric expressions *R*. The difference between the two examples above is not in their *CG*s, but in their *R*’s; the use of the indefinite NP in (62a) adds a potential object for later reference, whereas the phrase in (63a) raises the 9-ball collection to salience, biasing it for future reference:

$D_{(62a)}$		$D_{(63a)}$	
<i>CG</i>	<i>R</i>	<i>CG</i>	<i>R</i>
$\{p_1, p_2, p_3\}$	$\langle \mathbf{1-ball} \rangle$	$\{p_1, p_2, p_3\}$	$\langle \mathbf{9-balls} \rangle$

The tension that I have attempted to draw attention to has to do with what the driving mechanism behind discourse is. For an unstructured view, the only core tool needed is the *CG*. For a structured view, more is needed. The examples I’ve chosen have to do with interrogative speech acts and anaphora.

To make the tension even clearer, note that strictly the above discourse snapshots violate the standard definition of *CG*. This is because, presumably, the status of *R* in each case is known (importantly: mutually accepted) by all discourse participants. So to strictly adhere to its definition, the *CG* will include propositions about the state of the conversation itself. When there are different objects available for reference, or when there are different questions currently salient, the discourse participants will have different knowledge about the discourse contexts, and so there will be different propositions in the respective *CG*s. Excluding these propositions from the *CG* in structured representations of discourse is understandable, and highlights an important potential departure from the traditional way of conceiving of *CG*. Whether this kind of re-conception is shorthand or something more substantial is explored in the next section.

4.2 Pseudo-Structure

The point of the discussion above was to provide a general overview of two perspectives on how the *CG* can be used in modeling discourse. One, the unstructured view, takes the *CG* to be the central mechanism. The other, the structured view, takes it to be one among many streams of information necessary to capture the nuances of everyday communication.

This section is devoted to explaining the Stalnakerian response to considerations in favor of structure. The aim is to articulate a kind of middle-ground, deflationary view of discourse structure, according to which there is no need to posit any *real* additional structure to a discourse context beyond the *CG*. This deflationary view relies on a technical point about the nature of propositions in the *CG*. The thought is that any prediction made by a structured analysis can be accommodated by a purely *CG*-model so long as the appropriate propositions are mutually recognized in the conversation.

Positions that accommodate data like those above with unstructured, purely *CG*, contexts posit what I will call *pseudo-structure*, which is what allows for the functional effects

of structured information without explicitly using any. That is, there are ways to posit a *CG* that permits the kinds of operations that are prioritized by approaches to discourse that resemble *D* above.

The general idea is this. There are substantive facts about the status of a discourse at any given time, such as what its currently open questions are and what objects are available for anaphoric reference. Those substantive facts about the status of a discourse context will be mutually accepted by all involved in the conversation, and so those facts (propositions) will be part of the conversation's *CG*. In virtue of being part of the *CG* (being mutually accepted), those propositions license the same kinds of actions and interpretations by the interlocutors that structural facts would. So there is no technical need for discourse structure.

An early defence of this approach in light the kind of data above is from Stalnaker (1998), who responds to the anaphora criticism from Kamp (1988). His response is that these data demand structured representations at the level of linguistic interpretation, not at the level of communication.

The abstract framework, by itself, says nothing about which particular facts make salient, or available for pronominal reference, an individual who inhabits the domains of the possible worlds that define the context. Among the facts that might be relevant are facts about the words used to express certain propositions in the discourse, since these are facts that are available to the participants in the discourse, and so are reflected in each of the possible worlds that are members of the relevant context set. [...] whatever the explanation, it will appeal to facts about English, or about the practice of speaking English, that can be presumed to be available to competent English speakers, and so they will be facts that distinguish the possible worlds that define the different contexts. (pp. 11–12)

Put differently: the structural facts necessary for explaining anaphora and other phenomena are formally no different from ordinary facts about everyday life that will be present in

some *CG*s and not in others. On this view, semantic and pragmatic research reveals which facts are in the *CG*s for particular speech communities. More precisely, the explanation of the difference between (62a) and (63a) will be as follows. There is some proposition *P* whose content is (or contextually entails) that the speakers can refer to a ball. *P* is entertained when (62a) is uttered due to the conventions of English. As a result of the assertion of (62a), *P* is automatically accepted by the discourse participants as cooperative speakers of English. That is, just as we cannot help but incorporate the goat's sudden presence into the mutually recognized facts of the conversation, we cannot help but incorporate facts about how the conversation is to proceed.

Explaining the possibility of anaphora in this way replicates *R* from *D* without positing extra structure above and beyond the *CG*. Instead of there being a stack of salient drefs in the context that make certain referential expressions succeed, there are propositions in the *CG* about what kinds of speech acts are possible. In this way, the functional effects of structure are replicated, and this *pseudo*-structure seems to provide an explanatory mechanism for the data.

Stalnaker (2014) provides a more recent proposal that makes especially clear the reliance on pseudo-structure I have in mind. He introduces the notion of *derived contexts* in order to explain various phenomena. I will focus on the simplest: supposition. Suppose kangaroos had no tails. When you suppose with me that scenario, neither of us actually believes that it is true. Instead we hold fixed our current scenario (our current context), and introduce another that is not actual to temporarily explore. For Stalnaker, the mechanism that allows for this is the basic *CG*'s evolving to include propositions according to which the discourse participants are supposing some alternate, derived, context is true. That propositions of that kind are *CG* is the explanation for the existence of derived contexts—they only exist because interlocutors mutually accept that they do. When details are added to the imagined situation, the derived context updates in virtue of a proposition's being added to the basic *CG*. That proposition will detail what the speakers are temporarily supposing. When the

supposition ends, the basic *CG* will be updated with the right proposition saying as much. I'll call this kind of proposition, which brings pseudo-structure to a discourse, a *meta-proposition*. In the example of supposition, the meta-proposition that we come to mutually accept might be: *In this conversation we are temporarily pretending that kangaroos have no tails*.

Stalnaker focuses his discussion on entire derived contexts, and does not talk in terms of the specific propositions that interlocutors accept that serve as their basis. But given definition of the common ground (and given that, for the Stalnakerian, there is nothing to communication beyond what happens in the *CG*), one can see that meta-propositions must be the object of mutual acceptance in order for pseudo-structure to arise.

In principle this strategy—positing complicated propositions that make reference to the discourse of which they are a part—could be applied to any candidate element of a discourse context. Instead of representing currently salient questions separately from the *CG*, it is perfectly easy to conceive of there being *CG* propositions that specify the evolution of questions and topics in the discourse. *A should now tell B where A moved the yoga mat, or the current topic is post-punk music*, for example.

Similarly, there can be mutually accepted propositions about which things are salient and what the individual commitments of the interlocutors are. *The lone missing marble is at the center of attention*, or *A doubts B's claim that some zebras have spots*, for example.

The strategy of appealing to what I am calling *pseudo-structure* is not merely theoretical. I noted above Stalnaker's 1998 implementation of this strategy with respect to available anaphoric referents. Stalnaker (2014, pp.92-93) explicitly uses derived contexts to explain a case that mixes QUD and individual commitments, wherein a person's beliefs are the subject matter of a conversation. Green (2017) elaborates on this tactic, arguing that nearly any communicative phenomenon can be accounted for on a purely *CG*-approach. But instead of explicitly noting the pseudo-structural mechanisms (what I've called *meta-propositions*) underlying this possibility, Green merely presents the approach as a light expansion of Stalnaker's program. It ostensibly involves a partition on the *CS* for questions and varying layers

of acceptance type, defining different sets of propositions. But if these are to be vindications of the *CG* as the chief explanatory mechanism of communication, as Green forcefully advocates, then these expansions must happen by way of meta-propositions of the type discussed in this section. The reason is that the *CG* can contain propositions and nothing more. So any maneuver to explain any data must go by way of some kind of mutually accepted proposition. What makes this kind of approach possible is the wide nature of propositions—they can be about anything. So they can be about the very conversation whose *CG* houses them.

Of course one can adopt a view of context as unstructured while allowing for contexts to be formally treated as if they were structured. This would involve thinking (plausibly) that there are methodological reasons for acting as if conversationalists track multiple streams of information. But one can consistently take some approach for methodological reasons while still endorsing another as the actual root explanation. This is to say that one might coherently believe that the *CG* is the only true information state in a discourse (thanks to the mutual acceptance of meta-propositions that enforce pseudo-structure), while at the same time taking no issue with the linguist's modeling discourse in a structured way for the sake of efficiency or readability.

The point is that the *CG* can contain enough information to classify the state of discourse at a high level of complexity. The route I have taken in explaining this position perhaps makes the point seem a little strange—indeed, I will emphasize the strangeness even more when presenting my argument against pseudo-structure later on. But when approached ignoring considerations of structure, something about it seems right. Given the intuitive plausibility of *CG*, of course it would be mutually accepted that some question is at issue, and of course I would only treat something as salient if I accepted that you accepted it as salient also. Conversation is (ideally) cooperative, and the *CG* is the stage for cooperative communication.

4.2.1 Meta-Propositions

I have given an abstract overview of the deflationary account of discourse structure. I have gestured at a series of requirements that have gone largely unnoticed in the literature.⁶ I would now like to make these requirements more concrete, so as to better evaluate the prospects for the deflationary view. I will use the following exchange between Dani and Francis as a touchstone going forward:

- (64) **Keys:**
- a. *Dani*: I lost my keys again.
 - b. *Francis*: I saw them by the coffee pot.
 - c. *Dani*: I checked there already.
 - d. *Francis*: Did you drive yesterday?
 - e. *Dani*: Yeah, I did.

To evaluate the potential for an unstructured *CG* approach to (64), I'll use the following model:

$\mathcal{D} = \{w_1 \dots w_{10}\}$	
$\text{lost}(D, \text{keys})$	w_1-w_9
$\text{saw}(F, \text{keys}, \text{by-coffee-pot})$	w_1, w_2, w_4-w_7, w_9
$\text{search}(D, \text{keys}, \text{by-coffee-pot})$	w_2-w_5, w_8-w_{10}
$\text{drive}(D, \text{yesterday})$	w_3-w_8
$\text{salient-object}(\text{keys}, \text{convo})$	w_2-w_4, w_6, w_7-w_{10}
$\text{salient-location}(\text{by-coffee-pot}, \text{convo})$	w_2-w_5, w_8-w_{10}
$\text{goal}(@ \in \text{drive}(D, \text{yesterday}), \text{convo})$	w_1-w_4, w_{10}

⁶Sometimes though theorists explicitly incorporate aspects of what I discuss here. One notable example, which I discuss shortly, is Roberts (2012–1998).

For the purposes of this discussion, $\text{salient}(keys, convo)$, $\text{salient-location}(by\text{-coffee-pot}, convo)$, and $\text{goal}(@ \in \text{drive}(D, yesterday), convo)$ are the conditions that are most important. They are all what I've called *meta-propositions*—propositions about the current state of the discourse. The propositions having to do with salience determine that something (an object or location) is available for future anaphoric reference (e.g. by 'them' or 'there'). $\text{goal}(@ \in \text{drive}(D, yesterday), convo)$ establishes a goal for the discourse, to determine whether the actual world is one in which Dani drove their car yesterday. The idea behind these propositions is that they detail facts about possible ways the world can be. There are some worlds in which the keys are salient in the conversation above, and some in which they are not. There are some worlds in which there's an accepted goal of determining whether Dani drove yesterday, and some where the goal is different. This model in hand, let's walk through the conversation again, this time monitoring the progression of the *CS* (which for simplicity starts as the whole domain) as each new asserted proposition and covert meta-proposition is added to the *CG*:

$$(65) \quad CS_0 = \mathcal{D}$$

- a. *Dani*: I lost my keys again.

$$\begin{aligned} CS_1 &= CS_0 \cap \text{lost}(D, keys) \cap \text{salient-object}(keys, covno) \\ &= \{w_2-w_4, w_6-w_9\} \end{aligned}$$

- b. *Francis*: I saw them by the coffee pot.

$$\begin{aligned} CS_2 &= CS_1 \cap \text{saw}(F, keys, by\text{-coffee-pot}) \cap \text{salient-location}(by\text{-coffee-pot}, convo) \\ &= \{w_2, w_4, w_6, w_7, w_9\} \end{aligned}$$

- c. *Dani*: I checked there already.

$$\begin{aligned} CS_3 &= CS_2 \cap \text{search}(D, keys, by\text{-coffee-pot}) \\ &= \{w_2, w_4, w_9\} \end{aligned}$$

- d. *Francis*: Did you drive yesterday?
 $CS_4 = CS_3 \cap \text{goal}(@ \in \text{drive}(D, \text{yesterday}), \text{covno})$
 $= \{w_2, w_4\}$
- e. *Dani*: Yeah, I did.
 $CS_5 = CS_4 \cap \text{drive}(D, \text{yesterday})$
 $= \{w_4\}$

Each assertion adds the asserted proposition to the *CG*, thereby shrinking the *CS*. In addition, the appropriate meta-propositions are added to the *CG* at various stages. The NP “my keys” in Dani’s opening assertion, along with Dani’s likely intonation and de-meanour, adds a proposition according to which the keys are salient in the current conversation: *salient-object(keys, covno)*. That the keys are mutually recognized as salient from then on explains why Francis can anaphorically refer to the keys with ‘them’. The same pattern obtains with the location next to the coffee pot. Francis’s question makes it commonly accepted that Dani is to say whether they drove yesterday.

Capturing the flow of conversation in this way brings out three noteworthy features about an unstructured view of context. First, a huge portion of what we say will update the *CG* (and therefore the *CS*) with meta-propositions that introduce pseudo-structure. This is the case even (or especially) for speech acts like questions, which are now functionally identical to assertions in that they remove previously live possibilities from the *CS*. The possibilities they remove are those in which our conversation went differently (e.g. where the question wasn’t asked or had no effect). Second, the speech acts that enforce pseudo-structure must be indexical; they introduce meta-propositions that make reference to the current conversation, dictating what is currently salient, or what the current conversational goals are. Finally, much more complexity is needed in order to capture a context’s pseudo-structural evolution.

In particular, all meta-propositions must be temporally indexed. Imagine Francis were to continue the conversation by saying “Here they are! Wow, when did you get this?” while holding a new key-chain toy on Dani’s keys. The discourse would have a new interrogative goal, to resolve when Dani got the new key chain, and would have a new salient object for resolving ‘this’. In order to avoid having contradictory propositions in the CG —about what the goal is and what is salient—those pseudo-structural facts must be tied to a particular section of time. None of these features are particularly troublesome, but they are necessary details of the unstructured picture that are rarely observed.

One influential proposal that may exemplify an embracing of this deflationary view, meta-propositions and all, is from Roberts (2012–1998). To show how her proposal works, I will explain a simplified version of her picture of discourse contexts.

Simplified QUD Context

A Discourse $D = \langle M, Q, CG \rangle$:

- $M = \langle m_1 \dots m_n \rangle \approx$ list of moves
- $Q = \langle q_1 \dots q_n \rangle$:
 - $q = \{p_1 \dots p_n\} | \langle q, q \rangle$
 - $Q \approx$ ordered stack of questions/sub-questions
- $CG = M \mapsto \mathcal{P}(\mathcal{P}(\mathcal{W})) : \forall m_i, m_{j>i} \in M$:
 - $\llbracket m_i \in M \rrbracket \in CG(m_j)$
 - If $m_i \in Q$, $\llbracket m_i \in Q \rrbracket \in CG(m_j)$
 - $\forall p \in CG(m_i) : \llbracket p \in CG(m_i) \rrbracket \in CG(m_k)$

In this simplified version of Roberts’s proposal, the CG is a function from moves to sets of propositions. This is merely a formal way of easily capturing the temporal evolution of the mutually accepted propositions. There are three requirements on the CG that enforce the

inclusion of meta-propositions. The first requires that at any given stage of the conversation, for each prior move, there is a proposition in the *CG* according to which that move was made. Similarly, at every stage, there will be a separate meta-proposition describing what the current Questions Under Discussion are. Finally, for every proposition p in the *CG*, there will be a proposition $q \neq p$ at every future stage whose content is that $p \in CG$. Regarding this final requirement, Roberts says,

“[It] is intended to capture the fact that at any given point in the discourse, the interlocutors have complete information about the information structure itself, including what moves have been made, which were questions and which assertions, which were accepted, what was in the common ground at the point a given move was made and what questions were under discussion at that point.” (p.15).

This articulation of the motivation highlights a key appeal of the deflationary view of discourse structure. It preserves the informal, intuitive description of the *CG*—that it is the collection of everything mutually accepted—while allowing for theorizing about how different information can interact.

4.3 Grounding Structural Facts

With this more detailed sketch of how pseudo-structure might be understood, I am now in a position to argue against it as the explanation of apparent structure in discourse. This first step of my argument targets the core of what makes pseudo-structure possible: that meta-propositions are not special in how they are admitted into the *CG*—they are propositions like any other that must be mutually accepted by all discourse participants. The general idea is this: that meta-propositions are not special coupled with the fact that they refer to the context itself means that they are only true when they are part of the *CG*. But it is implausible that they are made true by their inclusion in the *CG*. So an unstructured view

$$P \text{ is mutually accepted} \implies P \in CG \implies P \text{ is true}$$

Figure 4.1: Unstructured Order of Explanation

of context, which requires these features, is mistaken. More abstractly, my argument is that (a) contexts are genuinely unstructured if and only if (pseudo)structural facts about discourses (meta-propositions) are true in virtue of their being mutually accepted, but (b) they are true in virtue of something else.

4.3.1 Truth in Virtue of Acceptance

Beginning with the first part: on an unstructured approach, what ultimately grounds communication and the flow of conversation is the *CG*, which is defined as those propositions that are the objects of mutual acceptance by all conversational participants. For pseudo-structure to be imposed, there must be propositions in the *CG*, meta-propositions, whose content replicates structural divisions of information posited by structural theories of discourse. Like any proposition, those meta-propositions can be true or false. Prior to Francis asking the question above, $\text{goal}(@ \in \text{drive}(D, \text{yesterday}), \text{covno})$ is false, because the context is not organized according to that goal yet. In order for the context to be structured according to that goal, the proposition itself must be in the *CG*—mutually accepted as true by Francis and Dani.

Similarly, a deflationary view of structure must hold that facts about what is salient at any given point in a conversation ultimately depend on what facts are accepted as common ground. If it is not mutually accepted that x is salient, then x is not salient. But if x is accepted as salient, then it in fact is salient. For Francis and Dani, they must mutually accept the proposition $\text{salient-object}(\text{keys}, \text{covno})$ in order for that proposition to truly describe what is salient to them in the discourse.

This is the first step in my argument: for an unstructured theory of context, the truth of any meta-proposition depends on its inclusion on the *CG*. The schema is given in Fig.4.1:

for any meta-proposition P , that interlocutors mutually accept P implies that P is an element of their CG , which in turn implies that P is true because P describes the discourse context. That is, its truth depends on whether the conversation's participants hold the right propositional attitude toward it. If they do, then P is CG and therefore true; its inclusion in the CG means that the context does in fact conform to what the meta-proposition describes. If they do not mutually accept P , then the proposition is not CG and is therefore false; it wrongly describes the state of the context.

So for the unstructured theorist, a meta-proposition is true thanks to the propositional attitudes of the conversation's participants for three reasons. First, meta-propositions, like any proposition, can be true or false. Second, meta-propositions describe the state of the context. Third, there is nothing to the context beyond the CG . This last thesis is what makes a given theory of context unstructured.

4.3.2 Acceptance in Virtue of Truth

But there is an alternative order of explanation. Regardless of whether contexts are genuinely or merely pseudo-structured, a meta-proposition is true if and only if the discourse context it describes is in fact the way the meta-proposition says it is. For the unstructured theorist, this means that the meta-proposition itself must be in the CG . But for pictures on which context includes more than just the CG , a context can meet the descriptive content of a meta-proposition without mutual acceptance as a precondition. The structured order of explanation is given in Fig.4.2. For any meta-proposition P , its truly describing the state of the context implies that interlocutors mutually accept P , which in turn implies that P is part of the conversation's CG .⁷

The structured order of explanation is more plausible than the unstructured order of

⁷A meta-proposition's truth will only imply that it is mutually accepted in idealized, non-defective contexts. Strictly, this picture allows for a meta-proposition to be true while not mutually accepted. This is a benefit of a structured approach, and allows models of discourse to capture phenomena such as those discussed by Camp (2018). I discuss this in more detail in Sec.4.4

$$P \text{ is true} \implies P \text{ is mutually accepted} \implies P \in CG$$

Figure 4.2: Structured Order of Explanation

explanation. This to say that meta-propositions are accepted in virtue of their truth; they are not made true in virtue of their being accepted. To argue for this, I will discuss two types of meta-proposition: one regarding an object’s status as conversationally salient, and one regarding communicative conventions.

In (64), the conversation between Dani and Francis, Dani’s keys become conversationally salient. They are (or, more specifically, their location is) the topic of conversation thanks to Dani’s bringing them up. When Francis says “them”, the referent is the keys. It is quite clearly true that *salient-object(keys, covno)*. In virtue of what is that meta-proposition true? On an unstructured approach to context, it must be true in virtue of Francis’ and Dani’s accepting it as such. However, as I argued in the previous chapter, it would be wrong to conceive of conversational salience as ultimately reducible to propositional attitudes. Instead, conversational salience should be understood in terms of *attention*, a lower level psychological state. Specifically, when a conversation’s participants mutually attend to something, that thing qualifies as salient in their conversation—it bears all of the discourse properties that are relevant to semantic and pragmatic inferences about that thing. Furthermore achieving mutual attention is necessary for something to count as conversationally salient. In the last chapter, the upshot of these points was about what conversational salience amounts to, and what its psychological basis is. The same points are relevant to the discussion at hand when it is noted that something’s being salient is a fact about the state of a discourse at a given time. This means that salience is a structural or pseudo-structural feature of context, and so meta-propositions describing what is salient in a conversation will either correctly or incorrectly describe a context at a given time. If, as I have argued, something’s salience is grounded in the attentional states directed toward it and not in propositional attitudes, then a meta-proposition about salience is made true by those psychological states and not

by those propositional attitudes.

To put the point more simply, something is salient when we as conversationalists attend to it in the right way. The proper explanation holds that our attitudes about what is currently salient are responsive to the actual facts about what is salient to us at this moment, and those facts are true thanks to what we are attending to and how. When Dani says “I lost my keys again”, the keys are drawn to the center of cognitive attention for both Francis and Dani, and so in making the assertion Dani makes the keys salient.

The second, but more complicated, kind of meta-proposition that does not conform to the unstructured picture involves communicative conventions. The conventions I have in mind include, for example, those dictating what is grammatical in a language, those specifying what intonational pattern indicates sarcasm, and those involving the dynamic effect of speech acts. Thanks to these conventions, conversation flows easily. Implementations of these conventions will count as structural or pseudo-structural facts about a conversation. In any conversation, there will exist a true meta-proposition describing the order in which particular speech acts were made, for example. My claim is that a proper understanding of how the Common Ground is updated entails that the order of explanation required by a purely unstructured view of context is mistaken. The understanding stems from a subtle but important point from Stalnaker.

In general, given the assumption that the semantics of the language and the rules of the game are common knowledge among the players, we can conclude that when an assertion is made, it will be a manifest event that an assertion with a particular content was made by the person who made it. So the assertion changes the context in a way that depends on the fact that we are playing a game with these particular constitutive rules, but the change will not itself be the result of the application of the rules. That is, it is because of the constitutive rules of the game that a certain act counts as an assertion (in English) that φ , and if it is common ground that the participants are speaking English, then the

act will change the context by making it common ground that the speaker has just asserted that φ . But it is not a rule of the game that the context should change in this way. *This change is just a mutual recognition that an act of a certain kind has been performed.* (Stalnaker 2014, pp. 51–52, emphasis added)

Stalnaker's point is that it is important to distinguish a rule's implementation from its consequences. An implementation of the rules regarding assertion (what Stalnaker calls the *Essential Effect*) might be that a new proposition (that I am looking to buy a rug) comes to be mutually believed. But a consequence of those rules might be that an assertion with particular content has been performed. The fact that such an assertion, with such content, has been performed is what I have been calling a *meta-proposition*. For Stalnaker, this meta-proposition is added to the *CG* because we recognize it as a consequence of the rules that we have agreed to.

I think this is exactly right. But what is left implicit in Stalnaker's discussion is what is explicitly important to my argument. What is left implicit is that the consequences of the rules *are true* thanks to what the rules say, not thanks to what is accepted by the conversation's participants. That is, when one makes an assertion, the meta-proposition that they have made an assertion is true thanks to the conventions dictating what counts as an assertion. The fact that a question preceded the assertion is true because it in fact came first, not because the conversation's participants accept that it came first.

These points should be intuitive. Many facts about what a discourse context is like at a particular time will ultimately be grounded in particular communicative conventions. It is thanks to those conventions that particular meta-propositions are true, and it is thanks to their truth that conversational participants come to mutually accept them (and so add them to the *CG*). But the order of explanation that falls out of this is the one that is incompatible with a strictly unstructured view of context. Such a view requires that facts about a context ultimately be grounded in the conversationalist's accepting them.

4.4 Benefits of Structure

To help adjudicate between a genuinely structured and a deflationary unstructured view of context, I introduced the notions *pseudo-structure* and *meta-proposition*. Meta-propositions describe specific features of a discourse context, and their truth brings on apparent structure. I argued that an unstructured view of context must take a specific position on the truth of meta-propositions, whereas a structured view may maintain that different meta-propositions are true for different reasons, depending on what they describe. I argued that if meta-propositions are not all made true in virtue of their being mutually accepted into the Common Ground. This means that the explanation of the emergence of pseudo-structure from the point of view of an unstructured theory is untenable.

In this section I explore the theoretical utility in thinking of structure as genuine. I will discuss two broad benefits. The first is that a structured view allows for modeling cases of uncooperative speech and insinuation, inspired by arguments from Camp (2018). The second is that a structured approach to context better upholds certain motivations behind an unstructured approach.

4.4.1 Non-Ideal Conversation

I have argued that conversational contexts are structured in principle. Collapsing their theoretical function into only one information type that is updated in only one way is misguided. But there are also particular cases that naturally lend themselves to a structured approach to context. Camp (2018) argues on the basis of *insinuation* that we should understand conversational content in terms of several genuinely distinct information states, including the Common Ground, Conversational Record, and Conversational Score.

Typical cases of insinuation, she observes, involve making a conversational contribution that is admitted into the *CG* but not into the conversational record. Other cases, she argues, involve insinuations the performance of which are not even mutually accepted as having taken

place. Camp characterizes such cases of insinuation as involving the refusal to acknowledge manifestly obvious facts as such.

Without committing to specifics about knowledge, belief, or acceptance, one can imagine the utility in the following pictures:

- (66) *A* insinuates *P*, thereby making it *CG* that *P*, but the conversational record of discourse commitments remains unchanged.
- (67) *A* insinuates *P*, thereby performing a covert speech act on the conversational record, but the *CG* of mutually accepted facts remains unchanged.

The data that Camp (2018) offers could perhaps be explained by a purely unstructured picture of context. For example, one could posit additional propositions in the *CG* detailing what is acceptable to acknowledge. Such an explanation would then, of course, encounter the in principle problems that have been the focus of this chapter (involving the truth of those propositions). But the rough pictures in (67) and (66) are explanatory only when the different information states are truly distinct; the kinds of operations that update one need not necessarily update another. This capacity is exactly what a structured view of context enables.

4.4.2 Cooperative Social Lives

Central to Stalnaker's approach to theorizing about communication is that it is a joint enterprise that, at bottom, is not special. Given the kind of creatures we are, we can achieve coordinated mental states. Our capacity to recognize intentions and the deployment of conventions of meaning allows us to coordinate in increasingly complex ways. Strategically signaling intentions and deploying of conventions with another generates a collection of successfully coordinated-on facts that can in turn be used in reasoning about further acts of coordination. This collection is the Common Ground, and it is, on this view, the basis of

cooperation with other social agents. This neo-Gricean picture of communication holds that facts about what a particular context is like are ultimately no different than other facts that are coordinated on, and is what leads one to an unstructured view of context.

A consequence of my argument against an unstructured view of context is that it in fact abandons this motivation. To provide an explanation of apparent contextual structure, the unstructured theorist must hold that facts about contexts (meta-propositions) are actually unique in how they are made true. On such a view, the facts are true in virtue of their being accepted as true, and so they achieve a special status. As cooperative agents, we then respond to *ordinary* facts in the world by accepting them as true, but the order of explanation is reversed for facts about our conversation.

Even from the neo-Gricean, Stalnakerian perspective, then, a structured view of context should be attractive. Facts about conversations are grounded in the conversations themselves, not in our attitudes about them. Conversations have the features they do because of more basic facts about our psychologies, conventions, and more. So in conceiving of structural facts as genuinely structural, the theorist maintains a central starting motivation.

4.5 Conclusion

I began by setting up a tension between two styles of approach for modeling and explaining discourse phenomena. On one style, the Common Ground is everything. It represents what is mutually known, and conversation is about learning together. The other style incorporates that same kind of mechanism, but does so alongside others. Those other mechanisms track information of a formally different type than that contained in the Common Ground, and function in various intuitive ways patterning with various phenomena. I made a deflationary view of discourse structure the focus. On the deflationary view, the use of discourse-referential meta-propositions recreate the exact structural patterns without needing to posit information states of different types thanks to the imposition of pseudo-structure.

I have argued that, while intuitive at some level, the deflationary picture fails in the details. In particular, it wrongly pushes aside psychological and conventional facts that explain why discourse contexts have certain features.

The upshot, then, is that properly structural views of discourse should not be avoided. In conversation, we keep track of and update various streams of information simultaneously and without effort. That record keeping and update practice is not explained by our having a mutual attitude directed at some complex context-referential proposition, but rather is explained by the existence of various communicative conventions, which are the object of study in contemporary semantics and pragmatics, and by facts about human psychology. This does not rule out the explanatory value of Common Ground; it is still true that we accept some things as background knowledge, and that we come to accept more in the course of conversation. But for some things, facts about the shape of discourse itself, their truth is not owed to what we accept, but to the facts about how we actually go about communicating.

BIBLIOGRAPHY

- Anderbois, S., A. Brasoveanu, and R. Henderson (2013). “At-issue proposals and appositive impositions in discourse”. In: *Journal of Semantics* 32.1, pp. 93–138.
- Armstrong, J. (2018). “Provincialism in pragmatics”. In: *Philosophical Perspectives* 32.1, pp. 5–40.
- (2021). “Communication before communicative intentions”. In: *Noûs*.
- Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Springer.
- (2004). “Discourse topic”. In: *Theoretical Linguistics* 30.2-3.
- Asher, N. and A. Gillies (2003). “Common ground, corrections, and coordination”. In: *Argumentation* 17, pp. 481–512.
- Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Cambridge University Press.
- Asher, N. and L. Vieu (2005). “Subordinating and coordinating discourse relations”. In: *Lingua* 115.4, pp. 591–610.
- Austin, J. L. (1956). “If and cans”. In: *Proceedings of the British Academy* 42, pp. 109–132.
- (1962). *How to Do Things with Words*. Oxford University Press.
- Bach, K. (1992). “Intentions and demonstrations”. In: *Analysis* 52.3, pp. 140–146.
- Bach, K. and R. M. Harnish (1979). *Linguistic Communication and Speech Acts*. MIT Press.
- Barwise, J. and J. Perry (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.
- Beaver, D. I., C. Roberts, M. Simons, and J. Tonhauser (2017). “Questions under discussion: where information structure meets projective content”. In: *Annual Review of Linguistics* 3.1, pp. 265–284.
- Bittner, M. (2009). “Tense, Mood, and Centering”.
- (2011). “Time and modality without tenses or modals”. In: *Tense Across Languages*. Ed. by R. Musan and M. Rathert, pp. 147–188.

- Camp, E. (2018). “Insinuation, Common Ground, and the Conversational Record”. In: *New Work on Speech Acts*. Ed. by D. Fogal, D. W. Harris, and M. Moss. Oxford University Press. Chap. 2, pp. 40–66.
- Clark, H. H. (1996). *Using Language*. Cambridge University Press.
- Clark, H. H. and S. E. Brennan (1991). “Grounding in communication.” In: *Perspectives on socially shared cognition*. American Psychological Association, pp. 127–149.
- Clark, H. H. and C. R. Marshall (1981). “Definite Knowledge and Mutual Knowledge”. In: *Elements of Discourse Understanding*. Ed. by A. K. Joshi, B. L. Webber, and I. A. Sag. Cambridge, UK: Cambridge University Press, pp. 10–63.
- Cohen, P. R. and C. R. Perrault (1979). “Elements of a plan-based theory of speech acts”. In: *Cognitive Science* 3.3, pp. 177–212.
- Cumming, S., G. Greenberg, and R. Kelly (2017). “Conventions of viewpoint in coherence in film”. In: *Philosophers’ Imprint* 17.1.
- DeRose, K. and R. Grandy (1999). “Conditional assertions and ‘biscuit’ conditionals”. In: *Noûs* 33.3, pp. 405–420.
- Devlin, K. (2006). “Situation theory and situation semantics”. In: *Handbook of the History of Logic*. Ed. by D. M. Gabbay and J. Woods. Elsevier, pp. 601–664.
- Ebert, C. and C. Ebert (2014). “Gestures, demonstratives, and the attributive/referential distinction”. In: *Semantics and Philosophy in Europe* 7. Berlin.
- Ebert, C., C. Ebert, and R. Hörnig (2020). “Demonstratives as dimension shifters”. In: *Proceedings of Sinn und Bedeutung* 24. Vol. 1, pp. 161–178.
- Elbourne, P. (2008). “Demonstratives as individual concepts”. In: *Linguistics and Philosophy* 31.4, pp. 409–466.
- Esipova, M. (2019). “Composition and projection in speech and gesture”. PhD thesis. New York: New York University.
- Farkas, D. F. and K. B. Bruce (2010). “On Reacting to Assertions and Polar Questions”. In: *Journal of Semantics* 27.1, pp. 81–118.

- Frege, G. (1892). “Uber sinn und bedeutung”. In: *Zeitschrift für Philosophie Und Philosophische Kritik* 100.1, pp. 25–50.
- Gauker, C. (2019). “Against the speaker-intention theory of demonstratives”. In: *Linguistics and Philosophy* 42.2, pp. 109–129.
- Gazdar, G. (1981). “Speech Act Assignment”. In: *Elements of Discourse Understanding*. Ed. by A. K. Joshi, B. L. Webber, and I. A. Sag. Cambridge University Press, pp. 64–83.
- Gómez, J. C. (1996). “Ostensive behavior in great apes: The role of eye contact”. In: *Reaching Into Thought: The Minds of the Great Apes*. Ed. by A. Russon, K. A. Bard, and S. Parkers. Cambridge University Press, pp. 131–151.
- (2005). “Joint Attention and the Notion of Subject: Insights from Apes, Normal Children, and Children with Autism”. In: *Joint Attention: Communication and Other Minds*. Oxford University Press, pp. 65–84.
- Green, M. (2016). “Conversation and common ground”. In: 174.6, pp. 1587–1604.
- (2017). “Conversation and common ground”. In: *Philosophical Studies* 174.6, pp. 1587–1604.
- Grice, H. P. (1957). “Meaning”. In: *The Philosophical Review* 66.3, pp. 377–388.
- (1969). “Utterer’s meaning and intention”. In: *The Philosophical Review* 78.2, pp. 147–177. reprinted in Grice (1991, pp. 86–116).
- (1991). *Studies in the Way of Words*. Harvard University Press.
- Groenendijk, J. and M. Stokhof (1991). “Dynamic predicate logic”. In: *Linguistics and Philosophy* 14.1, pp. 39–100.
- Grosz, B. J. and C. L. Sidner (1986). “Attention, Intentions, and the Structure of Discourse”. In: *Computational Linguistics* 12.3.
- Grosz, B. J., S. Weinstein, and A. Joshi (1995). “Centering: A Framework for Modeling the Local Coherence of Discourse”. In: *Computational Linguistics* 21.2, pp. 203–225.
- Hamblin, C. L. (1973). “Questions in montague english”. In: *Foundations of Language* 10.1, pp. 41–53.

- Harris, D. W. (2019). “We talk to people, not contexts”. In: *Philosophical Studies*.
- Heim, I. (1982). “The Semantics of Definite and Indefinite Noun Phrases”. PhD Thesis. University of Massachusetts, Amherst.
- Hobbs, J. R. (1979). “Coherence and coreference”. In: *Cognitive Science* 3.1, pp. 67–90.
- (1990). *Literature and Cognition*. Vol. 21. Center for the Study of Language and Information.
- Hobbs, J. R., M. E. Stickel, D. E. Appelt, and P. Martin (1993). “Interpretation as abduction”. In: *Artificial Intelligence* 63.1-2, pp. 69–142.
- Hunter, J. (2018). “Relating gesture to speech: reflections on the role of conditional presuppositions”. In: *Linguistics and Philosophy*.
- Hunter, J. and M. Abrusán (2017). “Rhetorical Structure and QUDs”. In: *New Frontiers in Artificial Intelligence*. Ed. by M. Otake et al. Vol. 10091. Lecture Notes in Computer Science. Cham: Springer International Publishing.
- Hunter, J., N. Asher, and A. Lascarides (2018). “A formal semantics for situated conversation”. In: *Semantics and Pragmatics* 11.10.
- Kamp, H. (1988). “Comments on Stalnaker”. In: *Contents of Thought*. Ed. by R. Grim and D. Merrill. University of Arizona Press, pp. 156–181.
- Kamp, H. and U. Reyle (1993). *From Discourse to Logic*. Springer Netherlands.
- Kaplan, D. (1989a). “Afterthoughts”. In: *Themes from Kaplan*. Ed. by J. Almog, J. Perry, and H. Wettstein.
- (1989b). “Demonstratives”. In: *Themes from Kaplan*. Ed. by J. Almog, J. Perry, and H. Wettstein.
- Karttunen, L. (1969). “Discourse Referents”. In: *Proceedings of the 1969 Conference on Computational Linguistics*.
- Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Vol. 104. Center for the Study of Language and Information.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press. 400 pp.

- Kendon, A. (2017). “Pragmatic functions of gestures”. In: *Gesture* 16.2, pp. 157–175.
- King, J. C. (2001). *Complex Demonstratives: A Quantificational Account*. Cambridge, Mass: MIT Press.
- (2008). “Complex demonstratives as quantifiers: objections and replies”. In: *Philosophical Studies* 141.2, pp. 209–242.
- (2014). “Speaker Intentions in Context”. In: *Noûs* 48.2, pp. 219–237.
- Kratzer, A. (1989). “An investigation of the lumps of thought”. In: *Linguistics and Philosophy* 12.5, pp. 607–653.
- (2002). “Facts: particulars or information units?” In: *Linguistics and Philosophy* 25.5/6.
- van Kuppevelt, J. (1995). “Main structure and side structure in discourse”. In: *Linguistics* 33.4.
- Lascarides, A. and M. Stone (2009a). “A formal semantic analysis of gesture”. In: *Journal of Semantics* 26.4, pp. 393–449.
- (2009b). “Discourse coherence and gesture interpretation”. In: *Gesture* 9.2, pp. 147–180.
- Lederman, H. (2018). “Uncommon knowledge”. In: *Mind*.
- Lepore, E. and M. Stone (2015). *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford University Press.
- Lewis, D. (1979). “Scorekeeping in a language game”. In: *Journal of Philosophical Logic* 8.1, pp. 339–359.
- Lücking, A., T. Pfeiffer, and H. Rieser (2015). “Pointing and reference reconsidered”. In: *Journal of Pragmatics* 77, pp. 56–79.
- MacFarlane, J. (2011). “What is Assertion?” In: *Assertion: New Philosophical Essays*. Ed. by J. Brown and H. Cappelen. Oxford University Press. Chap. 4, pp. 79–96.
- Maier, E. (2009). “Proper names and indexicals trigger rigid presuppositions”. In: *Journal of Semantics* 26.3, pp. 253–315.
- Mann, W. C. and S. A. Thompson (1988). “Rhetorical structure theory: toward a functional theory of text organization”. In: *Text* 8.3.

- McNeill, D. (1992). *Hand and Mind*. The University of Chicago Press. 428 pp.
- (2005). *Gesture and Thought*. Chicago: University of Chicago Press. 318 pp.
- Michaelson, E. and E. Nowak (2022). “On salience-based theories of demonstratives”. In: *Salience: A Philosophical Inquiry*. Ed. by S. Archer. Routledge. Chap. 4, pp. 70–88.
- Mount, A. (2008). “Intentions, gestures, and salience in ordinary and deferred demonstrative reference”. In: *Mind & Language* 23.2, pp. 145–164.
- Murray, S. (2014). “Varieties of update”. In: *Semantics and Pragmatics* 7, pp. 1–53.
- Murray, S. and W. Starr (2018). “Force and Conversational States”. In: *New Work on Speech Acts*. Ed. by D. Fogal, D. W. Harris, and M. Moss. Oxford University Press. Chap. 9, pp. 202–236.
- (2020). “The structure of communicative acts”. In: 44.2, pp. 425–474.
- Neale, S. (1992). “Paul grice and the philosophy of language”. In: *Linguistics and Philosophy* 15.5, pp. 509–559.
- O’Madagain, C. and M. Tomasello (2019). “Joint attention to mental content and the social origin of reasoning”. In: *Synthese* 198.5, pp. 4057–4078.
- Pagin, P. (2004). “Is assertion social?” In: *Journal of Pragmatics* 36.5, pp. 833–859.
- (2009). “Assertion not possibly social”. In: *Journal of Pragmatics* 41.12, pp. 2563–2567.
- Peacocke, C. (2005). “Joint Attention: Its Nature, Reflexivity, and Relation to Common Knowledge”. In: *Joint Attention: Communication and Other Minds*. Ed. by N. Eilan, C. Hoerl, T. McCormack, and J. Roessler. Oxford University Press, pp. 298–324.
- Portner, P. (2004). “The semantics of imperatives within a theory of clause types”. In: *Proceedings of the 14th Semantics and Linguistic Theory Conference* 14, p. 235.
- (2007). “Imperatives and modals”. In: *Natural Language Semantics* 15.4, pp. 351–383.
- Portner, P. and B. H. Partee, eds. (2002). *Formal Semantics: The Essential Readings*. Wiley-Blackwell.
- Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford University Press. 258 pp.

- Reddy, V. (1996). “Omitting the second person in social understanding”. In: *Behavioral and Brain Sciences* 19.1, pp. 140–141.
- Reddy, V. and P. Morris (2004). “Participants don’t need theories”. In: *Theory & Psychology* 14.5, pp. 647–665.
- Reimer, M. (1991). “Demonstratives, demonstrations, and demonstrata”. In: *Philosophical Studies* 63.2, pp. 187–202.
- (1992). “Three views of demonstrative reference”. In: *Synthese* 93.3, pp. 373–402.
- Roberts, C. (2002). “Demonstratives as Definites”. In: *Information Sharing*. Ed. by K. van Deemter and R. Kibble. Stanford: CSLI Publications.
- (2011). “Solving for Interpretation”.
- (2012a). “Information structure in discourse: Towards an integrated formal theory of pragmatics”. In: *Semantics and Pragmatics* 5.
- (2012–1998). “Information structure in discourse: Towards an integrated formal theory of pragmatics”. In: *Semantics and Pragmatics* 5.
- (2012b). “Information Structure: Afterword”. In: *Semantics and Pragmatics* 5.
- Rothschild, D. and S. Yalcin (2015). “On the Dynamics of Conversation”. In: *Noûs* 51.1, pp. 24–48.
- (2016). “Three notions of dynamicness in language”. In: *Linguistics and Philosophy* 39.4, pp. 333–355.
- Sbisà, M. (2002). “Speech acts in context”. In: *Language & Communication* 22.4, pp. 421–436.
- Schiffer, S. R. (1973). *Meaning*. Oxford University Press, USA, p. 200.
- Schlenker, P. (2018). “Gesture projection and cosuppositions”. In: *Linguistics and Philosophy* 41.3, pp. 295–365.
- Schlenker, P. and E. Chemla (2018). “Gestural agreement”. In: *Natural Language & Linguistic Theory* 36.2, pp. 587–625.

- Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- Simons, M., J. Tonhauser, D. Beaver, and C. Roberts (2010). “What projects and why”. In: *Semantics and Linguistic Theory 2010*. Vol. 20. Linguistic Society of America, p. 309.
- Siposova, B. and M. Carpenter (2019). “A new look at joint attention and common knowledge”. In: *Cognition* 189, pp. 260–274.
- Snider, T. (2017). “Anaphoric Reference to Propositions”. PhD thesis. Cornell University.
- Soames, S. (1985). “Lost innocence”. In: *Linguistics and Philosophy* 8.1, pp. 59–71.
- Stalnaker, R. (1973). “Presuppositions”. In: *Journal of Philosophical Logic* 2.4.
- (1978). “Assertion”. In: *Pragmatics*. Ed. by P. Cole. New York: Academic Press, pp. 35–332. reprinted in Portner and Partee (2002, pp. 147–161).
- (1998). “On the Representation of Context”. In: *Journal of Logic, Language and Information* 7.1, pp. 3–19.
- (2002). “Common ground”. In: *Linguistics and Philosophy* 25.5/6, pp. 701–721.
- (2014). *Context*. Oxford University Press. 256 pp.
- Stojnić, U. (2018). *Discourse, Context, and Coherence*. Oxford University Press. Chap. 5, pp. 97–124.
- Stojnić, U., M. Stone, and E. Lepore (2013). “Deixis (even without pointing)”. In: *Philosophical Perspectives* 27, pp. 502–525.
- (2017). “Discourse and logical form: pronouns, attention, and coherence”. In: *Linguistics and Philosophy* 40, pp. 519–547.
- Stone, M. and U. Stojnić (2015). “Meaning and demonstration”. In: *Review of Philosophy and Psychology* 6.1, pp. 69–97.
- Strawson, P. F. (1950). “On referring”. In: *Mind* 59.235, pp. 320–344.
- Syrett, K. and T. Koev (2014). “Experimental evidence for the truth conditional contribution and shifting information status of appositives”. In: 32.3, pp. 525–577.

- Thomason, R. H. (1990a). “Accommodation, Meaning, and Implicature: Interdisciplinary Foundations”. In: *Intentions in Communication*. Chap. 16, pp. 325–363.
- (1990b). “Accommodation, Meaning, and Implicature: Interdisciplinary Foundations for Pragmatics”. In: *Intentions in Communication*. Ed. by P. R. Cohen, J. L. Morgan, and M. E. Pollack. MIT Press. Chap. 16, pp. 325–363.
- Tonhauser, J. (2012). “Diagnosing (not-)at-issue content”. In: *The Semantics of Underrepresented Languages in the Americas (SUCLA)*. Vol. 6, pp. 239–254.
- Txurruka, I. G. (2003). “The natural language conjunction ‘and’”. In: *Linguistics and Philosophy* 26.3, pp. 255–285.
- Wettstein, H. K. (1984). “How to bridge the gap between meaning and reference”. In: *Synthese* 58.1, pp. 63–84.
- Wu, W. (2014). *Attention*. Routledge.
- Zeevat, H. (1999). “Demonstratives in discourse”. In: *Journal of Semantics* 16.4, pp. 279–313.