NON-IDEAL FOUNDATIONS OF LANGUAGE

Jessica Keiser

NON-IDEAL FOUNDATIONS OF LANGUAGE

This book argues that the major traditions in the philosophy of language have mistakenly focused on highly idealized linguistic contexts. Instead, it presents a non-ideal foundational theory of language that contends that the essential function of language is to direct attention for the purpose of achieving diverse social and political goals.

Philosophers of language have focused primarily on highly idealized linguistic contexts in which cooperative agents are working toward the shared goal of gaining information about the world. This approach abstracts away from important issues like power, ideology, social position, and diversity of goals which are crucial to explaining linguistic phenomena both at the semantic and pragmatic levels. This book begins by examining the work of some of the pioneers of this tradition—primarily David Lewis, Paul Grice, and Robert Stalnaker. The author shows that various problems have their source in idealizations made at the foundational level of linguistic theory and proposes to rebuild from the ground up. She presents a non-ideal foundational theory of language which retains the major insights of traditional frameworks while rejecting the social idealizations that guide them. Then, she explores the social and political applications of her account to issues such as dog whistling, propaganda, racist speech, silencing, and manipulation.

Non-Ideal Foundations of Language will appeal to researchers and advanced students in philosophy of language who are interested in the social and political applications of language, as well as traditional metasemantic theory.

Jessica Keiser is Lecturer in Theoretical Philosophy at the University of Leeds, UK. Her published work has appeared in *Journal of Philosophy*, *Mind and Language*, *Philosophy and Phenomenological Research*, *Synthese*, *Inquiry*, and *Philosophical Studies*.

Routledge Studies in Contemporary Philosophy

For more information about this series, please visit: https://www.routledge.com/Routledge-Studies-in-Contemporary-Philosophy/book-series/SE0720

NON-IDEAL FOUNDATIONS OF LANGUAGE

Jessica Keiser

CONTENTS

ACKNOWLEDGMENTS

Writing a book is hard, and I couldn't have done it without the support of family and friends. My writing support group—especially Andy Peet, Thomas Brouwer, Haixin Dang, Marina Baldissera Pacchetti, and Rich Rowland—provided encouragement and a sense of community while I was trying to write in isolation during the pandemic. Helen Steward's company during long walks on the Ilkley moor between writing sessions kept me sane and grounded. I received a lot of support from my friends and colleagues at Leeds, where I won the lottery in terms of department collegiality. Most importantly, I'm grateful for the support of my partner Xavier, who has been there with me through all the ups and downs, my friend Sigma, who always understands, and my parents, who believe in my ability to do hard things.

1

INTRODUCTION

1.1 Idealization in the philosophy of language

I became interested in foundational questions in the philosophy of language while attending graduate school at Yale, where I first encountered Lewis' seminal *Convention*. In that book, Lewis aims to answer the question: what makes it the case that a public language belongs to a given community? His answer is that members of the community participate in conventions of truthfulness and trust in that language, where this involves coordinating assertion with belief in order to serve a shared interest in acquiring information. According to this picture, the essential function of public language is to coordinate information exchange among cooperative agents. This idea perplexed me, as it didn't map onto my own experience of language. To me, language appeared to serve many complex social functions, some of which were cooperative and some of which were adversarial. Though I was completely drawn in by many of the ideas in Lewis' book, I wondered how we could have such different judgments about the core function of our shared linguistic practices.

This is when I first started thinking, at least in a nebulous way, about idealizations in theory building, and how they can go wrong. It seemed to me that Lewis' picture of language as cooperative information exchange involved an idealization that missed—rather than carved out the core features of—its target. The more I looked around, the more I realized that this conception of language was not specific to Lewis; rather, most philosophers of language (at that time) appeared to at least tacitly accept it as a guiding background assumption. As Longino (1990) notes, the background assumptions that guide theory building are often invisible to the

theorists who adopt them, especially when they are deeply entrenched. These assumptions can generate what Camp (2019) calls a "perspective" about a given domain of inquiry. Perspectives influence theorists by "guiding what information they take to warrant explanation and what they dismiss as irrelevant or un-reliable; which explanations they find compelling; and what predictions they make about counterfactual contingencies and future events".[1] She points out that though such perspectives can aid in understanding, they can also be perniciously self-reinforcing, leading theorists into distortions that can be difficult to discern from the inside.[2] I began to suspect that the characterization of public language as a mechanism for cooperative information exchange was of this pernicious sort.

Recently, more and more theorists working on social and political issues in the philosophy of language have started to interrogate and challenge this idealized perspective on our linguistic practices.[3] The majority of this work, however, has taken place within the domains of speech act theory and the semantics/pragmatics of particular expression types like slurs or generics, or has been concerned with the active project of altering our current linguistic representations (e.g., conceptual engineering and metalinguistic negotiation). My project builds on this movement while shifting the focus of its application to a more foundational level of the theory of language, dealing with issues such as the nature of meaning, discourse structure, and linguistic conventions. Currently, there is a disconnect between work at this foundational level—which tends to adopt the conception of language as cooperative information exchange—and work being done at higher levels of linguistic theory (e.g., speech act theory, semantics/pragmatics, conceptual engineering, etc.)—which commonly rejects it. This creates an uncomfortable incongruity in our overall theory of linguistic communication, given that work in the latter domains often implicitly rely on foundational theories that are premised on a conflicting picture of the nature of language. My project aims to incorporate the insights of recent work in social/political philosophy of language into a new foundational theory that rejects the idealized conception of language as cooperative information exchange. If successful, this project will be a step toward developing a more integrated theory of public language that reflects our social and political realities.

The book is divided into two parts. In the first I look at the ways in which the idealized conception of language as cooperative information exchange shows up at the foundational level of the philosophy of language, particularly in the work of Paul Grice, David Lewis, and Robert Stalnaker. I show that this conception of language was a point of departure for each of these theorists, shaping their work in problematic ways. As parents often say to children, *when others pick on you, it is usually a sign that they like you*. This is certainly true in the present case; Grice, Lewis, and Stalnaker are the three philosophers whose work has most profoundly influenced my own,

and the reason I pick on them in particular is because I find so much value in their work. My aim is not to undermine their projects, but rather to build on their most fundamental insights while doing away with a pernicious idealization that created a handicap for each of them. This positive project is developed in the second part of the book, where I first provide a sketch of a non-ideal foundational theory of language, and then develop its details by contrasting it with the work of Grice, Lewis, and Stalnaker.

The positive proposal conceives of linguistic communication as a complex social affair characterized by a multiplicity of goals and interests. Though linguistic communication serves many different functions, I argue that one in particular plays a primary explanatory role: this is the goal of attention-direction, which I use as a central concept in developing a broader foundational theory of language. This theory can be roughly divided into three interrelated domains: the theory of meaning, the pragmatic theory, and the metasemantic theory. The theory of meaning seeks to articulate a very basic act of meaning that figures into the metasemantic and pragmatic theories: in performing this act, agents attempt to direct the attention of others. Such acts of attention-direction, however, are not limited to contexts of cooperative information exchange, and can be used for diverse social and political goals. While this picture is inspired by Grice's work in various ways, it allows that acts of meaning can be covert, and that we need not rely on assumptions of cooperation in order to interpret one another. The metasemantic theory aims to answer this question: In virtue of what does a language L belong to a linguistic community? Here I draw on Lewis' general picture of linguistic convention, while rejecting various details that arise from his conception of language as cooperative information exchange. Rather than being grounded in cooperative conventions of information exchange, public language use is grounded in conventions of attention-direction that are complex and varied with respect to their level of cooperation. The pragmatic theory is an extension to the theory of meaning in the sense that it explores the goals and effects of a broader range of speech acts that extend beyond attention-direction. I demonstrate how we might represent these speech acts in a framework that draws on Stalnaker's work in formal pragmatics, while incorporating a more diverse range of mental states and rejecting the requirement that these states must be shared.

I recognize that in orienting the discussion around Grice, Lewis, and Stalnaker, I ignore a range of other approaches to foundational issues in the philosophy of language.[4] This is an unfortunate necessity, as the literature is too vast to address the way my project is situated within its entire scope. I orient the project in relation to the work of these three philosophers in particular for several reasons. First, though I reject a common idealization that they each adopt, I share with them an approach to theorizing about language that can be characterized by the following methodological principles:

(1) *Autonomy of Pragmatics*: According to this principle, named by Stalnaker (2014), general facts about the nature and structure of communication (which I call *pragmatic facts*) can be theorized independently of conventional devices governing specific linguistic practices.[5] These more general pragmatic facts are analyzed in terms of facts about the mental states, behavior, and relations between language users.

(2) *Pragmatics First*: According to this principle, pragmatic facts are more foundational than semantic facts in the sense that the latter are analyzed in terms of the former. The result is a reductive picture of semantic facts according to which they are ultimately reducible to more general and familiar facts about the mental states, behavior, and relations between language users.

These principles characterize what I take to be the right big-picture way to think about the nature of language. Focusing on the work of Grice, Lewis, and Stalnaker allows me to situate my work within this particular theoretical framework. Another reason for orienting the discussion around their work is that it has been fundamental not only in shaping my own project, but in shaping the philosophy of language as a whole. Most theorists working within this subdiscipline are both familiar with, and reliant on its basic conceptual machinery. My hope is that orienting my project in relation to theirs will allow readers to see how my approach can be adopted without the need to abandon wholesale the theoretical tools and conceptual frames with which they are most familiar.

1.2 Two senses of idealization

There are two ways in which we might use the term *ideal*, and its counterpart, *non-ideal*. One is specifically normative—something is ideal if it conforms to some normative standard, and is non-ideal if it fails to conform. An alternative use is in application to theories that either employ, or are devoid of some relevant theoretical idealization. Idealized theories misrepresent their target phenomenon in some way, and for some purpose (more on this below). Though both uses of these terms will be relevant to my project here, my aim is to develop a non-ideal theory in the latter sense of the term. In particular, I offer a theory that does not mischaracterize our linguistic practices to fit the mold of cooperative information exchange. Importantly, it doesn't follow that my proposal will be devoid of any theoretical idealizations whatsoever—rather, it is non-ideal *in this particular respect*. There is also a sense in which this project is concerned with the normative senses of *ideal/non-ideal*. It adopts as a background assumption that speech is normatively non-ideal in certain respects, e.g., that it often exhibits failures of cooperation which serve to promote harm and injustice.

Such non-ideal cases of speech are treated as relevant and important data points that an empirically adequate theory must be able to accommodate.

These different senses of ideal/non-ideal can often be disambiguated in context by noting what kinds of things they are applied to. When we are talking about ideal/non-ideal *theories*, the theoretical sense of these terms is often at issue. But this need not be the case, as there are also ways in which theories can be *normatively* ideal/non-ideal. To add to the confusion, this can sometimes be a consequence of whether or not they are ideal in the theoretical sense. A theory that serves to perpetuate a problematic ideology, e.g., would be non-ideal in the normative sense. Various feminist philosophers and philosophers of race have pointed out ways in which theoretical idealizations can lead to this result. Charles Mills (2005), for example, notes that theories of justice which incorporate theoretical idealizations that involve, e.g., modeling human beings as normatively ideal agents can undermine the fundamental aim of such theories, which is to create a more just world:

> If we start from what is presumably the uncontroversial premise that the ultimate point of ethics is to guide our actions and make ourselves better people and the world a better place, then the [ideal] framework above will not only be unhelpful, but will in certain respects be deeply antithetical to the proper goal of theoretical ethics as an enterprise. In modeling humans, human capacities, human interaction, human institutions, and human society on ideal-as-idealized-models, in never exploring how deeply different this is from ideal-as-descriptive-models, we are abstracting away from realities crucial to our comprehension of the actual workings of injustice in human interactions and social institutions, and thereby guaranteeing that the ideal-as-idealized-model will never be achieved.[6]

Feminist epistemologists and philosophers of science have also noted that theoretical idealizations involved in the use of, e.g., androcentric models, can reinforce sexist stereotypes and sexist practices. Androcentric models characterize the phenomenon of inquiry from the perspective of men's lives, treating other perspectives as irrelevant. Such theories often result in rendering women's activities and interests invisible, and in reinforcing gendered power relations. Ideal theories which represent language as cooperative information exchange, I want to suggest, may be normatively non-ideal in similar respects. Such theories often render invisible the experiences of marginalized groups, and abstract away from linguistic realities that are crucial to our understanding of the mechanisms of linguistic injustice, making it more likely that those injustices will persist.

The conception of language as a mechanism for cooperative information exchange reflects the scientistic mindset of the founding fathers of the

philosophy of language, who saw language as a tool for developing and recording our final theory of the world.[7] As Craige Roberts (2012)—who has done influential work in extending Stalnaker's pragmatic framework—puts it, the goal of linguistic exchange on this conception is to answer the "Big Question": *What is the way things are*?[8] This characterization, however, best corresponds to the interests of a very specific and privileged set of people. While members of marginalized groups have as much of an interest as anyone else in gathering information and understanding the world around them, they have historically been denied entry into the halls of scientific institutions. Moreover, members of such groups are systematically subjected to various forms of hostility, violence, manipulation, and other kinds of injustices through the use of language. A conception of language according to which it essentially functions as a mechanism for cooperative scientific enquiry is completely divorced from these kinds of experiences. One might be tempted to respond by pointing out that those in privileged groups participate in their fair share of linguistic conflict. Indeed, competition and conflict are coded as "male", and theories which focus on these features have in many cases been accused of being androcentric.[9] However, while members of privileged social groups experience hostile uses of language, they are less likely members of marginalized groups to experience it *systematically*. Thus, it is more coherent from the former perspective to treat such uses of language as aberrations from the norm that can be abstracted away from for the purposes of inquiry. For those whose experience of linguistic practice involves systematic injustice, this way of characterizing our linguistic practices may seem far less plausible. It should be no surprise, then, that the growing field of social/political philosophy of language has been spearheaded by members of groups who are accustomed to being barred from information exchange, whose ways of exchanging information do not conform to the practices of institutional science, who systematically experience harmful uses of language, and to whose linguistic contributions others systematically fail to respond in cooperative ways.

The normatively non-ideal social and political implications of theoretically idealized theories are often hidden from view. For this reason, it is especially important to seek out and interrogate them. As Longino (1990) notes:

> The ideological or political value of the assumptions framing a scientific research program may not be apparent to practitioners within the program and … practitioners are not all motivated to pursue a research program because of the political or social implications that it may have. But that such a program thrives and has an audience is often very much a function of its ideological power. Researchers who do not examine the presuppositions of their work may, therefore, quite unintentionally be furthering some ideological program as well as a scientific one.[10]

Accordingly, we would do well to dig deeper into the background assumptions that frame our theories of language, and to ask whether these theories are adequate to the purposes we would like them to serve.

1.3 Idealization in theory building

Though I've been speaking about theoretical idealization, I haven't yet said much about what this involves. While there are different ways of thinking about idealization, I have found Weisberg's (2007, 2013) characterization especially clear and helpful, and will rely on that account here.[11] Weisberg's discussion focuses primarily on scientific models, while I am concerned with philosophical theories. Thus, I will often substitute the term *theory* where Weisberg uses *model*; I don't anticipate that this will make a difference, as the theories I am concerned with qualify as models in Weisberg's sense.[12] I follow Weisberg in taking a theory to be idealized when it fails to represent some aspect of its target.[13] There is also a psychological component to idealization; idealizations are made deliberately, and for certain purposes.

This characterization suggests that there are three important components to recognizing and evaluating idealization: the theory, the target phenomenon, and the theorist(s)'s intentions. Given that idealizations are partially determined by the intentions of the theorist, one way to taxonomize them is by sorting them according to the theorist's reasons for failing to represent certain aspects of the target phenomenon. Though Weisberg identifies three such categories, throughout this book I focus on the following two: minimalist and Galilean idealization. I argue that the conception of language as cooperative information exchange, as it is manifested in the work of Grice, Lewis, and Stalnaker, involves both of these idealizations.

Weisberg characterizes minimalist theories as those which contain "only those factors that make a difference to the occurrence and essential character of the phenomenon in question".[14] Because Weisberg is concerned primarily with scientific models, his discussion revolves around causal factors. For our purposes, we can accept this characterization, but broaden our conception of relevant factors to those which play a metaphysically determining role. That is, I assume that philosophical theories are concerned not only with causal relations, but with certain asymmetrical metaphysical relations that are sometimes called *grounding* relations. I want to remain neutral on various issues connected to the grounding literature, so I do not commit myself to any theses about, e.g., whether such relations are unified or analyzable.[15] I assume simply that they are asymmetrical relations, and that they are importantly related to explanation, but further details should not matter for our purposes here. A theory of language which used the conception of language as cooperative information exchange as a *minimalist idealization*, then, would look roughly like this: First, it would represent

only those features of the target phenomenon which involved cooperation and information exchange—i.e., non-cooperative uses of language and uses other than information exchange would be left out. Second, the theorist's reasons for representing the target this way would be that they thought that alternative uses of language played no role in metaphysically determining the target phenomenon.

The second type of idealization we'll consider is *Galilean idealization*. Weisberg describes Galilean idealization as a pragmatic and non-permanent practice. In this case, theorists introduce distortions into the theory in order to render it more theoretically tractable, with the future goal of de-idealization. As the simplified theory is better understood, more complexity is systematically added, bringing it closer to accuracy. So, in contrast to minimal theories, Galilean theories are not necessarily restricted to all and only factors that are considered to be essential to the phenomenon in question—rather, they are restricted to those which would allow the development of a tractable theory.[16] A theory of language which used the conception of language as cooperative information exchange as a Galilean idealization, then, would look roughly as follows: Like the minimalist idealization, it would represent only those features of the target phenomenon which involved cooperation and information exchange—i.e., non-cooperative uses of language and uses other than information exchange would be left out. However, here the theorist's reasons for representing the target this way would be in order to develop a tractable theory, with the goal of eventually adding alternative uses of language to the theory. One way to engage in Galilean idealization is by simple trial and error, however throughout this book I'll be thinking of this practice in a more substantial sense—that is, I assume that the idealization of language as information exchange is adopted as a principled and strategic starting point, rather than as a place that looks as good as any other.

Because I critique the use of such idealizations throughout this book, it will be helpful to think a bit about their success conditions. This will help us determine whether the theories that utilize these idealizations have been a success, or whether these idealizations have hampered them in some way. These success conditions will differ depending on what kind of idealization is at issue, given that different types of idealizations are intended to serve different purposes. Of course, theories are measured relative to many different criteria (such as simplicity, fruitfulness, empirical adequacy, etc.), and ideal theories are no different in this regard. What sets them apart is that they provide distorted representations of their target. For this reason, it will be helpful to consider their success conditions in relation to empirical adequacy in particular.

The success conditions of minimalist idealization are perhaps the most straightforward to characterize, even if difficult to establish. A minimalist theory is successful if it does in fact represent the essential features of its

target phenomenon. In the case of a minimalist theory of language which employs the conception of language as cooperative information exchange, the theory will be successful if cooperation and information exchange turn out to be all and only the uses which play a role in metaphysically determining the target phenomenon. How do we go about establishing whether a theory succeeds in this regard? Failure is determined in the usual way: by providing counterexamples to show that the idealized theory fails to represent a factor that plays a determining role. While success is difficult to establish, it is a sign that the theory is on the right track if non-represented phenomena (such as non-cooperative uses of language) can be shown to be derived from represented phenomena (such as cooperative uses); this would show that the theory has captured the essential features of the broader target phenomenon even though it has restricted its representational scope to a limited range of factors.

The success conditions of Galilean idealization are less straightforward, given that this kind of theory is understood in advance to be inadequate, and given that it is temporary. The idea is to develop it into a better theory by making small systematic changes. The best way to judge the success of Galilean idealization, then, might be in retrospect by considering whether the initial distorted theory provided a suitable starting position for developing a more adequate theory. I take it that this would be the case if the theory were able to be systematically altered in relatively minor ways, without having to undergo substantial demolition and reconstruction. Rather, the tools developed in the initial theory would either be altered slightly, added to, or both. Now, of course there is a sense in which a Galilean theory that needs to be completely rebuilt might be said to be successful if something is learned from the process. While this is certainly true, it's not the sense of success that I'm thinking about. Rather, I consider Galilean theories to be successful if they can be brought to accuracy through minor adjustments and additions, rather than reconstruction.[17] Finally, success should be considered relative to a specific idealization. A particular type of Galilean idealization in a theory will be unsuccessful if we have to reconstruct the tools specific to that particular idealization, even if other aspects of the theory are kept intact. Of course, it's always hard to determine whether a Galilean idealization is successful, even in retrospect—sometimes we simply haven't yet found the right strategy for revising the theory. But if, over time, the idealized theory is never brought to accuracy with adjustments, this is a worrisome sign.

1.4 Some clarifications

Critiques of ideal theory are sometimes met with exasperation, as if they are coming from a place of complete naivety. Critics of idealization, the complaint goes, don't understand that idealization is a useful and necessary part

of theory building. *Of course ideal theories don't represent all aspects of the target phenomenon—they aren't meant to!* This line of argument, however, misunderstands the critique. Most critics of ideal theory, including myself, are not opposed to idealization *tout court*, but to certain ways of idealizing that they find to be pernicious in certain respects. Idealization in theory building is unavoidable; as Cappelen and Dever (forthcoming) point out, a theory of language which tried to represent every linguistic phenomenon would be a useless mess.[18] Thus, I want to be very clear that my critique is not directed toward the practice of idealization in general—indeed, my own proposal will be idealized in various respects. Rather, it will be directed toward a specific set of idealizations in the philosophy of language which reflect a conception of language as cooperative information exchange. Such idealizations are pernicious in the sense that they distort our understanding of the nature of public language, and prevent us from developing the hermeneutical resources needed to articulate the ways in which language can function to perpetuate unjust social systems.

Another confusion that has arisen in defense of ideal theory concerns the distinction between idealization and scope limitation. Cappelen and Dever, for instance, claim that the practice of idealization can never determine the adequacy of a theory—rather, these practices merely determine the theory's scope:

> These choices about how to idealize/abstract/simplify our data aren't choices that will make the difference between right and wrong theories. Whichever way we choose to idealize (assuming that we then do our theory building well) we will end up with a correct theory. But the choices will make a difference in what our theory ends up being about. Idealize in one way, and we end up with a theory of cross-sentential anaphora. Idealize in another way, and we end up with a theory of the role of presupposition in creating plausible linguistic deniability.[19]

This characterization mistakes idealization for scope limitation. Idealization, at least in the way I will be understanding it here, involves a disconnect between the target phenomenon and the theory. An idealized theory will be narrower than the target phenomenon (at least in some respects) in failing to represent some aspects of that phenomenon. For instance, a theory of implicature might be idealized in the sense that it fails to represent non-cooperative cases. So, what is represented by the theory—namely, cooperative cases of implicature—is narrower than the target phenomenon, which is implicature generally. This restriction is deliberate, and may be done for various purposes, as we've discussed above. For instance, in the case of minimalist idealization, the idea would be that factors involved in cooperative cases capture the essential features of the broader phenomenon. Or in

the case of Galilean idealization, the motivation is pragmatic—we simplify by considering only cooperative cases in order to get a tractable theory off the ground, with the end goal of eventually introducing non-cooperative cases.

A theory of limited scope, in contrast, does not exhibit this disconnect between target and theory. A theory of *cooperative implicature* which failed to represent non-cooperative cases would not be an idealized theory— rather it would simply be a theory whose intended representational scope was limited to certain instances (the cooperative ones) of a broader phenomenon (implicature). But because the theory and target are aligned, this would not be a case of idealization. Granted, a theory of limited scope may be indistinguishable from an ideal theory; however, idealization is determined not just by features of the theory, but also by the target phenomenon and the intentions of the theorists.[20]

The topic of this book will be idealization, not scope limitation. What I hope to offer is not a theory with a broader scope than those which I critique, but rather a non-ideal theory of the same target phenomena. My hope is that this theory will offer an improvement over its counterparts with respect both to empirical adequacy and to usefulness in social/political projects. At the very least, I hope that the book will encourage philosophers of language to be more self-conscious about the idealizations that they are making, and the purposes that they serve. As Appiah (2017) notes,

> … an idealization involves ignoring the truth in a way that is useful for some purpose … In every case, it is worth asking, I think, whether we can identify the falsehoods we are treating as true, why it might be useful to proceed despite their falsity, and for which purposes it is useful.[21]

Notes

1 Camp (2019, p. 30).
2 Camp (forthcoming, p. 14).
3 Not all of the challenges have been recent. Wittgenstein (1953), for instance, questioned whether there is any unifying function of language, and Austin (1962) brought out the importance of functions other than information exchange.
4 For a helpful overview that contrasts two basic types of approaches, see García-Carpintero (2012a) and García-Carpintero (2012b).
5 See Stalnaker (2014, p. 1).
6 Mills (2005, p. 170).
7 See, e. g., Wittgenstein (1961), Frege (1879), Carnap (1937), Quine (1960).
8 Roberts (2012, p. 5).
9 See Anderson (1995) and Keller (1992) for discussion.
10 Longino (1990, p. 133).
11 While competing accounts may differ in their details, I don't expect such details to be relevant to the main threads of argument throughout the book.

12 For Weisberg, models "consist of an interpreted structure that can be used to represent a real or imagined phenomenon." Weisberg (2013, p. 15). The structures of the theories I examine are made up of states (usually involving mental states such as beliefs, intentions, preferences, expectations, etc.) and transitions between them.

13 Weisberg (2013, p. 98).

14 Weisberg (2013, p. 100).

15 For overviews of the literature, see Trogdon (2013), Raven (2015), Raven (2020), and Bliss and Trogdon (2021).

16 Of course, if the theory is going to be of much use, theorists will likely strive to include at least some essential factors, though without necessarily aiming to be comprehensive.

17 The process of successful Galilean idealization is reminiscent of Kuhn (1962)'s mopping-up operations in normal, paradigm-based science. Thanks to Thomas Brouwer for this point.

18 Cappelen and Dever (forthcoming, p. 3).

19 Ibid., p. 3.

20 Though structurally different, theories of limited scope and idealized theories may obscure understanding for similar reasons. E.g., a narrow-scope theory of cooperative implicature and a narrow-scope theory of non-cooperative implicature may be adequate to their purposes taken individually, but together fail to reveal the more general explanations common to both phenomena. Idealized theories may similarly fail to bring out the explanations needed to generalize beyond the idealization. Thanks to Thomas Brouwer for this point.

21 Appiah (2017, p. 170).

References

Anderson, Elizabeth (1995). Feminist epistemology: An interpretation and a defense. *Hypatia*, 10(3):50–84.

Appiah, Kwame Anthony (2017). *As If: Idealization and Ideals*. Cambridge, MA: Harvard University Press.

Austin, John Langshaw (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

Bliss, Ricki, and Trogdon, Kelly (2021). Metaphysical grounding. In Edward N. Zalta (ed.) *Stanford Encyclopedia of Philosophy*. (Winter 2021 Edition) https://plato.stanford.edu/archives/win2021/entries/grounding/

Camp, Elisabeth (2019). Perspectives and frames in pursuit of ultimate understanding. In Stephen R. Grimm (ed.), *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology*. New York: Oxford University Press.

Camp, Elisabeth (forthcoming). Perspectival complacency, perversion and amelioration. In Wayne Riggs and Nancy Snow (eds.), *Open-mindedness and Perspective*. Oxford: Oxford University Press.

Cappelen, Herman, and Dever, Joshua (forthcoming). On the uselessness of the distinction between ideal and non-ideal theory (at least in the philosophy of language). In *Routledge Companion to Social and Political Philosophy of Language*. New York: Routledge.

Carnap, Rudolf (1937). *The Logical Syntax of Language*. London: Routledge.

Frege, Gottlob (1879) Begriffsschrift. In Michael Beaney (ed.) (1997). *The Frege Reader* (pp. 46–78). Oxford: Blackwell.

García-Carpintero, Manuel (2012a). Foundational semantics I: Descriptive accounts. *Philosophy Compass*, 7(6):397–409.

García-Carpintero, Manuel (2012b). Foundational semantics II: Normative accounts. *Philosophy Compass*, 7(6):410–421.

Keller, Evelyn Fox (1992). *Secrets of Life, Secrets of Death*. New York: Routledge.

Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Longino, Helen E. (ed.) (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

Mills, Charles W. (2005). "Ideal theory" as ideology. *Hypatia*, 20(3):165–184.

Quine, Willard Van Orman (1960). *Word and Object*. Cambridge, MA: MIT Press.

Raven, Michael J. (2015). Ground. *Philosophy Compass*, 10(5):322–333.

Raven, Michael J. (2020). *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Stalnaker, Robert (2014). *Context*. Oxford: Oxford University Press.

Trogdon, Kelly (2013). An introduction to grounding. In Miguel Hoeltje, Schnieder Benjamin, and Alex Steinberg (eds.), *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence*, (Basic Philosophical Concepts) (pp. 97–122). Munich: Philosophia.

Weisberg, Michael (2007). Three Kinds of Idealization. *Journal of Philosophy*, 104(12):639–659.

Weisberg, Michael (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Wittgenstein, Ludwig (1961). *Tractatus Logico-Philosophicus* (trans. Pears and McGuinness). London: Routledge.

Wittgenstein, Ludwig Josef Johann (1953). *Philosophical Investigations*. New York: Wiley-Blackwell.

PART I
Ideal foundations

2

LEWIS

2.1 Introduction

One of David Lewis' earliest and most influential contributions to the philosophy of language was his theory of linguistic convention, developed first in his Harvard Ph.D. thesis and then subsequently in *Convention*. Here and in later articles, Lewis draws on ideas from game theory (most notably Schelling (1960)) to develop a general theory of convention.[1] He then uses this theory as the basis for an account of the specifically *linguistic* conventions that underlie public language use. Lewis' theory offers a story about how conventions can develop without explicit agreement, resolving a circularity worry raised by Quine and others in relation to linguistic conventions in particular—namely, that it would be impossible to get a public language off the ground if doing so required explicit agreement in a public language.[2] Lewis draws upon Schelling's (1960) ideas about salience and mutual expectations to show how agents can successfully and stably coordinate their behavior in the absence of any mode of explicit communication.

Lewis shares a number of methodological commitments with both Grice and Stalnaker, including adherence to a principle that Stalnaker calls the *Autonomy of Pragmatics*. According to this principle, certain general facts about the nature and structure of linguistic communication (let's call these *pragmatic* facts) can be theorized independently of conventional devices governing specific linguistic practices. According to Lewis, the actions that constitute communication in a language *L* are both metaphysically and conceptually independent of any specific linguistic conventions. Such actions, however, may become conventionalized within a community through repetition and the development of certain kinds of expectations

(more on this below). Like that of Grice and Stalnaker, Lewis' work also conforms to a *Pragmatics First* picture of linguistic communication in the sense that he takes semantic facts to be ultimately reducible to pragmatic facts. For Lewis, the fact that a community speaks the language it does can be analyzed in terms of more general facts about the actions and mental states of its members. Unlike Grice and Stalnaker, however, Lewis does not appear to adopt a methodological preference for pragmatic explanations of linguistic phenomena over those which introduce complications to the semantic theory. Rather, his metasemantic picture carries with it a commitment regarding the location of the semantics/pragmatics boundary that is independent of such methodological preferences; facts which are determined by conventions of use are semantic, regardless of their complexity. If one accepts Lewis' theory, settling the question of whether some linguistic phenomenon is semantic or pragmatic becomes an empirical, rather than methodological issue.

Most notably for my purposes here, Lewis shares with Grice and Stalnaker an idealized conception of language, according to which the essential purpose of linguistic communication is to facilitate cooperative information exchange.[3] According to Lewis, language is a way of resolving a situation in which each agent's preferences regarding her own actions are determined by her expectations of what others will do, and where everyone is better off by acting together. For Lewis, the particular problem that language use aims to solve is communication, which he equates with information exchange. The rough idea, echoed in Stalnaker's work, is that everyone has an interest in being able to pass on their own beliefs and to acquire knowledge from others—thus, we are all better off if we can coordinate on an efficient and reliable way of exchanging information. Public language serves this function, as information exchange is more efficient and reliable when everyone decodes information in the same semantic system.

Lewis' metasemantic theory aims to give the grounding conditions of public language use—that is, an account of the facts in virtue of which a language L is a language of a population P. This theory exhibits two categories of idealization described by Weisberg (2007, 2013): Gallilean idealization and minimalist idealization. Minimalist idealization, we recall, aims to capture the essential character of the phenomenon in question, which— in Lewis' case—is linguistic communication. Though this phenomenon is broad, Lewis assumes that this phenomenon can be adequately characterized in terms of the more limited category of information exchange, claiming that "the foremost thing we do with words is to impart information".[4] This methodological approach takes for granted that a metasemantic account which gives the conditions under which a population uses a language to exchange information will have captured the essential features of linguistic communication as a whole.[5]

Lewis also makes use of Galilean idealization, which involves the purposeful introduction of simplifying distortions in order to render the theory more tractable. Once the theorist has come to understand the simplified system, they can systematically remove the distortions, bringing the theory closer to accuracy. Lewis begins by focusing on assertion, which is the speech act most naturally conducive to information exchange. Once his theory of assertoric linguistic conventions is adequately developed, he extends the same basic theoretical machinery to questions and directives, as well as non-literal and non-serious speech. The background assumption is that linguistic communication falling outside the scope of information exchange is nonetheless conceptually and metaphysically derivative of information exchange. Thus, it can be accounted for by introducing complexity to the original account without having to restructure its basic theoretical tools.

I argue below that both of these idealizations are misguided. Cooperative information exchange does not exemplify the essential character of communication. Thus, a metasemantics developed according to this conception offers a distorted—rather than simplified—picture of public language use. Moreover, because alternative ways of communicating are not always derivative of information exchange, the theoretical tools used to represent the latter cannot be adapted to accommodate the former. Rather than introducing complexity to a metasemantics tailored to an idealized conception of language as cooperative information exchange, we must rethink some of its basic commitments and theoretical categories.

2.2 Lewis' theory

Before presenting problems for Lewis' theory, it will be helpful to outline it in more detail—especially given that the positive proposal developed in the second part of the book will draw heavily on what I take to be its core insights. In this section I describe Lewis' general theory of convention and its application to public languages, giving special attention to how this theory maps onto a conception of language as cooperative information exchange. In the remainder of the chapter, I argue that this theory faces challenges with respect to its predictions about the conditions under which it is possible to use a language, as well as its predictions about language change.

2.2.1 Convention

Lewis suggests that conventions are developed to solve particular types of games of interdependent decision, which we'll call *strategy problems*. Strategy problems are situations involving multiple agents, and where each agent's preferred action depends on what they expect the others in the situation to do. Suppose that you and I both want to meet for lunch. Each of

our preferences about our lunch venue will depend on what we expect the other to do; if I expect you to go to the sandwich shop, I will prefer to go there, but if I expect you to go to the falafel truck, I will prefer to go there instead. Your preferences, similarly, depend on what you expect me to do.

Lewis is particularly interested in strategy games that involve multiple equilibria—that is, multiple solutions in which no player would stand to gain by unilaterally changing their course of action. This focus on games with multiple equilibria serves to capture the arbitrary element of conventions, as well as their reliance on interdependent decision making. For instance, if we are mutually aware that the falafel truck is the only venue open for lunch, my choice of that venue will not be arbitrary—nor will it require extensive reasoning about what you will do. If several solutions are available, however, the choice between the falafel truck and the sandwich shop will have a degree of arbitrariness, and will require that we reason about what the other will do in order to ensure that we coordinate on the same outcome.[6]

The label "strategy" might invoke the idea of conflict or competition, but this need not be the case. Rather, strategy problems can fall anywhere along a scale between (and including) what Schelling (1960) calls "pure collaboration games", in which agents stand to gain or lose together, and "zero-sum games", in which one agent's loss is another's gain. Schelling refers to strategy problems located in the middle of the scale as "mixed motive games", because they involve both elements of collaboration and of competition.[7]

Our lunchtime strategy game would be one of pure collaboration if the only thing informing our preferences were a common interest in meeting at the same place. In this case we will each be indifferent about whether we meet at the sandwich shop or the falafel shop; if we end up at the same place, we both win, and if we end up in different places we both lose. It would be a zero-sum game in the case that our preferences were informed instead by conflicting interests: perhaps I am indifferent about where to eat and care only to end up at the same place as you. However, you are avoiding me—so you are also indifferent about where to eat and care only that we do not end up at the same place. In this case, any scenario in which we end up together is my gain and your loss, while the reverse is true in any scenario in which we end up apart.

Lewis is focused particularly on those strategy problems in which "coincidence of interest predominates"—that is, situations which are located on the collaboration end of the scale. So, while he allows that there may be mixed motives to some degree, he is focused on cases in which preferences involving collaboration are stronger than preferences involving competition. Imagine a case in which you prefer sandwiches while I prefer falafel, but we share an even stronger preference to eat together rather than alone. In this situation, our common interests will trump our conflicting interests

in determining what we each decide to do. In the end, we'll both prefer to go for lunch where the other goes—wherever that may be—rather than to eat alone at our preferred venue.[8]

Conventions arise when a group of agents develop a regularity of solving a recurring strategy problem in the same way over an extended period of time. When a solution has worked in the past, it may become the most salient solution in the case that agents are faced with the same strategy problem in the future. Each agent will expect the others to choose the most salient solution, which in turn provides them reason to prefer that same solution themselves. The more a particular solution is adopted, the more agents' expectations and preferences reinforce that behavior; this is how conventions emerge. Though Lewis's account of convention is complex, many of its details will not be relevant to the discussion below. To keep things simple, then, we'll set aside those details and use the following rough characterization: A convention is a self-reinforcing regularity in behavior, adopted as a solution to a strategy problem characterized by a predominance of common interests, and sustained by expectations of conformity.[9]

This picture of convention offers a solution to Quine's puzzle because it doesn't presuppose a public language; while conventions can be established by explicit agreement, this need not be the case. Suppose that last week we met at the falafel truck by accident. This week, we both want to lunch together again, but have not discussed where to meet. Given that we successfully met at the falafel truck last week, this may become the salient solution; we may each reason that the other person will choose the falafel truck since it worked last time, and this will inform our own preferences to go there. If we are successful, then we may reason the same way the following week and the week after that—until we have developed a convention for lunching together at the falafel truck.

2.2.2 Application to language

According to Lewis, public language use is conventional, in that it is a self-reinforcing regularity in behavior, adopted as a solution to a strategy problem characterized by a predominance of common interests, and sustained by expectations of conformity. Once we adopt this assumption, the following questions arise: 1) What is the strategy problem that public language has emerged to solve? 2) What is the regularity in behavior that serves as a solution to this problem? 3) What are the common interests underlying this regularity and how do they lead to expectations of conformity? In this section we'll explore Lewis' account of linguistic conventions by looking at the answers he provides to these questions. Just like in the case of his general account of convention, Lewis' account of specifically linguistic conventions is complex and detailed. Because not all of these details will be relevant to

our interests here, I again adopt a simplified and somewhat schematic version of the theory which will allow me to highlight those aspects which are related to the idealized conception of language as cooperative information exchange.

If public language use is conventional in the Lewisian sense, the first question we might ask is: what is the strategy problem that it has emerged to solve? Though Lewis' answer to this question is nuanced, the general idea can be characterized as follows: language emerged as a solution to the strategy problem of communication. As we'll see below, a lot rides on how we characterize the notion of *communication*. To get a handle on things for now, however, we can think communication in a language L as a way of encoding and decoding contents. Communication involves two roles, that of the speaker and that of the audience. The equilibrium points of the game are those in which the speaker encodes her message in the same language that the audience uses to decode it. Achieving communication involves interdependent decision making. The speaker's choice of how to encode her message will depend on her expectations of how her audience will decode it, and vice versa; if I expect that you will decode my message according to a language L, then I will encode it in L. And if you expect that I will encode my message in L, you will decode it according to L. Communication problems also have multiple equilibria in the sense that there are different options for solving them; we will successfully communicate as long as you decode my message in the same language that I encode it, regardless of which among many languages we choose.

Perhaps the most natural way of thinking about communication is in terms of a speaker and audience of a particular exchange. However, Lewis notes that there are also "diffuse" strategy problems that occur across exchanges over time between present speakers and past speakers, present speakers and past audiences, present audiences and past speakers, and present audiences and past audiences. The equilibria of the diffuse strategy problems are those in which we encode/decode messages in the same language as speakers/audiences have done in the past. If successful, we set up our current conversational partner's preferences and expectations in such a way that she will be inclined to encode/decode messages in that same language. Solving the "diffuse" strategy problem contributes indirectly to our ability to solve the more direct strategy problem that we face with our present conversational partner in that it provides a salient equilibrium point.

The next question we might ask is: what is the regularity in behavior that serves as a solution to this problem? This question essentially asks for an analysis of the behavior that constitutes communication. Lewis' answer reveals his commitment to the picture of language that he shares with Grice and Stalnaker: communication—or at least the type of communication that determines linguistic meaning—amounts to information exchange. Lewis

gives a more precise analysis of information exchange in terms of truthfulness and trust in a language, but for our purposes here we can abstract away from such details.[10] The concerns I raise below target the identification of metasemantically relevant behavior with information exchange, rather than Lewis' particular analysis of it. Thus, going forward I frame the discussion in the more general terms of information exchange, rather than truthfulness and trust.

Lewis' identification of information exchange with the regularity constituting linguistic convention exhibits both Galilean and minimalist idealization. It is Galilean in the sense that he builds the metasemantic framework to fit information exchange, and then makes minor adjustments to its basic theoretical machinery (in this case, truthfulness and trust) in order to capture a broader range of speech acts, such as commands and questions. After all, questions and commands may be considered relevant to the overarching goal of information exchange, even if less directly so than assertion. Other speech acts, however, including jokes and story-telling are considered too far removed from information exchange to be relevant. Indeed, Lewis goes as far as to say that in performing such acts, language users "are not communicating at all".[11] This brings out the sense in which Lewis' methodology involves minimalist, in addition to Galilean idealization: he takes it for granted that information exchange characterizes the essential features of linguistic communication as a whole.

Now that we've identified the strategy problem that (according to Lewis) conventions of public language have emerged to solve, and the regularity in behavior used to solve it, the final question we might ask is: What are the common interests underlying this regularity and how do they lead to expectations of conformity? Recall that strategy problems can fall anywhere along a scale between pure collaboration games and zero-sum games. Lewis claims that conventions are stable solutions to games that fall on the collaborative end of the scale—those that are characterized by a predominance of common interests. That is, while players may be motivated by a variety of considerations, their collaborative motives outweigh their competitive motives. On Lewis' picture, there is an overarching collaborative motive, which is to communicate. An interest in this outcome is shared by all parties; everyone wins if communication is achieved and everyone loses if it is not. While there may be other motives involved, they are outweighed for each player by the motive of achieving communication. He considers a case in which one prefers "the language of his fathers to the language of their conquerors", concluding that the preference for communication will be stronger: "There are few who would give up communication out of piety to the mother tongue, if it came to that."[12] That is, I might prefer L over another language L*, and you might have the opposite preference. However, these conflicting preferences are outweighed by our common preference for

encoding/decoding messages in the same language. Thus, each of us will choose the language that we expect the other to use, because our interest in using the same language outweighs our preferences for using any language in particular.

Because of these common interests, a regularity of using any particular language L will be self-sustaining. The more members of a community observe a regularity of encoding/decoding messages in a particular language L, the more they will expect one another to conform to it. These expectations provide each of them with a reason to conform to the regularity themselves, and to prefer general conformity from others.[13] The most efficient strategy for ensuring successful communication in any random exchange between community members is if all community members are in the habit of using the same language. So maximal—rather than partial—conformity to encoding/decoding messages in the same language is in everyone's interest. Moreover, these preferences hold not only with respect to conversational partners, but also with respect to speakers and audiences across the community over time. This is because the past linguistic behavior of community members influences the preferences and expectations of one's present and future conversational partners. Thus, complete convergence is the most efficient way to ensure successful communication throughout the community over the long run. Given community members' overarching common interest in communication, convergence on a single language will form what Lewis calls a *coordination equilibrium*: no member is better off by switching to a different language unilaterally. Moreover, no member is better off if any *other* community member switches to a different language unilaterally.[14] This makes convergence on a language L a particularly stable solution to the strategy problem of communication.

Because it will become important later, I want to take a moment to contrast the weaker notion of an equilibrium—a solution to a strategy problem in which no agent is better off by changing her strategy unilaterally—with the stronger notion of a *coordination equilibrium*, in which no agent is better off if any agent (including herself) changes their strategy unilaterally. Solutions to pure collaboration games are always coordination equilibria, while solutions to strategy problems on the more conflictual end of the scale may only be equilibria in the weaker sense. Here is an example from Lewis of the latter kind of case:

> Suppose … that many of us are nasty people who want to dress like the majority but also want to have a differently dressed minority to sneer at. We still achieve an equilibrium when we all dress alike, but it is not a coordination equilibrium: nobody wishes he himself had dressed otherwise, but the nasty ones wish that a few other people— say, their worst enemies—had dressed otherwise. The regularity

whereby we achieve this equilibrium is not a genuine convention by my definition, because the element of conflict of interest prevents it from being a means of reaching a *coordination* equilibrium.[15]

For Lewis, all conventions—including linguistic conventions—are solutions to strategy problems characterized by a predominance of common interests. Their solutions are not mere equilibria, but rather coordination equilibria; not only does no one stand to gain by deviating from the convention unilaterally, but no one stands to gain from anyone else deviating unilaterally. Thus, for Lewis, language users are not like the nasty party-goers described above; they do not wish there to be a minority of nonconformers whom they can sneer at. Rather, everyone wants everyone to converge on the same language.

2.2.3 Three Lewisian idealizations

In the following sections we'll look at various problems that arise in connection to Lewis' answers to the foregoing questions, which I call the *three Lewisian idealizations*.[16]
   The three Lewisian idealizations:

1.   The regularity in action constituting public language use is taken to be information exchange.
2.   The sole (or, at least, the predominant) function of public language is taken to be that of solving the strategy problem of communication.
3.   Public language is taken to be rooted in common interests.

These three idealizations are interrelated, and flow directly from Lewis' pre-theoretical assumption that language is essentially a mechanism for cooperative information exchange. In what follows, I outline several ways in which these idealizations prove to be problematic. First, I argue that by taking the relevant regularity in action involved in linguistic conventions to be information exchange, Lewis' account delivers the wrong modal predictions. Second, I argue that empirical data concerning language change is incompatible with Lewis' claims that linguistic conventions are characterized by a predominance of common interests and that they are solutions to a single strategy problem—namely, communication.

2.3 Possible linguistic conventions

As we saw above, Lewis claims that information exchange in a language L is the regularity in behavior that constitutes linguistic convention. While he recognizes that there are other uses of language that fall outside the scope

of information exchange, he does not consider them to be metasemantically significant. Rather, he claims that while such uses "must find a place somewhere in a full account of the phenomenon of language, it may be inexpedient to burden the specification of £ [the language of a population] with them".[17] The general idea is that one of the principle aims of metasemantic theory is to separate out the kinds of use that are relevant to determining the public language of a linguistic community from those that are not:

> If L is a possible language, and P is a population of agents, what relation must hold between L and P in order to make it the case that L is an actual language of P? The obvious answer is: the members of P must use the language L. But that answer is neither informative nor clear. They must use L in a certain way. If everyone in P used L by telling lies in L or by singing operas in L (without understanding the words), they would be using L but not in the right way; L would not be their language.[18]

For Lewis, the relevant kind of use is information exchange; other uses can be bracketed off from metasemantic theory and squeezed into "a full account of the phenomenon of language" at a less foundation level. Insofar as we are interested in the kind of use that determines the public language of a linguistic community, however, we should focus on information exchange. As far as I'm aware, Lewis does not provide any justification for this claim. Rather, as with Grice and Stalnaker, he seems to simply adopt—without argument—the starting assumption that information exchange is the characteristic aim of language use. In this section, I suggest that there are at least *prima facie* reasons to doubt this claim—and that in the absence of positive reasons in its favor, we ought to reject it. The reasons for doubt come in the form of counterexamples: examples of possible (and in some cases actual) linguistic communities that the Lewisian account deems impossible.

It seems possible for a community to use a language—English, for instance—non-literally (e.g., ironically or metaphorically) more often than not. Lewis' account counterintuitively predicts that it is not possible for English to be the public language of such a community. Though they may be participating in a practice of exchanging information, this information fails to map onto English in the right way; this community fails to exhibit a regularity of using its expressions to exchange information involving the contents that the English assigns to them. (For readers who are having trouble picturing this scenario, *Star Trek: the Next Generation* provides a fictional example of such a community. In the 102nd episode, the crew encounters a vessel manned by aliens who only speak in metaphors. Their Tamarian language is therefore untranslatable by the universal translator.)[19] Lewis (1975) considers this possibility, and offers a consolation prize. Though his theory can't

deliver the intuitive prediction that English could be the public language of such a community, it can offer a different public language in its place. He suggests that contrary to appearances, English would only be a simplified approximation to this community's language—call it English*—which may be obtained by specifying certain systematic departures from English.[20] (A simplified example: if *s* means *p* in English, and is systematically used ironically by this community, then *s* means *not p* in English*).[21] Lewis suggests that this counterintuitive prediction is mitigated by the fact that English* is suitably related to English.[22] However, functions are easy to come by; there will be some such mapping between English* and any language you please. Lewis provides no principled reason to think English, in particular, plays a privileged role in determining the language of this community. The mapping from English to English* offers no help in mitigating the counterintuitive prediction of Lewis' account if it is trivial.[23] Another problem with this solution is that it implausibly assumes that non-literal use will display enough systematicity to be able to derive English* from English; certain forms of non-literal speech—such as metaphor—tend to be highly unsystematic, and there is no reason to assume that a principled mapping could be specified. [24]

This solution, moreover, fails to carry over to yet another possibility—that of a community using a language primarily for non-serious speech, which we can characterize liberally as any speech that is not aimed at information exchange (e.g., jokes, stories, poetry, etc.) If information were the only metasemantically relevant use of language, then it would be impossible for a community to share a public language without exchanging information in that language. Such linguistic communities seem not only possible, but actual. Ancient Hebrew and Latin, for example, are taught in schools, though rarely—if ever—used to exchange information. Though no longer anyone's native language, they are nonetheless used in a robust sense; contemporary users can decode texts that are written in those languages, and use these languages to encode novel contents. The regularity in behavior underlying the use of these languages—at least at present—is not information exchange. The consolation prize that Lewis offers in the case of non-literal speech is not available here. While non-literal speech may be used to exchange information indirectly, non-serious speech does not involve information exchange at all. Therefore, we will not be able to point to a regularity of information exchange in some alternative language L* that can be derived from such a community's language L. In response to this problem, Lewis suggests that we simply revise the metasemantic theory so that it is restricted to what he calls "serious communication situations":

> We may say that a serious communication situation exists with respect to a sentence a of £ whenever it is true, and common knowledge between a speaker and a hearer, that (a) the speaker does, and the

> hearer docs not, know whether a is true in £; (b) the hearer wants to know; (c) the speaker wants the hearer to know; and (d) neither the speaker nor the hearer has other (comparably strong) desires as to whether or not the speaker utters a.[25]

This move, however, is an *ad hoc* fix. In the absence of principled, independent reasons for thinking that contexts of information exchange are uniquely relevant for determining public language meaning, the adoption of such a restriction carries the risk of introducing distortions into the theory that ultimately serve to cloud—rather than to enhance—our understanding of the phenomenon of public language use.[26]

It might be tempting to brush off such modal considerations on the basis that the counterexamples are too remote to be taken seriously, or that intuitions about them are shaky. While I share this circumspection about using far-flung possibilities to guide theorizing, they can nonetheless be instructive. These examples provide at least defeasible consideration in support of the possibility of using a public language without exchanging information in that language. Though defeasible, they should be taken seriously in the absence of countervailing considerations. Our metasemantic theory should not rule out by *fiat* the possibility of these, or other kinds of linguistic communities without principled and independently motivated justification for doing so.

2.4 Language change

In the previous section we looked at problems with Lewis' suggestion that the regularity underlying linguistic conventions is information exchange in a language L. In this section we'll look at challenges to his claims that linguistic conventions are characterized by a predominance of common interests and that they are solutions to a single strategy problem of communication. Drawing on empirical work in sociolinguistics, I argue that facts about language change are contrary to what Lewis' account predicts. Before presenting this data, let me say a bit about the process of language change, and the general kinds of patterns that we would expect to see if Lewis' account were correct.

Linguists distinguish between what we might think of as internal vs. external drivers of change:[27] We might think of internal drivers of change as involving adjustments to the semantic system that is already in place, which serve to improve its function. Internally driven changes occur because we are constantly making gradual adjustments to our linguistic conventions in order to improve the efficiency and communicative power of our public language. Since these features may be in conflict, considerations of communicative efficiency and communicative power constantly trade off against one another in a process of language building and language erosion.

External drivers of change, in contrast, are driven by changes in contact. Recall that conventions are constituted and sustained by expectations. Because we expect different things from different people (based on our observations of their behavior and other available information), a change in our contacts can alter our expectations about the linguistic behavior of those around us. This, in turn, may lead to an alteration in our own linguistic behavior, which is conditional on what we expect others to do. Thus, language change is externally driven by changes in contact between language users.

Given how external and internal drivers of language change function and interact, I suggest that Lewis's account would predict the following three patterns. First, we should expect the absence of what I call *divergence without distance*—that is, divergence in language that does not result from diminished contact between language users. If communication were the only strategy problem that public language emerged to solve, we would not expect to see linguistic divergence without a change in the frequency in contact, as this would decrease—rather than increase—communicative efficiency. It would suggest that there are competing interests at play; that public language is serving some function other than communication. (A quick note on why we might expect divergence to result from distance: If there were decreased contact between two communities, internal drivers of change might send them off in different directions. There's no single "right" way for a linguistic convention to undergo developments that would increase communicative efficiency and power. So, if two communities separated, such developments would be independent and would likely go in different directions. But if there were no change in contact, we'd expect any such internal changes to be uniform across the community—so no divergence.)

Lewis' account also predicts the absence of what I call *non-neutral convergence*—or convergence along socioeconomic axes. On this account, we'd expect that if two previously independent linguistic communities were to come into frequent contact, their languages would ultimately converge, given a shared interest in communication. But if communication were the only function of public language, as Lewis would have it, we would not expect this convergence to pattern asymmetrically along socioeconomic axes. This type of pattern would suggest that there were competing socioeconomic interests contributing pressure toward converging on the language of one group in particular. Moreover, such asymmetries may also suggest a lack of common interests: if one group is less incentivized to solve the problem, then the solution may disproportionately favor them. (As the conclusion of a price negotiation in the sale of a house will favor the seller in a seller's market.)

Finally, Lewis' account predicts that linguistic conventions will always be *coordination equilibria*. Coordination equilibria, recall, are solutions to strategy problems in which there is a general preference for general

conformity. In the case of language, this means that on the condition that the majority are using some language L, everyone prefers that everyone else (including themselves) conforms to L. (In analogy to the Lewis' party-goers case, no one wants there to be a differently speaking minority to sneer at.) This prediction stems from Lewis' idea that language use is grounded in a predominance of common interests. Solutions to strategy problems that are characterized by a predominance of common interests are coordination equilibria; so, if linguistic conventions are not coordination equilibria, they are not—contrary to Lewis' assumptions—characterized by a predominance of common interests.

Nonetheless we do see all three of these patterns, and they are fairly common. In what follows, I go through some examples, explaining in more detail why such occurrences would be surprising on Lewis' account. In particular, I argue that these patterns threaten the second and third Lewisian idealizations, according to which the sole (or at least the predominant) function of public language is to solve the strategy problem of communication, and according to which public language is rooted in common interests.

2.4.1 The Northern Cities Shift

The Northern Cities Shift is one of the most dramatic changes in English pronunciation to occur in the last few centuries. It is taking place around the Great Lakes area, roughly from Milwaukee to Syracuse. (If you've ever watched *Fargo*—either the movie or the television series set in a small Minnesota town, these characters exhibit an early version of the accent—though this region hasn't fully shifted.) The change in question involves the pronunciation of vowels. It is called a "chain shift" because vowels tend to change together; a change in one causes a domino effect on the others. (If one vowel starts to impede on the space in your mouth reserved for another, the latter will shift to make room for it, and so on. They must occupy relatively distinctive places in order to avoid ambiguity.) It's unclear exactly when this change started, but it was first documented by linguists in the 1960s.[28]

One influential hypothesis concerning the cause of this change (compatible with others, since the causes of language change are complex and multifactorial) is that the phenomenon is a linguistic version of "white flight". Gerard Van Herk (2008) argues that it was triggered by the Great Migration in the middle of the 20th century, when millions of African Americans moved from the South to northern cities, dramatically changing their demographic composition. He argues that the vowel shift was at least partially an attempt by white residents to differentiate their speech from new Black residents—a linguistic form of the residential white flight that was occurring at the same time in these same cities, as white residents rapidly fled neighborhoods that came to be occupied by new Black residents.

Support for this hypothesis is drawn from correlations between participation in the Northern Cities Shift and the speed and degree of Black migration—as well as the degree of residential white flight—in certain locations. Van Herk (2008) suggests that these correlations also account for the unexpectedly sharp and gerrymandered temporal, social, and geographic boundaries of the shift. I won't go into these details, since they won't matter for our purposes here. Instead, we'll examine some of the features of this change that are perplexing given Lewis' picture. For the sake of the discussion, I assume that Van Herk's hypothesis is correct. Though the cause(s) of the Northern Cities Shift is an open empirical question, we need only the possibility of linguistic white flight to cause trouble for Lewis' picture, which rules out such a scenario.

The first thing to note about the vowel shift is that it occurred from a very stable position. The vowel system of American English in this region had been consistent for a long time, suggesting that the change was not driven by internal factors concerning, e.g., improvements in efficiency or communicative power, but rather by external factors. However, this externally driven change occurs in a way that is inconsistent with the predictions that we should expect from Lewisian assumptions. Recall that Lewis' view predicts that linguistic convergence will follow from community convergence. Here we have the opposite—convergence between two linguistic communities instead leads to linguistic divergence. The speech of the white residents is moving away from, rather than toward, new Black residents.

This change comes at a communicative cost. Linguists have documented that the shifted language is often misinterpreted (especially out of context) by non-shifted speakers. This means that there will be a drop in communicative efficiency between shifted speakers and non-shifted speakers within the same community. The fact that this linguistic change impedes, rather than improves communication across the community suggests that it was driven by interests other than—and indeed, in competition with—communication. If the white flight hypothesis is correct, one such purpose would have been to differentiate social groups. The fact that this purpose was achieved at the cost of communication suggests that, contrary to Lewis, a common interest in communication is not necessarily the dominant factor in determining the public language of a community.

A second way that the Northern Cities Shift poses a challenge to the Lewisian picture is that it seems to exhibit a violation of the assumption of common interests: rather than a general preference for general conformity, white speakers appear to have a preference for at least some degree of non-conformity, so that they can establish a group identity *in opposition to* another subset of the broader linguistic community. It is not a coordination equilibrium, but rather like Lewis' example of nasty party-goers who prefer a differently dressed minority to sneer at.

2.4.2 Martha's Vineyard

A similar phenomenon—documented by sociolinguist William Labov—concerns language change on Martha's Vineyard, an island off of Cape Cod.[29] In the 50s and 60s, this island was in the process of transforming from a traditional rural fishing economy to one more reliant on tourism. The increased contact with mainlanders initially confirmed Lewisian predictions in that the speech of the islanders started to converge with that of the mainlanders. However, in the 60s, Labov found that many young islanders—specifically those working within the previously autonomous island fishing and farming industries, whose livelihoods were being threatened by the mainland-dependent tourist economy—started to speak more like their grandparents. In other words, they started to undo the convergence with mainlanders by taking on patterns of speech that were characteristic to the island prior to the convergence.

Like in the case of the Northern Cities Shift, sociolinguists hypothesize that this change was driven by the desire to establish social identity among one group in opposition to another. The islanders were using language as a symbol of solidarity with one another *as islanders* in opposition to mainlanders, who posed a threat to their traditional economic system and way of life. Here again we see features that conflict with Lewisian predictions. As in the previous case, linguistic change on Martha's Vineyard is an instance of divergence without distance. The change doesn't seem to be driven by internal factors like tradeoffs between communicative power and efficiency, or else—given Lewisian assumptions about general preference for general conformity—we'd expect to see it occurring uniformly across the population. Rather, the change appears to be externally driven, but in a way that is inconsistent with what Lewis' account would predict. The speech of the islanders is moving away from the present speech patterns of the majority, and toward older speech patterns that have begun to fall out of use. This change comes at a communicative cost, as convergence on a single language is the best way to ensure communicative efficiency. In this case, the islanders are creating a divergence in speech patterns between subgroups, thereby weakening the communicative efficiency of the community at large. Of course, in spite of this divergence these subgroups can understand each other well enough to communicate, so the cost is limited; nonetheless there is movement from a more communicatively efficient system to a less efficient one, suggesting that communication is not the dominant factor underlying public language.

Second, this case challenges Lewis' assumption of a predominance of common interests across the community, given that there is not general preference for general conformity. The linguistic convention on Martha's Vineyard is not a coordination equilibrium, as those community members

identifying as islanders prefer linguistic behavior that goes against the *status quo*. They *disprefer* that the rest of the population conform to their behavior, as this would undermine the goal of establishing an identity which can be distinguished from other subgroups within the broader community.

2.4.3 Upward mobility

Linguistic change tends to be driven by the upwardly mobile. The lower middle class, for instance, tend to adopt prestigious linguistic forms, or forms associated with the upper class. (Linguistic forms associated with less powerful groups may exhibit what is called *covert* prestige. Adoption of covertly prestigious forms tends to be solidarity based—like in the case of Martha's Vineyard—rather than driven by social mobility.) Change driven by the lower middle class toward adopting upper class linguistic forms is a stable empirical pattern.

On the Lewisan picture, however, we would not expect to see stable patterns of language change breaking along socioeconomic lines. If the only function of language were communication—and if everyone had equal interest in communicating with everyone else—then linguistic convergence would be relatively neutral with regard to socioeconomic factors. There would be no reason for the lower middle class to consistently accommodate their speech patterns to that of the upper class, rather than meeting somewhere in the middle. This asymmetrical convergence suggests, again, that language is serving a non-communicative function relating to social identity. The act of signaling identity with a more powerful group through speech can aid upward movement, given that symbolic displays of power—rather than, e.g., displays of economic recourses—can still buy some degree of status.[30] The more powerful group has no such incentive to signal identity with the less powerful group.

Although I am suggesting that communication is not the *only* metasemantically relevant function of language, it is surely one such function. However, asymmetrical convergence casts doubt on the idea that all members of the linguistic community have roughly the same interests in achieving it. There is some cost to adapting to the speech patterns of a different group; it is not only cognitively demanding, but may also signal disloyalty to one's one group, which can be costly when these groups are in conflict. The upwardly mobile may have a greater incentive to bear this cost, as communicative efficiency with a more socioeconomically powerful group may improve access to economic resources (through, e.g., promotion to managerial positions which require communicative competence with both workers and executives). The upper class, in contrast, have less to gain from communicative efficiency; thus, they can count on the lower middle class to put in the work of accommodating their speech in order to achieve communicative success.

Finally, language patterns across socioeconomic classes fail to qualify as coordination equilibria, suggesting a lack of a predominance of common interests. As the speech of the lower middle class merges with that of the upper class, the latter tends to "flee" in response. This pattern can be clearly seen in naming practices, noted by Levitt and Dubner (2006) after examining data from California birth certificates. Lower class parents tend to give their children names that are trending among the upper class. This has the effect of weakening the prestige signal of the name, causing upper class parents to abandon it. Eventually, as Levitt and Dubner note, "it is considered so common that even lower-end parents may not want it, whereby it falls out of the rotation entirely. The lower-end parents, meanwhile, go looking for the next name that the upper-end parents have broken in."[31] Rather than exhibiting a general preference for general conformity, at least some community members reveal a preference for speaking differently than the masses.

2.4.4 Linguistic gender divide

In just about every culture, there is a marked difference between men's and women's speech forms. This is especially perplexing as there is very little distance between men and women, who tend to cohabitate and interact regularly throughout the course of their lives. Studies show that early in life, boys' and girls' speech patterns are more or less identical, and that they tend to pattern with women's speech. (This is likely a reflection of the fact that women tend to be the primary caregivers and the earliest linguistic influence.) But as they grow older—starting relatively early, but becoming the most pronounced in adolescence when gender identity takes on a greater significance—boys' and girls' speech styles start to diverge. Boys, for instance, start to show a stronger tendency toward vernacular speech patterns, while girls tend toward standard forms.[32] Many linguists hypothesize that this phenomenon can be at least partially explained by the fact that children are taking note of the social significance of speech styles and adapting to the group with which they identify.[33]

At first glance, this phenomenon looks like an instance of divergence without distance. Boys and girls use the same language as children but diverge as they approach adulthood, even though there is no decrease in contact. In one way of describing the case, it exhibits an absence of a coordination equilibrium, threatening Lewis' assumption of common interests. Rather than a general preference for general conformity across the community at large, each group prefers non-conformity by at least half of the population.[34] However, there is an alternative description that is more friendly to Lewisian assumptions: instead of developing divergent languages, the community moves *en masse* to a complementary convention in a more complex

language with gender-relativized rules: *if a woman speak this way, if a man speak that way*.[35] Thus understood, the case is compatible with there being a general preference for general conformity, and thus a predominance of common interests. Even so, it still threatens the assumption that the sole function of language is that of communication, as the move from a simple to a complementary convention involves a cost in communicative efficiency. The linguistic gender divide suggests that language is serving a non-communicative function, which is to establish and sustain social identity.

2.4.5 Linguistic colonialism

Linguistic colonialism is a loaded term that is often used to describe a multi-faceted phenomenon. For our purposes here, we can roughly characterize it as the phenomenon whereby colonized people tend to take on the language of their colonizers. This phenomenon fails to line up with the Lewisian picture given that it is an instance of non-neutral convergence; convergence between the language of colonizers and colonized is clearly and stably correlated with power, rather than being socio-politically neutral.

Recall that on the Lewisian picture, language is a solution to the strategy problem of communication, and everyone has a common interest in achieving this aim. Moreover, peoples' interests are roughly aligned in the sense that they don't much care one way or another how (i.e., in what language) communication is achieved, as long as it is successful. But as I argued above, if this were the case, we should expect socioeconomic neutrality with respect to which language is adopted when linguistic communities converge given that each language is equally suited for this function.[36] The phenomenon of linguistic colonialism suggests instead that an interest in communication is not the sole, or even the overriding factor determining the common language of the converged communities. If anything, there will be a communicative cost to converging in favor of the language of the colonizers, who are often in the minority; this direction of convergence will involve more effort for more people.[37]

Like others we've been discussing, this phenomenon also points to a failure of the sort of common interest that Lewis took to undergird public language use. Though both subgroups likely have some degree of interest in communicating with one another, their incentive structures are misaligned. For the colonizers, the benefits of the social control that follow from having their language adopted as the *lingua franca* may outweigh the benefits of easy communication with the colonized. In contrast, for the colonized group, communication with the colonizers is often a non-negotiable prerequisite for sustaining a living and getting access to crucial resources. It is unsurprising, then, that a solution to their linguistic negotiation would favor the colonizers (more on this in Chapter 6).

2.5 Conclusion

Throughout this chapter I have discussed various problems generated by three idealizations that arise from Lewis' conception of language as a mechanism for cooperative information exchange. The first idealization, according to which the regularity in action constituting public language use is information exchange, leads to counterintuitive results with respect to the possibilities of sharing a public language. At least on the face of things, it seems as though we could share a public language without using that language to exchange information. Of course, we should expect that idealizations will deliver the wrong predictions in some cases—what matters is that they do their job as idealizations. In the case of minimalist idealization, this job is to capture the essential feature of the target phenomenon; i.e., the features that have a metaphysically determining role. However, I have argued that a practice of exchanging information is not essential to public language use if it is possible for there to be linguistic communities who do not exchange information in their public language. Even when focusing on actual, rather than merely possible cases, it is highly questionable whether information exchange characterizes something explanatorily central to the bulk of our linguistic interactions; this is something that is merely assumed by the Lewisian account, rather than argued for. Without a principled reason for adopting such an assumption, the examples presented in Section 2.3 suggest that identifying metasemantically relevant use with information exchange threatens instead to distort our understanding of the grounding conditions of public languages. Moreover, it fares no better as a Galilean idealization, given that Lewis' attempt to generate the right predictions by looking for systematic departures from the basic case of information exchange was unsuccessful. All of this suggests that a viable metasemantic account will need a different conception of what kinds of communicative behaviors are meaning-determining, which will allow for the possibility that a community can share a public language without using it to exchange information. I propose such an account in Chapters 5 and 6.

According to the second Lewisian idealization, the sole (or at least the predominant) function of language is communication. But as we've seen, this idealization conflicts with empirical data concerning linguistic change. Languages sometimes change in ways that impose a cost on the communicative efficiency of the group, suggesting that it is serving alternative—even competing—functions. One might be tempted to reply here that Lewis' account is compatible with the idea that we use language to do things other than communicate. While this is true, communication is nonetheless taken to be the only function that plays a privileged grounding role. For Lewis, the ways that we use language in order to solve the strategy problem of communication are the ways that determine our public language. While

we may then exploit this language to serve other ends, it is its *communicative* role that determines which public language belongs to a given community. Against this idea, the data on language change suggests not only that we use public language for functions other than communication, but that these alternative uses of language are *metasemantically relevant*; they play a role in determining which is the public language of a linguistic community. Again, however, it is important to acknowledge that we are dealing with an idealization, and faulty predictions are to be expected; what matters is whether the idealization is doing its job. In the case of minimalist idealization, the relevant job would be to capture the essential character of public language use. In this regard, the second Lewisian idealization seems to have failed. The idea that public language serves to solve the problem of communication cannot explain the phenomenon of public language, as it fails to account for systematic patterns in natural language use. When it is considered as a Galilean idealization, I am a bit more sympathetic. As I argue in Chapter 6, I think there is something right in saying that language is used—perhaps even necessarily so—to solve the strategy problem of communication. I argue, however, that we ought to be pluralists about the function of public language, without reserving a privileged place in the metasemantic theory for the role of communication.

According to the third Lewisian idealization, we recall, public language is rooted in common interests. That is, all members of any given linguistic community have a shared interest in solving the problem of communication, and few—if any—competing interests. As a result, they face roughly the same payoff structures with regard to its various possible solutions. Whatever solution they choose, then, will be a coordination equilibrium in the sense that there will be general preference for general conformity. Patterns with regard to language change, however, reveal that in many cases there is in fact a lack of general preference for general conformity; many language users are rather more like Lewis' nasty party-goers who prefer to set themselves apart from a subset of the broader group. In such cases, public language use is not a coordination equilibrium, suggesting that it is not serving to solve a strategy problem that is characterized by a predominance of common interests. Thus, the idea that public language use is rooted in common interests also fails as a minimalist idealization; it cannot provide a viable account of the essential character and occurrence of public language, as it fails to account for systematic patterns in natural language use. How does it fare as a Galilean idealization? Is there some way of adding complexity to this idea while retaining the spirit of the overall picture? Here's one way that one might be tempted to do so: In each of the cases that challenge this idealization, there are at least two subgroups of the broader linguistic community which have competing interests. However, perhaps taken individually, each of these communities share a common interest

(such as establishing an identity in opposition to the competing subgroup). Thus, while public language might not be characterized by common interests between members of *all* its users, the response goes, it will nonetheless be characterized by common interests when we move down to the level of the subgroups that make up the broader community. The problem with this response is that the subcommunities that make up a broader linguistic community are highly complex and intersectional; I suspect that each time we move down a level, we will encounter the same problem, and it strikes me as doubtful that we would ever reach a "bottom" level made up of discrete, but fully cooperative subgroups. Even if we could, cooperation between these subgroups could not serve as the main explanatory factor, given that in many cases it will be explanatorily central to note that they are cooperating *for the purpose of competing with* a different group. The third Lewisian idealization, too, fails both as a minimalist and a Galilean idealization. In Chapter 5, I propose a metasemantic account that retains Lewis' insight that public language is a solution to a strategy problem, while rejecting the constraint that this problem must be characterized by a predominance of common interest. In Chapter 6, I show how cases that were problematic for Lewis can be accommodated once this constraint is relaxed.

Notes

1 See also Lewis (1975) and Lewis (1992).
2 See Quine (1969) and Russell (1921).
3 Lewis, Grice, and Stalnaker were each thinking of contexts of information exchange in roughly the following way: the speaker has some information that the audience lacks, the speaker wants to share this information with the audience, and there are no stronger or overriding ulterior motives at play. The notion of "information exchange" may have broader applications in fields like information theory, but I'll be using it in this more restrictive sense.
4 Lewis (1980, p. 80).
5 By "essential features", I mean the features that play a metaphysically grounding role.
6 See O'Connor (2019) on how arbitrariness—and thus conventionality—comes in degrees.
7 Schelling (1960, p. 89).
8 Formally, Lewis characterizes such contexts as those in which the difference between agents' payoffs for any particular option aren't bigger than the difference between payoffs for different options. See Lewis (1969, p. 14.)
9 Lewis also includes the following conditions: (1) the regularity R is a solution to a recurring coordination problem faced by members of the population P, (2) the expectation of conformity to R provides members of P with reason to themselves conform, (3) there is a general preference among P of general conformity to R on condition that most members conform, (4) there are alternative regularities that would have solved the same coordination problem, and (5) these conditions are common knowledge among P. See Lewis (1969) and Lewis (1975) for details (my presentation relies more heavily on the later account). Though the spirit of Lewis' account of convention is standardly taken to be correct, some

of these conditions are controversial. For discussion, see Burge (1975), Skyrms (1996), Marmor (1996), Marmor (2009), Millikan (1998), Gilbert (1989), and Gilbert (2008).

10 To be truthful in some language L is to try never to utter its sentences unless they are true. To be trusting in a language L is to expect truthfulness in L from others—that is, to expect that they will try not to utter sentences of L unless they are true. See Lewis (1975, p. 7).

11 Lewis (1975, p. 28).

12 Lewis (1969, p. 50).

13 Note that for Lewis, communication *simpliciter* is not conventional; it is possible to exchange information without sharing a public language. Rather, what is conventional is communication in a *particular* semantic system *L*. That is, rather than having a convention of exchanging information, a linguistic community has a convention of exchanging information *in some particular language*.

14 This is compatible with a preference for everyone moving *en masse* to a different language.

15 Lewis (1979, pp. 45–46).

16 I don't mean to suggest by this that these are the *only* idealizations exhibited by Lewis' account, nor that these are the only problematic features of the account. Other problematic features have been discussed at length by various theorists, including Burge (1975), Skyrms (1996), Marmor (1996), Marmor, (2009), Millikan (1998), Gilbert (1989), and Gilbert (2008).

17 Lewis (1980, p. 28).

18 Lewis (1969, pp. 176–177).

19 https://memory-alpha.fandom.com/wiki/Darmok_(episode) Thanks to Ed Elliot for this example.

20 Lewis (1975, p. 183).

21 This is a simplification and not meant to be a serious characterization of irony. See Stern (2000, pp. 71–76) and Camp (2006) for discussion.

22 Here is a common response: *Isn't Lewis' prediction in fact correct? If we used English non-literally often enough, wouldn't this eventually lead to language change, i.e., from English to English*?* If what began as non-literal speech became fully conventionalized, this would properly be characterized as language change—however, it would no longer be an instance of the example under consideration because such uses would no longer be non-literal.

23 Cf. Szabó (2019).

24 Friends of Lewis might think that an appeal to grammar will help here, but I doubt this is the case. For Lewis, the grammar of the language is determined by patterns of use, where use is taken to be information exchange. By stipulation, there are no patterns of information exchange in L. Though there are patterns of information exchange in L*, it is hard to see how these patterns will determine a grammar that will lead us back to L (especially given that non-literal speech is not systematic.)

25 Lewis (1975, p. 29).

26 An alternative solution to both examples is to suggest that these communities at one point in time participated in conventions of information exchange in L—and that as non-literal/non-serious speech became dominant, their language retained its original meanings. Getting the right results here would necessitate a revision to Lewis' account, according to which meaning is determined by past, rather than present conventions of use. This doesn't strike me as a promising way to go, as it would fail to accommodate language change. Furthermore, the counterexamples can be revised to stipulate that the conventions in use were always predominantly non-literal/non-serious. Thanks to Thomas Brouwer for pressing me on this point.

27 Drivers of linguistic change are many and varied; I'll just be focusing on a few that are especially relevant to testing the predictions of Lewis' theory.
28 See Fasold (1969), Labov, Yaeger & Steiner (1972).
29 See Labov (1963).
30 See Bourdieu (1991), Chapter 7 for discussion.
31 Levitt and Dubner (2006, p. 202).
32 See Trudgill (1972), Eckert (1997) and Labov (2006) for discussion.
33 See Labov (1963), Labov (2002), and Eckert (2000).
34 The reason I say "at least half" is that these differences are intersectional. Eckert, e.g., claims that she "cannot conceive of speakers as … "doing jock" and "doing male" separately, rather than "doing male jock". So male jocks are not only preferring non-conformity from women, but also from male burnouts (where burnouts are the opposing social group to jocks in Eckert's (2000) study of Belten High).
35 For discussion of complementary conventions, especially as they relate to gender, see O'Connor (2019).
36 Of course, people in power often adopt an ideology according to which their language is inherently better suited for communication, but this is a *post-hoc* rationalization rather than a linguistic reality.
37 This is not to say that there won't be cases in which there are communication-driven reasons to favor the language of the colonizers—rather, that we should not expect such cases to be the norm.

References

Bourdieu, Pierre (1991). *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.

Burge, Tyler (1975). On knowledge and convention. *Philosophical Review*, 84(2):249–255.

Camp, Elisabeth (2006). Metaphor and that certain 'je ne sais quoi'. *Philosophical Studies*, 129(1):1–25.

Eckert, P. (1997). The whole woman: Sex and gender differences in variation. In N. Coupland, and A. Jaworski (eds.), *Sociolinguistics*. Modern Linguistics Series. London: Palgrave. https://doi.org/10.1007/978-1-349-25582-5_18

Eckert, Penelope (2000). *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Oxford: Blackwell.

Fasold, Ralph (1969). *A Sociolinguistic Study of the Pronunciation of Three Vowels in Detroit Speech*. Washington: Mimeographed.

Gilbert, Margaret (1989). *On Social Facts*. Abingdon: Routledge.

Gilbert, Margaret (2008). Social convention revisited. *Topoi*, 27(1-2): 5–16.

Labov, William (1963). The social motivation of a sound change. *WORD*, 19(3):273–309. https://doi.org/10.1080/00437956.1963.11659799

Labov, William (2002). Driving forces in linguistic change. International Conference on Korean Linguistics, August 2, 2002, Seoul National University.

Labov, William (2006). *Atlas of North American English*. Berlin: De Gruyter.

Labov, William, Yaeger, Malcah, and Steiner, Richard (1972). *A Quantitative Study of Sound Change in Progress*. Philadelphia, PA: U.S. Regional Survey.

Levitt, S. D., and Dubner, S. J. (2006). *Freakonomics*. New York: Harper Trophy.

Lewis, David K. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Wiley-Blackwell.

Lewis, David K. (1975). Languages and language. In Keith Gunderson (ed.), *Minnesota Studies in the Philosophy of Science* (pp. 3–35). Minneapolis: University of Minnesota Press.

Lewis, David K. (1980). Index, context, and content. In Stig Kanger and Sven Öhman (eds.), *Philosophy and Grammar* (pp. 79–100). Dordrecht: Reidel.

Lewis, David K. (1992). Meaning without use: Reply to Hawthorne. *Australasian Journal of Philosophy*, 70(1):106–110.

Marmor, Andrei (1996). On convention. *Synthese*, 107(3):349–371.

Marmor, Andrei (2009). *Social Conventions: From Language to Law*. Princeton, NJ: Princeton University Press.

Millikan, Ruth Garrett (1998). Language conventions made simple. *Journal of Philosophy*, 95(4):161–180.

O'Connor, Cailin (2019). *The Origins of Unfairness: Social Categories and Cultural Evolution*. Oxford: Oxford University Press.

Quine, W. V. O. (1969). Foreword. In David K. Lewis (ed.), *Convention: A Philosophical Study* (pp. xi–xii). Cambridge, MA: Wiley-Blackwell.

Russell, Bertrand (1921). *The Analysis of Mind*. London: Allen and Unwin.

Schelling, Thomas (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Skyrms, Brian (1996). *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Stern, Josef (2000). *Metaphor in Context*. Cambridge, MA: MIT Press.

Szabó, Z. (2019). Semantic explanations. In E. Lepore and D. Sosa (eds.), *The Oxford Studies in Philosophy of Language* (pp. 240–276). Oxford: Oxford University Press.

Trudgill, Peter (1972). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Van Herk, Gerard (2008). The Northern Cities Shift as linguistic white flight. University of Pennsylvania Working Papers in Linguistics 14(2): Article 19.

Weisberg, Michael (2007). Three Kinds of Idealization. *Journal of Philosophy*, 104(12):639–659.

Weisberg, Michael (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

3
GRICE

3.1 Overview

Though Paul Grice's groundbreaking work in the theory of language and communication has many facets, it can be roughly separated into two philosophical projects: that of explaining speech production and interpretation in terms of general principles of rationality and cooperation, and that of defining expression meaning in terms of speaker meaning.[1] As Neale (1992) notes, attention to Grice's work tends to focus on his theory of speech production/interpretation and its implications for issues at the semantics/pragmatics interface, at the expense of his theory of expression meaning and its connections to metasemantics. This might be partially explained by the fact that Grice explored the latter only insofar as it helped elucidate the:

> My primary aim is rather to determine how any such distinction between meaning and use is to be drawn, and where lie the limits of its philosophical utility … Any serious attempt to achieve this aim will, I think, involve a search for a systematic philosophical theory of language, and I shall be forced to take some tottering steps in that direction.[2]

Most of his efforts were spent refining the notion of speaker meaning in response to ongoing criticisms, and to developing concomitant notions such as *implicature* and *what is said*. His ideas about how expression meaning reduces to speaker meaning were left as more of a rough sketch, and perhaps thus lost their due place in in the popular conception of Grice's

philosophical legacy; however, they were taken up and developed by theorists working on foundational issues in the philosophy of language—including Stalnaker and Lewis, among others—who cite Grice as a major influence.[3]

Stalnaker, for example, inherited from Grice his adherence to a methodology according to which pragmatic explanations of linguistic phenomena are preferred to those that complicate the semantics. The justification for this approach is that, "if the sense is predicted by the pragmatics, then adding it to the semantics does no explanatory work".[4] Grice's theory of meaning and interpretation provided a blueprint for this methodology; by marking a clear distinction between speaker meaning and expression meaning, we can adopt a pared down semantics while explaining linguistic phenomena in terms of speaker meaning together with general principles of rationality.

While his ideas about the semantics/pragmatics boundary were a major influence on Stalnaker's work, Grice's ideas about the foundations of expression meaning are reflected in Lewis' work on linguistic conventions.[5] On this picture, meaning and interpretation are characterized in terms of the actions and mental states of language users, which can be specified independently from conventional features of language (in accordance with the principle Autonomy of Pragmatics). Conventional language use, in turn, can be explained in terms of certain kinds of regularities in those actions and mental states.[6] Thus, Grice's approach to linguistic theory conforms to what I've been calling a *Pragmatics First* methodology: Public language is reducible to pragmatic facts, which can be explicated in terms of the mental states, behaviors, and relations between language users.

Another methodological approach that Grice shares with Lewis and Stalnaker is his adoption of an idealized picture of language as cooperative information exchange.[7] To see how this idealization functions in Grice's work, we'll consider its component parts in turn, starting with cooperation. On Grice's view, cooperation is at the heart of successful communication, playing a crucial explanatory role in both speech production and communication. On the side of production, this role comes out in Grice's definition of speaker meaning (of which I use a simplified version):

> *Grice Meaning*: In uttering U, S means something iff for some audience A, S utters A reflexively intending to produce a belief in A at least partially on the basis of A's recognition of this intention.

A reflexive intention is self-referential in the sense that part of the intention is for that very intention to be recognized.[8] The term *reflexive intention* doesn't come from Grice, but it allows for a less cumbersome version of Grice's theory, which he formulated by appealing to a series of

higher-order intentions: 1) the speaker intends to produce a belief in A, 2) intends that the audience recognize that (1), 3) intends that the audience recognize that (2), *ad infinitum*. The idea here is that the speaker has no "sneaky" intentions; i.e., no intention to deceive the audience about one of her lower order intentions. Grice recognizes that his account is idealized in the sense that for reasons of cognitive limitations, the speaker's higher-order intentions may not be infinite. The main point is that, however far they go, there will not be any intentions to deceive the audience among them. According to Grice's account, then, speech production is essentially cooperative in the sense that acts of meaning are transparent in nature. Speakers do their best to help their audience recognize what they are up to in performing such acts, and rely on this intention-recognition to achieve their communicative goals; on this account, recognition of the speaker's intention provides the audience with a reason to perform the intended response.[9]

Cooperation also plays a central role in Grice's theory of interpretation. He suggests that interlocutors operate under the assumption that they are all following the *Cooperative Principle*—a norm according to which one must make their conversational contribution conform to what is required by the mutually accepted purpose of the conversation. (Though the purpose of the conversation could in principle be anything, Grice tends to assume that it will normally be information exchange—more on this below.) This norm both guides speech production and facilitates successful interpretation; audiences assume that speakers are following it and interpret them in such a way as to be consistent with this fact. When a speaker appears to be violating the Cooperative Principle by what she says, for instance, audiences will assume that she is indirectly communicating something that in fact adheres to it. This idea forms the basis of Grice's famous theory of conversational implicature. A speaker can say one thing while implicating another, and the Cooperative Principle allows the audience to identify the implicated content.

In arguing against the usefulness of the distinction between ideal and non-ideal theory in philosophy of language, Cappelen and Dever (forthcoming) suggest that Grice's theory (and indeed all idealized theories) is simply one of limited scope; it is not a theory of communication in general, but of communication in cooperative contexts. This is a misrepresentation of Grice's work, which drastically understates his rich and ambitious project. The scope of Grice's account was significantly more broad; the phenomenon that he wanted to explain was rational communication.

> It is the rationality or irrationality of conversational conduct which I have been concerned to track down rather than any more general characterization of conversational adequacy.[10]

He not only took rational communication to form the bulk of our communicative exchanges, but to play a crucial role in explaining how conventional language could get off the ground in the first place (more on this below). The connection to cooperation is as follows: while linguistic meaning is underpinned by rational communication, rational communication is, in turn, underpinned by cooperation. Grice claims that the *Maxims of Conversation*, which flesh out the demands of the Cooperative Principle, "are such that in paradigmatic cases, their observance promotes and their violation dispromotes conversational rationality".[11] If the scope of the theory were cooperative communication rather than rational communication, this appeal to cooperation in explaining the phenomenon at issue would be significantly less interesting and informative. Thus, Grice's theory involves a minimalist idealization, which introduces a deliberate restriction to only those features that are thought to play a role in metaphysically determining the target phenomenon; his thought was that by restricting our attention to cooperative contexts, we can uncover the essential features of the subject of inquiry, namely rational communication as a whole.

The second element of Grice's idealized picture of language involves thinking of linguistic communication as aimed at information exchange. In contrast to the element of cooperation, the element of information exchange seems to function more as a Galilean idealization in Grice's theory—however, there are glimpses of minimalist idealization here as well. Like Stalnaker and Lewis, Grice seems to consider contexts of information exchange to provide a reasonable starting point for theorizing about linguistic communication. At some stage, however, a theory that is initially designed to characterize such contexts must be adapted to cover other types of speech; this strategy can be seen in his early definition of speaker meaning. Here, again, is a simplified version of that definition:

GRICE MEANING: In uttering u, S means something iff for some audience A, S utters u reflexively intending to produce a belief in A at least partially on the basis of A's recognition of this intention.[12]

Here we see the conception of communication as information exchange in the characterization of the speaker's goal—she is attempting to get the audience to believe something, which is generally how we go about exchanging information. But Grice recognizes that there are other uses of language, and so refines the definition in a way that "drop[s] the pretense that we have to deal only with "informative" cases".[13] Here is a simplified version:

GRICE MEANING REVISED: In uttering u, S means something iff for some audience A, S utters u reflexively intending to produce some response in A at least partially on the basis of A's recognition of this intention.

The revised account is broadened to include an open-ended range of responses, rather than just belief (more on this below). Thus, in accordance with the methodology of Galilean idealization, Grice initially focuses on cases of information exchange in order to develop a working definition, and then systematically adjusts it in order to cover a broader range of cases.

The conception of language as information exchange also comes out in Grice's formulation of the Maxims of Conversation, which include the following:

QUANTITY: Be as informative as you possibly can, and give as much information as is needed, and no more.
QUALITY: Be truthful, and do not give information that is false or that is not supported by evidence.
RELATIO: Be relevant, and say things that are pertinent to the discussion.
MANNER: Be as clear, as brief, and as orderly as you can in what you say, and avoid obscurity and ambiguity.[14]

The maxims of Quantity and Quality seem directly related to the goal of information exchange, in that they instruct the speaker to give the information required, and to provide true and well-supported information. The maxims of Manner and Relation instruct the speaker to make the process of information exchange efficient by making their contributions clear and relevant. Grice notes the connection between these maxims and the particular conversational goal of information exchange:

> The conversational maxims … are specially connected (I hope) with the particular purposes that talk (and so, talk exchange) is adapted to serve and is primarily employed to serve. I have stated my maxims as if this purpose were a maximally effective exchange of information; this specification is, of course, too narrow, and the scheme needs to be generalized to allow for such general purposes as influencing or directing the actions of others.[15]

This quote suggests that Grice is thinking of his account of the maxims of conversation as involving Galilean idealization. They are stated as though the purpose of linguistic communication were to exchange information, but ultimately the theory will need to be adapted to cover other purposes.

In other remarks, however, Grice seems to indicate that the conception of language as information exchange is also functioning as a minimalist idealization; it is intended to capture the essential features of the phenomenon under exploration, namely rational communication. In *Meaning Revisited*, he provides an explanation for why language would be a beneficial tool for rational creatures. Assuming that languages contain expressions that

correspond to types of psychological states, an utterance of such an expression could facilitate the transfer of a type of psychological state from the mind of one rational agent to that of another. Here's where the central role of information exchange comes in: Grice seems to be thinking of the relevant psychological state as belief, and suggests that such transfers would be beneficial only if the beliefs in question were true. That is, only if they facilitated information exchange:

> If all these transfers involved mistaken beliefs, it is not clear that one would regard this communication mechanism as beneficial, even if the appropriate psycholinguistic correspondences held … so a general condition would be that soul-to-soul transfers, so to speak, are beneficial provided that states transmitted are ones which correspond with the world.[16]

Though at times the conception of language as information exchange seems to function merely as a Galilean idealization, these remarks reveal a hint of minimalist idealization as well. They suggest that information exchange is not just one among many uses of language, but rather a privileged type of use that ultimately explains why language would be a useful tool for rational creatures like ourselves.

Below we'll explore various problems for Grice's theory of linguistic communication insofar as it relies on the conception of language as cooperative information exchange. Though I do think that the element of information exchange is problematic for Grice, I focus my attention mainly on the element of cooperation, and its relation to speech production and interpretation.[17] The problems I discuss have been raised by others, and so in this sense they are not new. However, it is important for me to highlight them here in order to show in Chapter 7 that my de-idealized proposal is able to overcome them. Moreover, I attempt to strengthen these critiques by responding to potential replies that the Gricean might be tempted to offer.

3.2 Interpretation

Asher and Lascarides (2013) have noted that, contrary to the Gricean picture, we seem to be able to interpret implicatures without relying on assumptions of cooperation. Recall that according to Grice, implicatures are calculated in the following way: The audience assumes that the speaker is following the Cooperative Principle and its associated maxims of conversation. That is, she assumes that the speaker will make her contribution such as required for the mutually accepted purpose of the conversation. If the speaker says something that appears to violate this principle, the audience will interpret her as implicating something that adheres to it. (Grice takes "what is said"

to be "closely related to the conventional meaning of the words" uttered, however the details need not concern us here.[18]) The assumption of cooperativity, then, is necessary for interpreting implicatures on this account:

> To assume the presence of a conversational implicature, we have to assume that at least the Cooperative Principle is being observed … to calculate a conversational implicature is to calculate what has to be supposed in order to preserve the supposition that the Cooperative Principle is being observed.[19]

But as many have pointed out, implicatures can be calculated even in the absence of the assumption of cooperation.[20] Conversational contexts including courtooms, political speeches, and marital spats are rife with examples. They are not only used in critiques of the Gricean theory of implicature, but are also commonly used in speech act theory in discussions of the lying/misleading distinction.[21] Asher and Lascarides (2013) provide the following example (borrowed from Solan and Tiersma (2005)):

BRONSTON: During a cross-examination of a defendant (Bronston) by a prosecutor, the prosecutor asks if he has ever had any bank accounts in Swiss banks, to which Bronston replies, "The company had an account there for about six months, in Zurich."

With this answer, Bronston implicates that he never had a Swiss bank account, which is false. On the Gricean account, recall, implicatures are interpreted by assuming that the speaker is following the Cooperative Principle. If the speaker says something that violates it (by violating one of its maxims), she must be implicating something that adheres to it. But this is not how the prosecutor could have calculated the implicature. What the speaker says violates the maxim of *Relevance*, in that it fails to directly answer the question. So, the prosecutor looks for a relevant implicature. So far, this process conforms to the Gricean model. However, the implicature that Bronston has never had a Swiss bank account violates *Quality*. The prosecutor knows this, and yet this is the implicature that he attributes to Bronston. At this stage, the Gricean story breaks down; the prosecutor attributes this implicature to Bronston not because he assumes that Bronston will offer a conversational contribution (at the level of what is meant, if not what is said) that contributes to their *shared* goals, but rather one that serves Bronston's own goals. Here, Bronston's goal is to contribute the (false) information that he never had any accounts in Swiss banks, so as to avoid conviction. Moreover, by contributing this information indirectly rather than directly, he is further serving his own interests by making it less likely that he will be convicted of perjury.

So, contrary to the Gricean story, we can interpret implicatures in non-cooperative contexts. Moreover, assumptions about conflicting—rather than shared—goals seem to play a role in the interpretative process. This problem for the Gricean model of implicature has been taken up by a number of linguists—especially those interested in game-theoretic models of communication—who have developed alternative accounts that don't rely on assumptions of cooperation (more on this below).[22] However, the entrenchment of the idealized conception of language as cooperative information exchange is revealed in the fact that theorists in mainstream philosophy of language have continued to rely primarily on the cooperation-dependent model of implicature in spite of the fact that alternative models have become available.

One reason for the continued reliance on this model may be that some philosophers don't consider the examples from non-cooperative contexts to pose a genuine threat. Below I consider a few potential responses which I've heard in personal conversations with proponents of the Gricean picture, and which are suggested in Grice's own work.

The first response goes as follows: Sure, the Gricean theory of implicature may not be an apt description of how we interpret implicatures in non-cooperative contexts. But these contexts are peripheral in the sense that cooperative contexts are the "default"; the Gricean theory is the correct theory for default contexts. This line of thought seems to appeal to minimalist idealization: The cooperation-dependent model of implicature earns its keep by correctly describing how we interpret implicatures in central, or "default" cases. Perhaps an alternative theory will be needed to explain the mechanism of interpretation in non-default cases, but such cases are of peripheral interest. By capturing the process of interpretation in default contexts, the Gricean theory is able to characterize its most important or essential features.

The problem with this response is that it is rarely accompanied by clarification of what is meant by "default", or support for the claim that cooperative contexts qualify as such. One idea might be that cooperative contexts are simply the most common. That cooperative contexts are more common than non-cooperative ones, however, is far from obvious; without supporting empirical data, it is difficult to take the claim seriously. Another idea is that cooperative contexts are default in the sense that interlocutors assume that a context will be cooperative unless given reason to think otherwise. But again, this a non-obvious empirical claim that would require support. For instance, we might expect under this assumption that implicatures made in non-cooperative contexts would take longer to process. Moreover, this picture appears to require a more complex account which posits one type of mechanism for implicature in default contexts and another for non-default contexts. Without principled reasons in its favor, such complexity

is unjustified—especially when there are alternative, unified theories on the market. In accounts developed by Asher (2012), and Asher and Lascarides (2013), e.g., implicatures are calculated by reasoning about the speaker's goals. In cooperative contexts, these goals happen to be shared with the audience and in non-cooperative contexts they are not; the way that beliefs about them help audiences to calculate implicatures, however, is the same in either case.

A second response is as follows: The alleged cases of implicature in non-cooperative contexts are misdescribed. While there may be elements of conflict in such contexts, they can also be characterized—at least on some level—in terms of cooperation. Thus, they are not real counterexamples to the Gricean model of implicature. This response was put forward by Grice himself, who considered the foregoing objection to his view.

> … we should recognize that within the dimension of voluntary exchanges (which are all that concern us) collaboration in achieving exchange of information or the institution of decisions may coexist with a high degree of reserve, hostility, and chicanery and with a high degree of diversity in the motivations underlying quite meager common objectives.[23]

Grice and his followers are certainly correct in pointing out that conversation, like most any other joint activity, can involve a mixture of cooperation and conflict. What the Gricean needs are principled constraints that would allow us to mark the distinction between cooperative and non-cooperative conversational contributions. One might be tempted to suggest that there is at least some way in which the speaker is cooperating if she bothers to respond at all, however the following example from Asher (2012) illustrates the problem with this strategy:

NEW YORK: A pedestrian stops a stranger on the street and asks, "Excuse me. Could you tell me the time please?" To which the stranger replies, "Fuck you!"[24]

To suggest that the stranger's reply is cooperative is to stretch the ordinary notion of cooperation too far; it provides so weak a constraint as to render the notion vacuous. Responsiveness alone, then, is ill-suited to characterize the minimal level of cooperation that underlies the Gricean theory.

Another strategy for characterizing cooperation comes from the work of David Lewis. Both Grice and Lewis take linguistic conversation to be a type of strategy problem in that interlocutors are involved in interdependent reasoning: each person's move depends on what they expect the other to do. As discussed in Chapter 2, strategy problems can range anywhere along

a scale from pure cooperation—where players gain and lose together—to pure conflict, where one person's loss is another's gain. Grice notes that conversations need not fall at the purely cooperative end of the scale in order to be characterized as cooperative, and this seems right. Lewis was able to accommodate this thought by his notion of "predominantly cooperative" contexts. Such contexts admit some level of conflict, but deserve the characterization of "cooperative" in virtue of the fact that they fall closer to the cooperative end of the scale, even if they are not always instances of the limiting case. Formally, such contexts are those in which the difference between agents' payoffs for any particular option aren't bigger than the difference between payoffs for different options.[25]

Lewis' proposal seems to be approximately what we are looking for: it allows that strategy problems can involve a mixture of cooperation and conflict, and provides principled constraints characterizing those mixtures that may reasonably be described as "cooperative". Nonetheless, it's too strong to be of much use to a Gricean who wants to insist that even the most meager amount of common objectives will do. Lewis' game theoretical framework also reveals the incongruity of such a position; if strategy problems may fall anywhere along a scale from pure cooperation to pure conflict, surely it is not apt to characterize every problem as cooperative which is not an instance of the limiting case of a zero-sum game.

A final problem for this response is that the Gricean needs more than just a principled way to characterize apparently non-cooperative conversations as cooperative; where there is a mixture of cooperation and conflict, the Gricean must also show that the cooperative elements are doing the explanatory work. But this does not seem plausible: even if the case of *Bronston* does involve some level of cooperation, it is the assumption of conflict that allows the prosecutor to calculate Bronston's implicature. That is, even if Bronston and the prosecutor share some goals, the prosecutor correctly interprets Bronston's implicature by reasoning about their *conflicting* goals. Bronston wants to avoid revealing the truth and getting convicted, while the prosecutor is aiming for this result; the prosecutor is aware of this and uses it as a basis for calculating the false implicature.

The data about implicatures in non-cooperative contexts should not be surprising. If conversation is a kind of strategy problem, as Grice and others would have it, then there is no independent reason to think that *rational conversations* in particular are characterized by cooperation. This would make conversations disanalogous to other sorts of strategy problems, which decades of research in game theory have shown to be guided by general norms of rationality and decision making regardless of where they fall along the scale between pure cooperation and pure conflict. There is no reason, other than a preconceived picture of language as cooperative information exchange, that we should expect linguistic communication to

be different in this regard. Rather, we should expect that we interpret others' conversational moves the same way that we interpret their moves in any other type of strategy problem: by reasoning about their information and their goals—regardless of whether these goals are shared by other conversational participants. If this is the case, then Grice was right to think that pragmatic reasoning is governed by more general norms of rationality and decision making, but wrong in restricting this set of norms to those governing cooperative contexts in particular.

3.3 Speech production

In the last section we focused on problems that the idealization of cooperativity caused for the *audience*-side of Grice's theory of communication. In this section we'll see that this idealization also creates problems for the *speaker*-side. According to Grice's theory of speaker meaning, speakers' communicative intentions are overt, in that they are intended to be recognized. This implies that acts of communication must be at least minimally cooperative; even if the speaker is being uncooperative on some level (e.g., lying to or insulting her audience), she must be cooperative in the sense of trying to let the audience know what her communicative intentions are. Below I discuss different kinds of speech acts that do not fit this mold. As in the last section, I draw on arguments already present in the literature. Reiterating them here, however, allows us to get a better grip on the range of phenomena that a theory of meaning must be able to accommodate, and to better understand how the assumption of cooperativity poses obstacles to this project.

3.3.1 Insinuation

One type of speech act that fails to fit with Grice's model of communication is insinuation. In the Gricean picture, saying and implicating are both species of speaker meaning, which differ in that the former bears a tighter connection to the literal content of the utterance. They are similar in that—given that both are species of meaning—both what is said and what is implicated must be overtly communicated. As Camp (2018) has noted, however, insinuations do not appear to fall neatly within either of these categories. They are not part of what is said, given that their content fails to bear the right relationship to the literal content of the utterance. On the other hand, they don't seem to be implicated, given that the speaker's intentions are apparently not fully overt. Insinuations are peculiar in the sense that the speaker has both an interest in letting the audience know what she is up to, and in preserving the ability to deny it. Here is an example from Lee and Pinker (2010) to illustrate:

INSINUATED BRIBE: "I'm very sorry, officer. But I'm actually in the middle of something right now, sort of an emergency. So maybe the best thing would be to take care of this here … without going to court or doing any paperwork."[26]

To the extent that the speaker wants to succeed in bribing the police officer, she has an interest in letting him know that this is what she intends. However, there is also a sense in which she has an interest in keeping these intentions hidden; if the officer is not amenable to the bribe, she will need a way to deny that she ever intended to bribe him. Thus, she must perform a balancing act by making her communicative intention recognizable, but in a way that is cloudy enough to allow her to retain plausible deniability in the case that things go south.

Do insinuations, then, pose a counterexample to Grice's theory of meaning? Camp argues that they do not; contrary to first appearances, they involve overt communicative intentions:

> … it appears that insinuating speakers do intend to produce cognitive effects in their hearers, and that these intentions play some sort of important role in getting the hearer to entertain [the insinuated content].[27]

Though Camp recognizes the existence of completely covert communicative acts (including subliminal advertising and dogwhistles), she thinks that insinuations are not of the same kind. Their characteristic feature is that while they are overt, they are "off record" in the sense that the speaker is not prepared to own up to the insinuated content. Camp claims that they do not, however, involve the sort of "veiled higher-order intentions" that would preclude them from qualifying as overt speech acts.[28]

Camp is correct to note that insinuations involve at least some level of overtness; in the case of Bribe, for instance, we noted earlier that the speaker will not be able to succeed in bribing the policeman unless he is able to recognize her intention to do so. But it is less clear that insinuations are completely overt, in the sense of being devoid of any veiled higher-order intentions whatsoever. Lee and Pinker (2010) compared audience reactions to Insinuated Bribe with the following example of a direct bribe attempt:

DIRECT BRIBE: "I'm very sorry, officer. If I give you a fifty, will you just let me go?"[29]

Study participants identified the speaker as intending to bribe the officer in both cases. However, in the case of *Direct Bribe* they reported certainty about these intentions, while they assigned them a lower probability in the case of *Insinuated Bribe*.

This result coheres well with Dinges and Zakkou's (forthcoming) account of plausible deniability. As Camp notes, one of the distinguishing features of insinuations is that (when successful) the speaker retains plausible deniability about what she intended to communicate. In Camp's view, plausible deniability is compatible with complete overtness and the subsequent full recognition of the speaker's intentions; while the speaker denies her intention to have communicated the insinuated content, this intention may nonetheless be common knowledge between her and the audience. But this picture does not capture the difference in the audience's level of certainty about the speaker's intentions in the cases of *Insinuated Bribe* vs. *Literal Bribe*. According to Dinges and Zakkou, in contrast, the speaker has plausible deniability in case recognition of her intentions would fail to qualify as knowledge were the speaker to deny them.[30,31] (Dinges and Zakkou's counterfactual account allows for cases in which the audience knows what the speaker intended; in such cases, the speaker retains deniability because *if she were to issue a denial*, it would be sufficient to undermine that knowledge. For simplicity, I focus on the narrower range of cases in which the audience fails to have knowledge even without the speaker's denial.) On this picture, the audience may believe correctly that the speaker of *Insinuated Bribe* intended to offer a bribe, but if the speaker is to retain plausible deniability, she must plant enough doubt in the mind of the audience to ensure that this belief does not rise to the level of knowledge.[32] If retaining plausible deniability is a characteristic aim of insinuation, then insinuations are overt to the degree that the speaker intends that the audience recognize her intention, but covert to the degree that this recognition is intended to involve a certain level of doubt. This result is borne out in the audience reactions to Lee and Pinker's (2010) examples.

What we can learn from insinuations, then, is that overtness is not an all-or-nothing affair. Some level of overtness is compatible with some level of veiled higher-order intentions. The Gricean account assumes that acts of meaning are fully overt in the sense that the speaker has nothing to hide from her audience with regard to her communicative intentions. Insinuations suggest otherwise; while the speaker intends her audience to recognize her communicative intentions, she aims to keep them veiled enough to preserve plausible deniability. That is, she intends that the audience come to have some level of positive credence regarding her communicative intentions, without this credence rising to the level required for knowledge.

3.3.2 Dogwhistles

Following linguist Kimberly Witten (n.d.), we can think of dogwhistles as speech acts that are intentionally designed to allow for at least two plausible interpretations. One of these interpretations is private in the sense that it is targeted at a subset of the general audience and intended not to be

recognized by audience members as a whole. In her helpful discussion of dogwhistles, Saul (2018) notes that they may be categorized along two different dimensions: overt vs. covert, and intentional vs. unintentional. Like Saul, I'll focus my discussion here on intentional covert dogwhistles—those that are intentionally meant to produce a response in some targeted audience, and where the speaker intends that the targeted audience fail to recognize this intention.[33] As Saul notes, these kinds of dogwhistles are problematic for the Gricean account of speaker meaning. On Grice's account, speakers not only intend for the audience to recognize their intention to produce a response, but for this recognition to function as a *reason* for the audience to produce that response. Covert intentional dogwhistles fail to conform to this picture because the speaker neither intends for the audience to recognize their communicative intention, nor for such recognition to function as a reason for producing the intended response. On the contrary, as Saul points out, the audience's recognition of the speaker's intentions can tend to undermine the speaker's goal of producing the desired response.

White (2007) (cited in Khoo (2017)), for example, found that study participants harboring racist stereotypes (measured by a questionnaire) were more likely than others to oppose food stamp programs after reading "… programs like food stamps and Medicaid represent important safety-nets for many inner city families, keeping them from falling deeper into poverty". They did not, however, have this response after reading the same sentence with the phrase "inner city" replaced by "American" or "poor". This suggests that "inner city" is functioning as a dogwhistle which acts on racial stereotypes against Black Americans and promotes racist behavior. Interestingly, however, the same participants did not have this response after reading the same sentence with the dogwhistle replaced by "African Americans".[34]

Another example comes from Mendelberg (2001) (cited in Khoo (2017) and Saul (2018)).[35] In 1988, George H. W. Bush's campaign against Michael Dukakis ran an ad criticizing the prison furlough program that was in place during the latter's tenure as governor. The ad, intended to present Dukakis as soft on crime, features a story about a man who was granted a furlough under this program. During his furlough, William Horton (referred to as "Willie" in the ad, though he never went by that name) assaulted a white couple in their home, raping the woman. Though the ad never explicitly mentions race, it featured pictures of Horton, who is Black. It's frequent airtime on the news preceded a fall in Dukakis' poll numbers. In particular, data from the campaign shows that increased exposure to the ad correlated with an increased likelihood of racially resentful voters favoring Bush. As in the last case, however, making the racist intention of the ad explicit appeared to dull its effect; as Jesse Jackson began to publicly criticize the ad as racist, the correlation declined.

Some have suggested that the coded messages of dogwhistles are part of the semantic content of the utterance used. Stanley (2015), for instance, suggests that they are part of the "not-at-issue" content, while Mendelberg (2001) suggests that dogwhistles are ambiguous, with the coded content serving as one of the available meanings. Khoo (2017), however, shows that dogwhistles fail established tests for both ambiguity and not-at-issue content, suggesting that such contents are not semantically encoded. But neither do they qualify on the Gricean picture as impicatures, given that these messages are covert. The empirical evidence suggests that when the intention to communicate the coded message is recognized, the ultimate communicative intention to produce the desired effect on the audience (e.g., their voting behavior/support for certain policies) is undermined. Rather than being produced by the audience's recognition of the speaker's intentions, Khoo suggests that these effects are a result of surreptitiously exploiting audience's stereotypical beliefs. Covert dogwhistles lead the audience to draw inferences from speech without it being clear to them that they are doing so—nor that the speaker intends them to do so. As Saul notes,

> Any theory which includes the idea that uptake (recognition of the speaker's intention) is required for success will fail entirely as a way of accommodating implicit dogwhistles. Implicit [i.e., covert] intentional dogwhistles only succeed where uptake is absent; uptake prevents such dogwhistles from being effective.[36]

This suggests that a viable pragmatic theory must include covert speech acts—those that are intended to produce their result not on the basis of intention-recognition, but on the basis of surreptitiously exploiting attitudes and inference patterns of audience members.

3.3.3 Propaganda

Propaganda offers another example of speech that doesn't fit the paradigm of Gricean communication. According to Stanley's (2015) understanding, propaganda is speech that is presented as an embodiment of certain ideals, and tends to either support or undermine those same ideals through emotional or non-rational means.[37] Stanley's definition allows that this effect may be intended or not. Again, I'll focus my discussion on cases in which it is intended.

One thing to note here is that propaganda may, but need not, include dogwhistles. That is, propaganda may in some cases function, as Khoo suggests, by using certain words that are expected to elicit certain responses or inference patterns. In other cases the mechanism is similar, but the inference

may come not from words but from bundles of information. Fraser (2018) provides a nice discussion on how this may occur with certain uses of metaphor. On the structure view of metaphor, interpretation involves drawing on a source domain and mapping it to a target domain. We can consider an example from Tirrell's (2012) discussion of the Rwandan genocide to see how this works:

RWANDA: The Tutsi are *inyenzi*.

Here, the word "inyenzi" (*cockroaches* in Kinyarwanda) is not a dog-whistle—the sentence does not have its effect by operating on pre-existing associations between Tutsis and cockroaches, but rather *creates* those associations through the use of metaphor. According to the structure mapping model, there are two mechanisms involved in metaphor interpretation. One involves comparing the structures of the source and target domain and extracting an eligible mapping. Audiences, however, have a preference for a rich mapping—and so if one is not available, the audience may restructure either the target or the source domain in order to achieve it. This process may involve increasing the salience of aspects of the structured bundle that fit with the mapping and decreasing the salience of those that don't. In this example the source domain is the Tutsi people and the target domain is cockroaches. In order to understand the metaphor, the audience may restructure the target domain in a way that will provide a better mapping to the information associated with the source domain. As Tirell points out, the information associated with cockroaches in Rwanda includes that

> … cockroaches are pests, dirty, ubiquitous, multiply rapidly, are hard to kill, *ought* to be killed, show emergent tendencies when in groups, are resilient, carry diseases, can go long periods without food or water, tend to only emerge at night when they are hard to see.[38]

The repeated use of the metaphor, she argues, resulted in a reconfiguring of the information associated with the Tutsi people that mapped it onto the set of information associated with cockroaches—eventually contributing to the Rwandan genocide.

Such cases of propaganda are further instances of communicative goals that produce their intended effect not by intention-recognition, but by surreptitiously creating and exploiting automatic inferences. Insofar as we want our pragmatic theory to be able to illuminate this kind of speech—and, like Stanley, Saul, and others, I think that it should—we need to broaden our view of speech acts in order to include covert communicative intentions.

3.3.4 Advertising

Covert speech acts are also frequently used in the domain of advertising. Often, the ultimate goal of the communicative act depends for its success on the audience's failing to recognize that goal. None of us want to feel like we are being manipulated, or that we are automatons whose behavior is at the mercy of big corporations. The covert speech used in advertising is not mutually exclusive with dogwhistles and propaganda, but rather can serve as instances of both.

We'll look at one such case which covertly promotes the ideology of patriarchy. This example comes from a highly criticized ad for Bic pens that ran in South Africa in 2015.[39] (It is worth noting that more subtly sexist ads are ubiquitous and often pass without comment.) This ad features an attractive and smiling young woman with her arms crossed and wearing a suit. She looks like a woman who is successful in business, in love, and in life. To her side is a large script reading:

> "Look like a girl
> Act like a lady
> Think like a man
> Work like a boss"

At the bottom of the page. near the Bic logo, is written "#HappyWomensDay". This is a clear case of what Stanley calls "undermining propaganda". That is, it is presented as an embodiment of certain ideals, while undermining those same ideals using emotional and non-rational means. It purports to embody the ideals of feminism, by presenting what looks like a healthy, happy, successful, and empowered woman, and also purports to encourage other women to empower themselves. Moreover, the ad was published on Women's Day, and the hashtag at the bottom signaled recognition of, and support for that occasion.

However, it covertly undermines the ideals of feminism while purporting to promote them. First, the message is written as a command to women, reinforcing their subordinate position in a gendered social hierarchy. While encouraging them to pursue success, it does this by reinforcing patriarchal norms. For instance, the message is that in order to be successful, women must not age—they lose their value when they stop looking like girls. Second, it suggests that in order to be successful, their behavior must conform to that of a "lady", which is a dogwhistle for restricted behavior that conforms to a patriarchal ideal: modest, subdued, deferential, putting others' needs before one's own, etc. Perhaps most offensive is the suggestion that in order to be successful, a woman must think like a man, which reinforces the patriarchal idea that men are naturally more rational than women. Finally, the idea that to be successful a woman must "work like

a boss" suggests that her only path to success is to work harder than men. In this respect it echoes the much-criticized "lean-in" advice of Facebook's chief operating officer Sheryl Sandberg, who advises women to be more assertive and ambitious in the workplace.[40]

Again, the intentions behind the ad are covert; the ad is meant to operate on internalized misogny to create feelings of, e.g., shame about looking one's age, guilt about "unladylike" behavior, insecurity about thinking like a woman, and inadequacy about one's work efforts and status. Such feelings, if the advertiser is successful, will prompt a woman audience member to buy the pen in hopes that this will somehow help her better conform to the ideal expressed in the ad—a goal which would be undermined by the woman's recognition of this attempt to manipulate her feelings and behavior in ways that further entrench her subordinate position in a gendered hierarchy.

3.4 Conclusion

As we saw at the beginning of the chapter, the idealization of language as cooperative information exchange functions as both a minimalist and a Galilean idealization in Grice's work. My criticisms here were focused on the cooperative element, which—I argued—functions more as a minimalist idealization; Grice thinks that it captures the essential features of the target phenomenon, which is rational communication.

This idealization shows up in at least two places: the theory of speech production and the theory of speech interpretation. On Grice's account, audiences are able to interpret implicatures by assuming that the speaker is following the Cooperative Principle. This account wrongly predicts that implicatures cannot be interpreted in non-cooperative settings. False predictions, however, are not necessarily problematic for an idealized account—it depends upon the type of idealization in question. Grice is working with a minimalist idealization, which attempts to capture the essential features of the target phenomenon, namely rational communication. If there were reason to suppose that only implicatures made in cooperative contexts were characteristic of rational communication, then the theory would be adequate to its purposes in spite of its false predictions about non-cooperative contexts. I have argued, however, that there is no reason to suppose that such contexts are irrelevant to explaining the target phenomenon. Implicatures made in non-cooperative contexts appear to be ubiquitous; they are important for theoretical projects such as delineating the lying/misleading distinction, as well as social/political projects such as understanding and mitigating against propaganda and other forms of manipulative speech. Moreover, as Asher and Lascarides have argued, implicatures can be calculated in non-cooperative settings in broadly the same way as in cooperative settings. In both types of context, the calculation will exploit general

principles of rationality; the difference is merely in the input to those calculations. If she reads the situation correctly, then in the former case the audience will assume that her goals are misaligned with the speaker, while in the latter case she will assume that they are aligned. Thus, there are *prima facie* reasons for thinking that both non-cooperative and cooperative settings generate important data for a theory of rational communication. The restriction to cooperative settings in the name of minimalist idealization appears to be unjustified.

While I think that the idealization of cooperation served as a minimalist idealization for Grice, perhaps the problems for his theory of interpretation could be ameliorated by reconstruing it as Galilean idealization. Weisberg notes that this is a common practice in science: models may undergo various alterations as they are tested and developed. Some of these alterations may involve a change, e.g., in the target phenomenon: a model that is originally intended as a model of x may fail in that regard, but in the process of discovering this failure, it may be found to be suitable as a model of y.[41] Similarly, a model may fail as a minimalist idealization in that it is found to be unable to capture the essential features of the target phenomenon—but it may function better when conceived of as a Galilean idealization. Under Galilean idealization, essential features not captured by the original model may be accommodated through systematic, minor alterations to it. Grice himself makes remarks that are suggestive of Galilean idealization, claiming that non-cooperative contexts are derivative of cooperative contexts in the sense that in the latter, conversational participants are simulating the former.[42] Thus, one might adopt the theory under a Galilean idealization, hoping that non-cooperative cases can be understood by adding a little bit of complexity (such as the idea of simulation) to the standing theory. Again, however, this response is in need of support. In order for this strategy to be viable, we would need empirical evidence that in non-cooperative contexts conversationalists are indeed mimicking cooperative contexts; otherwise, the strategy is unprincipled and functions only to save the theory. Moreover, Asher and Lascarides consider this strategy and argue that the idea of simulating cooperative contexts fails to characterize at least some examples of how implicatures are calculated in non-cooperative contexts.[43] All of this suggests that the idealization of cooperation fails both as a minimalist and a Galilean idealization in a theory of interpretation.

As we discussed above, the idealization of cooperation also shows up in Grice's theory of speech production insofar as he defines meaning in such a way that the speaker's intentions must be overt. While the speaker may be non-cooperative in other respects, the theory entails that she is cooperative at least to the extent that she is trying to let the audience know what she is up to; she has no "veiled" communicative intentions. Here, as in Grice's theory of interpretation, the idealization is minimalist; Grice thinks that overtness

is an essential feature of communication. One might be tempted to claim that Grice's account succeeds as a minimalist idealization for the following reason: while covert speech acts do involve language in some way, only overt speech acts are properly characterized as communicative. Khoo (2017), for instance, suggests that the coded contents of dogwhistles are not communicated at all. So, one might suggest that they—and the propagandistic metaphors that function similarly by exploiting automatic inferences rather than intention-recognition—are not communicative acts. Accordingly, they do not belong in a theory that seeks to capture only those features that are essential to the character and occurrence of communication.

The problem with this response is that "communication" is a contentious term, and we would need some principled reason for accepting the claim that covert speech acts aren't communicative. Perhaps one could appeal to Grice's distinction between *natural* and *non-natural meaning* here. Though we use the word "meaning" for both kinds of cases—e.g., "Those three rings of the bell means that the bus is full" vs. "Those spots mean measles", respectively—they carve out importantly different kinds of categories. For Grice, natural meaning corresponds naturally occurring correlations; non-natural meaning corresponds to communicative acts, and is the key to understanding language use. (The bus driver, but not the spots, performs a communicative act.) Covert speech acts, however, exhibit the characteristic features of non-natural meaning that Grice uses as a diagnostic test: (1) The utterance does not entail the relevant content; (2) a conclusion can be drawn concerning what was meant by the utterance; (3) it is coherent to say that somebody meant something by the utterance; (4) a restatement can be found in which the verb "mean" is followed by a sentence or phrase in quotation marks; (5) no restatement can be found beginning with the phrase "The fact that…".[44] For illustration, consider an utterance of, "Medicaid prevents inner city families from falling deeper into poverty," used to covertly mean that Medicaid prevents Black families from falling deeper into poverty. First, the content that Medicaid prevents Black families from falling deeper into poverty is not entailed by the utterance. Second, a conclusion can be drawn about what was meant by this utterance. Third, it is coherent to say that *Medicaid prevents Black families from falling deeper into poverty* was meant by the speaker. Fourth, it's felicitous to say that this utterance means that Medicaid prevents Black families from falling deeper into poverty. Finally, it's infelicitous to say that "The fact that this utterance has been performed means that Medicaid prevents Black families from falling deeper into poverty." So, if acts of covert meaning pass Grice's initial diagnostic test for non-natural meaning, why did he ultimately exclude them from his definition? This is because he had the intuition that we don't use the word "mean" (in its non-natural, communicative sense, which he labels *meaning$_{NN}$*) to apply to covert acts:

> I might leave B's handkerchief near the scene of a murder in order to induce the detective to believe that B was the murderer; but we should not want to say that the handkerchief (or my leaving it there) meant$_{NN}$ anything or that I had meant$_{NN}$ by leaving it that B was the murderer, Clearly we must add that, for x to have meant$_{NN}$ anything, not merely must it have been "uttered" with the intention of inducing a certain belief but also the utterer must have intended an "audience" to recognize the intention behind the utterance.[45]

While Grice's restriction might get the right intuitive result in the case of the handkerchief, it gets the wrong result with respect to the acts of covert meaning considered above. This suggests that Grice's requirement of overt intentions rules out too much. (More on this in Chapter 7.) Moreover, the phenomenon of insinuation shows that we are not simply dealing with limiting cases of completely overt/completely covert speech acts, which can be respectively characterized as communicative and non-communicative uses of language. Rather, speech acts exhibit varying degrees of overtness; thus, even though the distinction is of theoretical importance, a theory of linguistic communication that restricts its scope to the limiting case will necessarily be impoverished.

Notes

1 See Neale (1992) for a useful overview of Grice's work in the philosophy of language.
2 Grice (1989, p. 4).
3 See also Schiffer (1972) and Loar (1981).
4 Grice (1989, p. 45).
5 Lewis was also influenced by Stenius (1967) here. See Reiland (n.d.) for discussion.
6 Grice was hesitant to invoke conventions and appealed instead to a "procedure in the repertoires of language users". See Grice (1989, pp. 123–127). This detail won't matter for our purposes here, however.
7 See Cappelen and Dever (2019) for further discussion of these idealizations in Grice's work.
8 This notion was introduced by Searle (1969, p. 47) and used by Harman (1974, p. 225) and Bach and Harnish (1979, p. 15).
9 Grice allows for the possibility of lying and misleading. In such cases, however, she deceives the speaker about the content of what she means, rather than her communicative intentions. For instance, in meaning *p* by her utterance of *u*, a speaker may intend to deceive her audience about *p*, but not about her intention to communicate *p*.
10 Grice (1989, p. 369).
11 Grice (1989, p. 370).
12 Over the years, many revisions to Grice's account have been proposed. See Huang (2017) and Moore (2018) for overviews.
13 Grice (1989, p. 219).
14 Ibid., p. 28.

15 Ibid., p. 28.
16 Grice (1989, p. 288).
17 I develop this criticism in Keiser (2020).
18 See Grice (1989, p. 25).
19 Grice (1989, p. 34).
20 See, e.g., Solan and Tiersma (2005) Asher (2012), and Asher and Lascarides (2013).
21 See, e.g., Saul (2012) and Stokke (2018).
22 See, e.g., Asher (2012), Asher and Lascarides (2013), and Franke, Dulcinati, and Pouscoulous (2020).
23 Grice (1989, p. 369).
24 Asher (2012, p. 48).
25 Lewis (1969, p. 14).
26 Lee and Pinker (2010, p. 800).
27 Camp (2018, p. 55).
28 Ibid., p. 55.
29 Lee and Pinker (2010, p. 800).
30 This is a simplified account which relies on the knowledge norm of assertion. Dinges and Zakkou (forthcoming) offer a more general account according to which it would not be epistemically proper for the audience to treat their recognition of the speaker's intentions as a reason for actions or emotions, were the speaker to deny them. Like Dinges and Zakkou, I am not committed to the knowledge norm of assertion, but will rely on the simplified account here for ease of exposition. The same arguments can be made, *mutatis mutandis*, using the more generalized account.
31 See D'Ambrosio (n.d.-a) and D'Ambrosio (n.d.-b) for a game-theoretic model of this kind of speech.
32 Dinges and Zhakkhou (n.d.) allow that the audience's belief can fail to rise to the level of knowledge either because the level of credence is too low (psychological deniability), or because it is not justified by the evidence (evidential deniability). For simplicity I'll focus on the first case, though similar remarks may be made about the second case.
33 I discuss non-intentional speech in Chapter 7.
34 Similar effects of the dog whistle "inner city" are noted in Hurwitz and Peffley (2005) (cited in Saul (2018)).
35 See Mendelberg (2001, chapters 5–8).
36 Saul (2018, p. 373).
37 See Stanley (2015, p. 53).
38 Tirrell (2012, p. 200).
39 https://www.bbc.co.uk/news/world-africa-33893386
40 Sandberg and Scovell (2013).
41 See Weisberg (2013).
42 Grice (1989, p. 370).
43 Asher and Lascarides (2013, p. 8).
44 See Grice (1989, pp. 213–214).
45 Grice (1989, p. 217).

References

Asher, Nicholas (2012). The non cooperative basis of implicature. In D. Béchet and A. Dikovsky (eds.), *Logical Aspects of Computational Linguistics, volume vol. 7351 of Lecture Notes in Computer Science. LACL 2012*, Berlin Heidelberg: Springer.

Asher, Nicholas, and Lascarides, Alex (2013). Strategic conversation. *Semantics and Pragmatics*, 6(2):1–62.

Bach, K., and Harnish, R. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Camp, Elisabeth (2018). Insinuation, common ground, and the conversational record. In Matt Moss, Daniel Fogal, and Daniel W. Harris (eds.), *New Work on Speech Acts*, chapter 2. Oxford: Oxford University Press.

Cappelen, Herman, and Dever, Josh (2019). *Bad Language*. Oxford: Oxford University Press.

Cappelen, Herman, and Dever, Joshua (2021). On the uselessness of the distinction between ideal and non-ideal theory (at least in the philosophy of language). In *Routledge Companion to Social and Political Philosophy of Language*. New York: Routledge.

D'Ambrosio, J. (n.d.-a). A theory of manipulative speech. (unpublished ms)

D'Ambrosio, J. (n.d.-b). On manipulative underspecification. (unpublished ms)

Dinges, A. and Zakkou, J. (forthcoming). On Deniability. *Mind*.

Franke, M., Dulcinati, G., and Pouscoulous, N. (2020). Strategies of deception: Under-informativity, uninformativity, and lies—misleading with different kinds of implicature. *Topics in Cognitive Science*, 12(2):583–607.

Fraser, Rachel Elizabeth (2018). The ethics of metaphor. *Ethics*, 128(4):728–755.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Harman, G. (1974). Meaning. *The Journal of Philosophy*, 71(7):224–229.

Huang, Yan (2017). Neo-Gricean pragmatics. In Yuan Hang (ed.), *The Oxford Handbook of Pragmatics* (pp. 47–78). Oxford: Oxford University Press.

Hurwitz, Jon, and Peffley, Mark (2005). Playing the race card in the post-Willie Horton era: The impact of racialized code words on support for punitive crime policy. *Public Opinion Quarterly*, 69(1):99–112.

Keiser, Jessica (2020). Language without information exchange. *Mind and Language*, 37(1):22–37.

Khoo, Justin (2017). Code words in political discourse. *Philosophical Topics*, 45(2):33–64.

Lee, James J. and Pinker, Steven (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review*, 117(3):785–807.

Lewis, David K. (1969). *Convention: A Philosophical Study*. Malden, MA: Wiley-Blackwell.

Loar, B. (1981). *Mind and Meaning*. Cambridge, UK: Cambridge University Press.

Mendelberg, Tali (2001). *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton: Princeton University Press.

Moore, Richard (2018). Gricean communication, language development, and animal minds. *Philosophy Compass*, 13(12):e12550.

Neale, S. (1992). Paul Grice and the philosophy of language. *Linguistics and Philosophy* 15(5),509–559.

Reiland, Indrek (n.d.). Meaningfulness, conventions, and rules. (unpublished ms)

Sandberg, Sheryl, and Scovell, Neil (2013). *Lean In: Women, Work, and the Will to Lead*. New York: Alfred Knopf, Inc.

Saul, Jennifer Mather (2012). *Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford: Oxford University Press.

Saul, Jennifer (2018). Dogwhistles, political manipulation, and philosophy of language. In Daniel Fogal, Daniel W. Harris and Matt Moss (eds.), *New Work on Speech Acts* (pp. 363–383). Oxford: Oxford University Press.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Schiffer, Stephen (1972). *Meaning*. Oxford: Clarendon Press.

Solan, Lawrence, and Peter Tiersma (2005). *Speaking of Crime: The Language of Criminal Justice*. Chicago: University of Chicago Press.

Stanley, Jason (2015). *How Propaganda Works*. Princeton: Princeton University Press.

Stenius, Erik (1967). Mood and language-game. *Synthese*, 17(1):254–274.

Stokke, Andreas (2018). *Lying and Insincerity*. Oxford: Oxford University Press.

Tirrell, Lynne (2012). Genocidal language games. In Ishani Maitra and Mary Kate McGowan (eds.), *Speech and Harm: Controversies Over Free Speech* (pp. 174–221). Oxford: Oxford University Press.

Weisberg, Michael (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

White, Ismail K. (2007). When race matters and when it doesn't: Racial group differences in response to racial cues. *American Political Science Review*, 101(2):339–354.

Witten, Kimberley (n.d.). Dogwhistle politics: The new pitch of an old narrative.

4

STALNAKER

4.1 Introduction

Among Robert Stalnaker's many influential contributions to the philosophy of language is his development of a formal pragmatic framework for modeling the structure of linguistic discourse. A model of discourse structure is not itself a theory of linguistic communication; rather, it *represents* a theory of linguistic communication. However, it is also generative; an exploration of the model may produce insights that lead to developments in the theory of linguist communication that it was initially designed to represent. Stalnaker, for instance, developed his model of discourse structure based on previously held commitments (strongly influenced by Grice) about the nature of speech production and interpretation. This model, however, generated developments in semantic and theory, as people realized that its tools could be used to define other kinds of speech acts and semantic rules. Prior to his work in this area, semantic theory tended to be respected as a serious, rigorous enterprise, while pragmatic theory was dismissed as informal and unsystematic. Stalnaker broke away from these restrictive and somewhat artificial associations by demonstrating that the domain of pragmatics is subject to the same level of careful, systematic theorizing as that of semantics, thereby helping to generate an explosion of new work in dynamic semantics and speech act theory.[1]

Stalnaker shares with Grice and Lewis a conception of language as a mechanism for cooperative information exchange, and his pragmatic framework is tailored to represent data that is primarily restricted to contexts which conform to this picture. The motivation for this restriction places the Stalnakerian model under two categories of idealization described by

Weisberg (2007, 2013): Gallilean idealization and minimalist idealization. Recall that Weisberg describes minimalist idealizations as aiming to capture the essential character of the phenomenon in question. While here the relevant phenomenon is linguistic discourse as a whole, Stalnaker takes information exchange to be the "essential" and "central" case:

> The point of a discourse—at least one central kind of discourse—is the exchange of information.[2]
>
> The reasons people talk to each other are of course varied and complex, but it seems reasonable to assume that there are some kinds of purposes that are essential to the practice and that are the principal reasons for speech in the most straightforward kinds of conversation. In a simple exchange of information, people say things to get other people to come to know things that they didn't know before.[3]

Thus, one motivation for tailoring the formal pragmatic framework to represent cooperative information exchange is that this type of exchange is taken to capture the essential character of discourse as a whole.

The Stalnakerian framework also exhibits what Weisberg calls *Galilean* idealization. This type of idealization introduces simplifying distortions into the model in order to render it more theoretically tractable. After theorists come to understand the simplified system, they can then systematically remove the distortions, bringing the model closer to accuracy. The Stalnakerian model exhibits Galilean idealization in the following way: Stalnaker assumes that not only is cooperative information exchange the "essential" type of discourse, but that other types of discourse "can be explained in terms of the same categories as a serious exchange of information".[4] Another justification for restricting the data set to cooperative information exchange, then, is the assumption that alternative types of speech can be modeled using the same basic theoretical tools. This will be a matter of introducing complexity to the existing framework rather than substantively restructuring it.

For Stalnaker, then, the conception of language as a mechanism for cooperative information exchange furnishes the starting point for theorizing about the nature and structure of discourse. The rough picture goes something like this: in the context of cooperative information exchange, linguistic communication is a matter of cooperating on achieving a shared goal that is motivated by a shared interest in learning about what our world is like. Because none of us have the capacity to learn a significant amount of information about the world on our own, our mutual interests will be best served by adopting the common goal of exchanging information. It is assumed that if a pragmatic model can adequately represent this "central" case, it has captured the essential character of discourse as a whole. Then

the same basic theoretical tools—perhaps with some minor adjustments and added complexity—can be used to represent outlier cases as well.

The process of exchanging information is represented in Stalnaker's framework using a single information set—the *common ground*.[5] The common ground has two notable properties. First the information in the common ground is *shared* in the sense that all members of conversation have some particular attitude toward that information. Second, the common ground is *public* in the sense that these attitudes are iterated (they all have the attitude toward the fact that they all have the attitude… toward that information).[6] In the central case, the common ground is a set of propositions compatible with the information that conversationalists commonly believe (and commonly believe that they commonly believe, etc.), and which serves as the background against which speech acts are both produced and interpreted.[7] When we produce speech acts in service of exchanging information, we are proposing to add to this set of information in order to learn more about what the world is like. The common ground also plays a role in interpretation; in order to understand how a speaker proposes to shape the common ground, her speech act must be evaluated against the existing common ground. Without common interests and common belief, Stalnaker suggests, communication would be impossible:

> 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.[8]

The idea here, following Grice, is that communicative intentions are essentially overt. In order for these intentions to be recognized, both speaker and audience must rely on information that is public and shared between them. And in order for the audience's recognition of the speaker's intentions to result in belief, which is the goal of information exchange, she must share common interests with the speaker. (Otherwise, e.g., she may not consider her to be trustworthy.) Thus, the possibility of communication requires both common beliefs and common interests.

In the central case of cooperative information exchange, the framework's foundational theoretical concept is *belief*: It is used to define other theoretical categories such as the common ground, and to explain the production, interpretation, and success conditions of speech acts. Stalnaker realizes, however, that the notion of belief is too strong to characterize noncentral cases. In keeping with the methodology of Galilean idealization, he

suggests that this can be done using the related, but weaker notion of *acceptance*. Acceptance is the attitude of treating something as true for some reason—where, in the central case, one treats something as true because one believes it. Thus, the framework can be made to accommodate non-central cases with this minor revision to the original model: Acceptance replaces belief as the basic theoretical building block of the framework. Common ground is now understood as the information that is commonly *accepted* for the purpose of the conversation, rather than commonly *believed*.

Replacing the notion of belief with that of acceptance removes a simplifying distortion. Though this move is intended to expand the scope of the framework beyond the central case of cooperative information exchange, the resulting picture still adheres to an idealized picture of linguistic communication as a cooperative activity motivated by shared interests and aimed at achieving a shared goal. This picture is baked into the model in the following interrelated ways, which I will sometimes refer to as the *five Stalnakerian idealizations*:[9]

1. Conversations are taken to have a *unique* goal.
2. This goal is taken to be *shared*.
3. All speech acts are taken to be aimed at updating, and as being interpreted against a *unique* information set.
4. This information set is characterized in terms of *shared* mental states.
5. This information set is characterized using a *unique* and *coarse-grained* type of mental state.

I argue below that these idealizations prevent the Stalnakerian framework from providing an explanatory representation of various linguistic phenomena, including non-literal speech, non-serious speech, deception, dogwhistles, and miscommunication. For the sake of simplicity, I use an early—and somewhat simplified—version of the Stalnakerian framework throughout most of the chapter. In the final section, however, I consider various proposals for modifying the traditional Stalnakerian framework, arguing that they fail to address the foregoing worries.

4.2 Non-literal speech

As mentioned above, when the Stalnakerian framework is restricted to the central case of cooperative information exchange, it uses belief as its foundational theoretical category. In an attempt to remove simplifying distortions and account for non-central cases, belief is replaced with the notion of acceptance for the purpose of the conversation. This move weakens the account in two ways. First, while it still presupposes uniformity of purpose *within* conversations, it allows for variation of purpose *across* them.

Second, it allows that these purposes do not always involve the production of belief, though it assumes that they uniformly involve the production of the closely related attitude of acceptance.

Speakers, however, may have a variety of conversational goals, not only across—but *within*—conversations. A consideration of non-literal speech illustrates that conversational goals may even vary within a single utterance.[10] Consider the following example from Donnellan, which Stalnaker uses to illustrate how acceptance can come apart from belief:

MARTINI: At a cocktail party, Alice says to Bob, "The man drinking a martini is a philosopher", knowing that the man is drinking Perrier. However, since she also knows that Bob believes that it is a martini, she believes that the best way to identify her intended referent is to use this description.[11]

Stalnaker notes that Alice and Bob can accept for the purpose of the conversation that the philosopher is drinking the martini, *even if both realize it to be false*; their pretense of belief is rational because it is "an efficient way for them to communicate something true—information about the man who is falsely presupposed to be the man drinking a martini".[12]

This example brings out the fact that a speaker may have a variety of goals in relation to a single utterance. Here, these goals are associated with different contents: the false content (that the man drinking the martini is a philosopher) which is literally expressed by the sentence, and the true content (that *that* man is a philosopher) which is ultimately communicated. The speaker utters a sentence expressing the false content with the goal of communicating the true content. The true content plays a different role in the speaker's goals; she communicates this content for the purpose of imparting information.

The same point can be illustrated by looking at a famous example from Grice, involving metaphor:[13]

COFFEE: A speaker utters the sentence "you are the cream in my coffee" to their lover, with the intention of communicating that she is fond of her lover.

This example appears to exemplify the kind of pretense discussed by Stalnaker; though the speaker doesn't actually believe that the audience is the cream in her coffee, it is rational for her to talk as though she did because this allows her to communicate true information—that she is fond of her lover. Here again, the speaker has a variety of goals in relation to a single utterance, relative to the following contents: the literal content of the sentence in its context (that the audience is the cream in the speaker's coffee), and the non-literal content (that the speaker is fond of the audience) which

is indirectly communicated. Again, the speaker utters a sentence expressing the literal content with the goal of communicating the non-literal content, while she communicates the non-literal content with the goal of imparting information.

Both of these examples pose problems for the first Stalnakerian idealization, given that each illustrates how a conversation—indeed, a single utterance—may involve a variety of goals. They also illustrate that different goals may be aimed at updating different sets of information, putting pressure on the third Stalnakerian idealization. Finally, these different sets of information seem to be associated with different attitudes, contra the fifth Stalnakerian idealization. In both cases, the speaker has different goals relative to the literal and non-literal contents of her utterance. With each of these contents, she aims to update a different set of information. Otherwise, the updated information set in the Martini case would include both the information that the man drinking the martini is a philosopher as well as the information that the man drinking the Perrier is a philosopher. In the *Coffee* case, it would include both the information that the speaker's lover is the cream in her coffee and that the speaker is fond of her lover.[14] This seems like the wrong result; an information set that combines the literal and non-literal contents does not accurately represent the conversational dynamics. Not only do the speaker's distinct goals correspond to updating distinct information sets, these information sets seem to be associated with different attitudes that cannot be adequately characterized using the single umbrella notion of acceptance. Given that both literal and non-literal contents may be accepted for some conversational purpose or other, the notion of acceptance is far too broad to shed explanatory light on the *nature* and *role* of non-literal speech. As many have pointed out, non-literal speech plays a specific role in linguistic exchange, which is to communicate non-literal content.[15] A model of conversation should be able to represent the fact that the non-literal content is the main point of the utterance and that the literal content is the *mechanism* though which the speaker communicates this point. Representing both literal and non-literal speech as aiming at updating the very same attitude of acceptance erases important distinctions in their conversational functions.

4.3 Non-serious speech

Let's think of non-serious speech as speech that is not ultimately aimed at information exchange. Examples include, e.g., jokes, storytelling, acting on a stage, etc. While non-literal speech illustrates that conversational goals may vary within a single utterance, non-serious speech illustrates that conversational goals may vary across a conversation—even when we are only considering the goals of a single speaker. A single conversation may

include speech that is aimed at exchanging information as well as non-serious speech; it is common, for instance, to intersperse joking with assertion. Consider the following utterances made by Michelle Wolf during the course of her 2018 White House Correspondent's Dinner speech:

WOLF JOKE: During her 2018 White House Correspondent's Dinner speech, Michelle Wolf joked, "Trump is so broke he looked for foreign oil in Don Jr.'s hair."

WOLF ASSERTION: At the end of her 2018 White House Correspondent's Dinner speech, Michelle Wolf asserted, "Flint still doesn't have clean water."

As characterized above, serious and non-serious speech serve different goals. Serious speech is aimed at information exchange, while non-serious speech is aimed at something else, such as entertaining one's audience. Thus, the example of Wolf's speech poses a problem for the first Stalnakerian idealization, according to which each conversation is guided by a unique goal. This fact generates further difficulties for the third and fifth idealizations, according to which this goal is aimed at updating a unique information set which is characterized by the single, coarse-grained notion of acceptance. If Wolf's various conversational goals were aimed at updating a unique information set characterized by acceptance, then—in the case of success—that set would include both the information that Trump is so broke he looked for foreign oil in Don Jr.'s hair, and that Flint still doesn't have clean water. Again, this seems like the wrong result; a model that represents the contents of jokes and assertions as belonging to a single information set does not accurately represent their distinct conversational functions. Finally, like non-literal speech, non-serious speech brings out difficulties in characterizing all speech acts as aiming to update a unique, coarse-grained attitude. Given that both jokes and assertions are accepted for some conversational purpose or other, the generic notion of acceptance fails to shed explanatory light on the *nature* and *role* of these different kinds of speech acts. Assertions are standardly put forward as candidates for belief, while jokes are put forward as objects of amusement. Because acceptance encompasses both of these finer-grained mental states, it masks the line between them.[16]

4.4 Deception

In contexts of linguistic deception, interlocutors have conversational goals that are at odds with one another. In such contexts, one party typically intends to exchange information while the other intends to provide misinformation. This feature of deceptive speech puts pressure on the fourth Stalnakerian idealization, according to which speech always aims to update

shared attitudes.[17] Recall that on the Stalnakerian model, speech acts are understood in terms of their intended effects on the information that is mutually accepted by conversational participants. Updating a shared information set, however, seems like the wrong way to represent the aims of certain kinds of speech acts. Deceptive speech, for instance, is designed to create a *distinction* between what is believed by the speaker and her audience.

It is interesting to note that it is partially in response to this sort of case that Stalnaker weakens the initial Galilean idealization, moving from a characterization of the common ground in terms of shared *belief* to that of shared *acceptance*. Deception poses a counterexample to the original account because the liar does not believe her assertion; thus, in the case of success, its content is not added to what is commonly believed by the interlocutors. However, in pretending to believe her lie, the liar genuinely accepts its content for the sake of conversation. Thus, deceptive speech can be accommodated by a general picture according to which speech acts are aimed at updating a shared information set, as long as this information set is characterized by the weaker attitude of *acceptance*, rather than *belief*.

The trouble with this appeal to acceptance is similar to that which arises in relation to non-literal and non-serious speech; it brings out problems with the fifth idealization, having to do with modeling conversational updates using a single, coarse-grained mental state. The notion of acceptance is attractive in that it is liberal enough to apply to a broad range of speech acts, however it does so at the cost of explanatory power. An update to the common ground is a byproduct—rather than the point—of deceptive speech.[18] The liar does not aim to create an alignment on the set of information that she and her interlocutor accept for the sake of conversation, but rather to create a *divergence* between their belief sets. If she is successful, their belief sets will diverge not only with respect to the content of the lie, but also with respect to what they take to be the beliefs shared between them. To see this, suppose that the audience understands both the force and the content of a deceptive assertion, but does not believe it. Suppose further that she does not explicitly challenge her interlocutor's assertion or in any way let on that she knows she has been lied to. In this case the lie has been added to the common ground; both liar and audience are treating it as true for their conversational purposes. On the Stalnakerian framework, the conversational impact of the lie stops here, at the update of the common ground; it fails to represent the fact that the liar's ultimate communicative aim—that of updating the audience's beliefs in a way that does not align with her own—has not been met. Thus, the aim of deceptive speech falls outside the scope of what a model can adequately represent using only sets of *shared* information. Moreover, the notion of acceptance for the sake of conversation is too weak a notion to capture the goal of deceptive speech. While acceptance may be a byproduct of deceptive speech, it aims at belief.

4.5 Dogwhistles

Another type of speech that puts pressure on various aspects of the Stalnakerian framework is dogwhistling. As in Chapter 3, I follow linguist Kimberly Witten (n.d.) in thinking of dogwhistles as speech acts that are intentionally designed to allow for at least two plausible interpretations. One of these interpretations is private in the sense that it is targeted at a subset of the general audience and intended not to be recognized by audience members as a whole. Recall that Saul (2018) makes further distinctions between different types of dogwhistles (e.g., covert vs. overt and intentional vs. unintentional). In Chapter 3 we looked at covert intentional dogwhistles. In this section, we'll limit our focus to intentional dogwhistles, while ignoring the covert/overt distinction. That is, we'll focus on cases in which this double-messaging is intentional, but ignore the question of whether it is overt in the sense that the group that is targeted with the private interpretation is intended to recognize the speaker's communicative intentions.

An example from Saul (2018) (quoted from Noah (2004)) of this kind of dogwhistle would be the following statement from George W. Bush from his 2003 State of the Union address:

BUSH DOGWHISTLE: "Yet there's power, wonder-working power, in the goodness and idealism and faith of the American people."

As Saul notes, the message intended for the non-fundamentalist Christian audience is a fairly straightforward claim (if couched in what Saul calls "fluffy political boilerplate") about the power of the goodness, idealism, and faith of the American people. But there is also at least one encoded message aimed specifically at the fundamentalist subset of the audience, which will go unnoticed by their non-fundamentalist counterparts. Among fundamentalists, the phrase "wonder-working power" refers specifically to the power of Christ.[19] Fundamentalist Christians, familiar with this use of the phrase, will likely interpret Bush as claiming that Christ's power is in the goodness and idealism and faith of the American people (as well as gain awareness of his status as an insider in the community).

Dogwhistles illustrate that in producing a single utterance, a speaker may have different goals relative to different audience members, contra the first Stalnakerian idealization. Like deception, this case also puts pressure on the fourth idealization, according to which speech acts are aimed at updating shared mental states. While the liar aims to create a misalignment between her own and her audience's mental states, in the case of dogwhistles the speaker aims to create a misalignment between the mental states of her audience members. This fact, in turn, puts pressure on the third idealization, according to which speech aims to update a single information set.

In order to model the speaker's goals, a pragmatic framework would need to represent the disjointed information states corresponding to at least two different subsets of the audience.

4.6 Miscommunication

On the Stalnakerian picture, the common ground plays two principal theoretical roles: it is "involved in an account of what language is used to do, and in an account of the mechanisms it uses to do it."[20] Thus far, we've been talking about the first role, having to do with the aim of speech production; speakers use language in order to update the common ground. In this section we'll look at issues relating to the second role, which has to do with audience interpretation; with respect to this role, the common ground functions as a means for resolving context-sensitive expressions. Before a speech act can *change* the common ground, it must first be interpreted against the *existing* common ground in order to resolve ambiguities and clarify its aim. In accordance with its fourth idealization, the Stalnakerian framework represents the interpretive process as taking place against a background of *shared* information.

Because of differences in experience and social positioning, however, we do not always share a substantial amount of relevant background information going into a linguistic exchange. This comes out especially clearly in the context of conversations about racial justice in America. As many have noted, the phrase "All lives matter" is commonly offered as a corrective to a *misinterpretation* of the "Black lives matter" slogan—one which wrongly assigns it the exclusive reading *only Black lives matter*. Anderson (2017a) points out that this misinterpretation arises because of differences in accepted background information; the exclusive reading of "Black lives matter" is only available when ignoring facts about the mistreatment of Black Americans within the criminal justice system and the broader historical context of the history of racial violence in the US.

Dotson (2011) calls the systematic failure to recognize such relevant background information "pernicious ignorance". This type of ignorance is systematic in that it "follows from a predictable epistemic gap in cognitive resources" and is pernicious in that it can generate harm to another person or set of persons.[21] The ignorance of "All lives matter" counter-protestors is predictable because they occupy social positions which generate incentives to ignore facts about social injustice. Such information may morally implicate them as well as oblige them to work to change a social system from which they benefit. As Anderson (2017a) points out (echoing Fricker 2007), "there are some situations … in which it serves the interests of the socially powerful to maintain ignorance or a misinterpretation of certain social experiences".[22]

On the Stalnakerian picture, such conversations are paradigmatic cases of communication breakdown. To the extent that interlocutors fail to have enough shared background information to be able to interpret one another in the intended ways, the context is "defective" and will result in miscommunication. Such communication breakdowns, however, fall outside the representational scope of the Stalnakerian framework. Because it models interpretation as occurring against the backdrop of a *shared* information set, it lacks the resources for representing interpretation processes which occur in defective contexts.[23]

This simplifying distortion is justified by Stalnaker's conception of linguistic communication as fundamentally cooperative. Because he assumes that conversations are cooperative activities contributing to shared goals (the second Stalnakerian idealization), he believes that we will have reason to recognize and repair defective contexts. Thus, such cases will be rare and likely to revert back to "normal":

> A defective context will have a kind of instability, and will tend to adjust to the equilibrium position of a nondefective context. Because hearers will interpret the purposes and content of what is said in terms of their own presuppositions, any unnoticed discrepancies between the presuppositions of the speaker and addressee is likely to lead to a failure of communication … Because communication is the point of the enterprise, everyone will have a motive to try to keep the presuppositions the same … it is not unreasonable … to assume that in the normal case contexts are nondefective (or close enough).[24]

These idealizations, however, leave the Stalnakerian framework without the theoretical tools to represent the interpretive processes that occur in defective contexts. In such contexts, there is no objective, shared information set that can do the work of determining the contribution of the speech act—instead, there is an information set that the speaker intended her speech act to be interpreted against, and a different information set against which the audience did in fact interpret it. One might wonder, at this point, why we need to be able to model speech that occurs in such contexts, if it results in miscommunication. It's one thing to suggest that a model of discourse should be able to adequately represent non-ideal cases of communication—but is it really a problem that *miscommunication* falls outside the scope of the representational capacity of a model of linguistic communication? Perhaps that's as it should be.

I suggest, however, that we *do* need to be able to represent miscommunication, for several reasons. First, it is commonplace. Some degree of misalignment in background information relevant to the conversation appears to be the norm rather than an aberration. And while such "defective

contexts" may resolve themselves in situations where interlocutors are coordinating on shared conversational goals (i.e., contexts that conform to the second idealization), in many cases interlocutors will lack the incentive to align their accepted information with that of their conversational partner. In such cases there is no reason to expect that a defective context will naturally resolve itself.[25]

Second, defective contexts are not homogeneous; different kinds of miscommunication have different social implications. Though both lead to a breakdown in communication, conversations in which we fail to share knowledge of a public language are importantly different from those in which we fail to share information due to our different positions in a social structure.[26] As Anderson (2017a) and Dotson (2011) point out (following a tradition of Black feminist thought which includes theorists such as Hill-Collins (2000) and Lorde (1984)), features of the latter context may both constitute and perpetuate systematic social injustice. If we have an interest changing conditions of injustice, we must seek a deeper understanding of these conditions. By treating miscommunication as a unified phenomenon and cordoning off defective contexts from the representational scope of a pragmatic theory, we rob ourselves of important explanatory resources. Looking for patterns in defective contexts, for instance, can help us to understand how conditions of social injustice can impact discourse, and vice versa. Anderson (2017a) makes a similar point when he says of the exclusive reading of "Black lives matter" that:

> One could say that this is simply a misreading, perhaps an uncharitable one. I think the proper response goes deeper than this, however … there is a gap in the collective interpretive resources that unjustly disadvantages the protestors' ability to express themselves intelligibly.[27]

These considerations suggest that an adequate pragmatic theory should be able to provide an explanatory model of the preconditions, processes, and effects of *miscommunication* as well as communication. This involves dropping the second Stalnakerian idealization, according to which interpretation takes place against the backdrop of a shared information set. Moreover, an adequate explanation of the persistence of certain kinds of defective contexts requires dropping the fourth idealization, according to which the goals of speech are shared. Defective contexts commonly fail to self-correct because they are characterized by misaligned goals which don't incentivize repair.[28]

4.7 Modifications

In this section we'll consider ways that proponents of the Stalnakerian framework have responded—or might respond—to the concerns raised

above, either by using resources within the existing framework, or by introducing minimal modifications in accordance with the method of Galilean idealization.

4.7.1 Presupposition sets

One might suggest that the Stalnakerian framework has resources that I've been ignoring, which could perhaps go some way in addressing the concerns raised above. For instance, in addition to the common ground, the Stalnakerian framework includes each conversational participant's presuppositions—i.e., the set of contents that each participant believes to be part of the common ground. (In nondefective contexts, participants' presupposition sets will be identical to the common ground.) However, while presupposition sets do contribute representational and explanatory power to the framework, they don't provide the resources needed to get around the worries expressed above.

First, presupposition sets don't add the resources needed to distinguish between different kinds of speech acts. The difference between the communicative aim of a joke vs. that of an assertion, for instance, has to do with the attitude that the speaker intends the audience to take toward its content—not with whether she intends the content to be mutually accepted by all participants vs. *merely believed by the audience* to be mutually accepted by all participants (similarly with literal vs. non-literal speech). Thus, presupposition sets do not provide the tools for representing the rich range of communicative aims that are present in ordinary speech.

Second, though presupposition sets need not be shared, they nonetheless remain too closely related to shared information sets to be able to ameliorate the worries raised against the framework's capacity to represent interpretation. For instance, not only can speech be interpreted against information that is not shared, it may also be interpreted against information that is not even *presumed* to be shared. This is true even in cases of communicative success. For instance, if I recognize that you have used a malapropism, then I may appeal to what I believe to be your false beliefs about language in order to interpret you. In this case I am drawing on my beliefs (not shared by you) about your beliefs (not shared by me) about language. Or perhaps I believe that you only used this malapropism because you falsely believe that *I* assign it the nonstandard meaning, and you are merely trying to accommodate me.[29] Here, in order to interpret you I will use my beliefs (not shared by you) about your beliefs (not shared by me) about my beliefs (not shared by you) about language. We rely on a broad range of information in interpreting one another's speech acts; while in some cases this information is shared, in other cases it is neither shared nor presumed to be shared.[30]

It is also unlikely that an appeal to individual presupposition sets will always be able to provide an adequate explanation of systematic patterns of miscommunication, such as the exclusive reading of "Black lives matter" discussed in Section 4.6. As I argued there, this process of interpretation cannot be adequately modeled as occurring against the backdrop of an information set that is shared between interlocutors. One might claim, however, that this misinterpretation can be explained by invoking the presuppositions of the ALM ("All lives matter") audience; they assign the exclusive reading because they assume that the information shared by themselves and BLM ("Black lives matter") protestors excludes America's history of racial violence. But this seems like the wrong explanation. Rather than failing to adequately understand the mental states of BLM protestors—and thus the information that is shared between them—the ALM activists appear to be discounting the perspectives and the lived experiences of BLM protestors altogether. Thus, the presupposition sets of ALM protestors do not appear to be a promising resource for explaining their systematic misinterpretation of the "Black lives matter" slogan. Rather, it is the fact that they are incentivized to ignore facts about the racist social structures from which they benefit, as well as the perspectives of those who seek to dismantle those structures. In order to model this, a pragmatic framework must be able to model mental states that are neither shared, nor presumed to be shared.

4.7.2 Derived contexts

Stalnaker suggests that differences between at least some speech acts can be represented by introducing additional information sets that are derived from the common ground (he calls these "derived" or "subordinate" contexts). There are two types of subordinate contexts: s*imple subordinate contexts* are subsets of the common ground, while *parallel subordinate contexts* are not. On this more complicated picture, the common ground is the *basic context*; subordinate contexts are derived from the basic context in the sense of being definable in terms of it. For instance, suppositions form a simple subordinate context containing information that is only temporarily added to the basic context. Assertions, then, could be distinguished from suppositions by the fact that the former updates the basic context while the latter updates a subordinate context. Subordinate contexts may also be used to model subjunctive conditionals: A claim that "if p then q" adds q not to the basic context, but to a subordinate context which includes the set of possibilities that are minimally different from the basic context, and in which p is true. In this case this subordinate context will be parallel rather than simple—that is, it will not be a subset of the common ground. Here again, subordinate contexts appear to provide the resources to distinguish between different kinds of speech acts—in this case, assertions

update the common ground, while subjunctive conditionals update subor-
dinate contexts. However, while this strategy may offer the resources for a
descriptively adequate representation of suppositions and certain kinds of
conditional statements, it's not clear that derived contexts offer a universal
fix. Suppositions and conditionals contribute to information sets that are
related to the common ground in a fairly straightforward and specifiable
way, but that may not be true of all speech acts. For instance, it's not clear
that there are specifiable rules for how the information set containing the
literal content of metaphors or jokes relates to the basic context set, which
would generate a principled way of demarcating these speech acts within
the framework. Moreover, we need the theoretical tools to represent the
various ways in which different information sets interact: suppositions tem-
porarily interact with information that is believed, while the information
from a story is generally kept entirely separate from it.[31] Literal content
interacts with non-literal content in an even more complex way, in that it
plays some role in *generating* the latter. Thus, we need the machinery to
not only show how these different sets of information are related in terms
of their conversational contributions, but also in terms of their pragmatic
roles of facilitating interpretation. While perhaps all this can be done, it's
not at all clear that what we end up with will be definable—in any sort of
substantive and explanatory sense—from the common ground.

4.7.3 Non-derived contexts

Others have suggested revisions to the traditional Stalnakerian framework
which involve additional information sets that are not clearly derived from
the common ground. For instance, Camp (2018) suggests that conversa-
tional moves may contribute to three distinct sets of information; 1) the con-
versational scoreboard, which closely corresponds to the content of what
is said in a conversation; 2) the common ground, which contains informa-
tion that participants are prepared to openly acknowledge; and 3) mutual
beliefs, which participants are not necessarily prepared to openly acknowl-
edge.[32] Which information set a speech act contributes to is determined by
the degree to which the speaker is liable to defend that contribution. Stokke
(2018) also suggests a solution which posits multiple information sets. He
defends a revision of the framework according to which non-serious speech
is added to an "unofficial" common ground, while assertions are added to
an "official" common ground.

These proposals are on the right track in the sense of recognizing that
an adequate representation of the function of various kinds of speech acts
will need to incorporate multiple information sets. However, their revisions
to the Stalnakerian framework are inadequate for several reasons. First,
both of these strategies characterize conversational dynamics using a small,

fixed number of information sets, while an adequate model will need to make use of a greater—and perhaps open ended—number of information sets. (Camp (2018) suppresses attitudes other than belief only for the sake of simplicity, so may welcome additional information sets corresponding to other kinds of attitudes.) This is because there will be information within *each* of the proposed categories of information sets that should be kept distinct. Different sets of content qualifying as "unofficial" in Stokke's view may nonetheless be playing distinct conversational roles. For instance, one can speak non-literally while joking. If both of these contents are added to a single "unofficial" common ground, the model will not be able to represent their distinct communicative roles.[33] Camp's proposal faces a similar issue, in that there are differences in the communicative roles of different types of speech that run orthogonal to the commitment level taken on by a speaker. Consider the use of metaphor within a joke: the difference between the literal and non-literal content is not a matter of speaker liability (presumably the speaker is not committed to either content), but rather communicative role. As we discussed earlier, the literal content serves as a mechanism for communicating the non-literal content.

4.7.4 Acceptance relativized

Lastly, I consider a way in which a proponent of the Stalnakerian framework could introduce a minor modification in order to handle worries I raised for its reliance on the single mental state of acceptance. I argued above that acceptance for the purpose of conversation is too coarse-grained a notion to delineate various kinds of conversational goals and the information sets that they aim to update. Assertions, for instance, are characteristically not aimed merely at acceptance, but at belief. One might respond to these concerns by suggesting the following minimal revision to the traditional Stalnakerian framework: instead of appealing to a generic notion of acceptance for the sake of conversation, the theory should appeal to acceptance for specific purposes. Assertion, then, might be characterized as aiming at acceptance for the purpose of belief, while joking may be characterized as aiming at acceptance for the sake of entertainment. Yalcin's (2007) idea of *conversational tone* is an example of such a strategy. Yalcin holds onto the traditional Stalnakerian idea that speech acts are aimed at updating the common ground, which is defined as common acceptance for the purpose of the conversation. He then suggests that each conversational purpose determines a conversational tone which specifies the appropriate attitude to hold toward the information in the common ground. So, e.g., in the context of information exchange the conversational tone will be belief. Though Yalcin's discussion focuses on a case in which there is only one conversational tone, he suggests that a conversation may have multiple tones. Extrapolating from

this idea, we can suppose that where there are multiple conversational tones, there will be multiple common grounds corresponding to each. This strategy provides the resources to distinguish between various types of speech acts by appealing to the conversational tone appropriate to them.

While this solution might improve the framework's representational power, the resources it adds renders the unifying notion of acceptance explanatorily redundant. Recall that in considering the case of deception, we saw that acceptance was a byproduct, rather than the aim of assertion. We also saw that the notion of acceptance fails to be useful in explaining audience uptake; finer-grained mental states such as belief play these roles. Once we have relativized acceptance to finer-grained attitudes such as belief or entertainment—e.g., through the notion of conversational tone—the notion of acceptance itself is no longer doing any significant explanatory work. Rather, it functions only as a tool for creating the appearance of theoretical unity, a virtue that is a hallmark of Stalnakerian framework. But there are reasons to think that a more substantive theoretical unity cannot be achieved, if the resulting framework is to be explanatorily adequate. We can see this by tracing the history of modifications to the original Galilean idealizations that led us to this point. Under the original idealization, the central theoretical concept was belief, following the picture of language as cooperative information exchange. The initial modification of changing the central theoretical concept from that of belief to that of acceptance allowed the framework to represent a greater range of speech phenomena, but at the cost of representational and explanatory power; acceptance captures a broader range of speech acts but fails to distinguish between them. As we've seen, the finer-grained notion of belief seems to be playing a central role in the aims of at least some speech acts, and the move from belief to acceptance papers over this fact. In order to remedy this situation, the final modification we've considered involves the strategy of relativizing acceptance to purposes involving finer-grained mental states. In doing so, it re-introduces the concept of belief—which is important for representing at least some speech phenomena—while adding other fine-grained propositional attitudes in order to represent speech acts for which the concept of belief is inappropriate. Once we reach this stage, however, it becomes apparent that we can do away with the second modification; after we have a range of fine-grained attitudes—including, but not limited to belief—there is no longer a substantive theoretical role for acceptance to play.[34]

4.8 Conclusion

We are now in a position to evaluate whether the foregoing concerns pose a major threat to the Stalnakerian framework, given that its guiding conception of language as cooperative information exchange is intended as an

idealization. Let's start by considering the success of this conception as a minimalist idealization. The difficulties outlined in this chapter would not be of great concern if they qualified as sufficiently peripheral cases. But it is a stretch to think that cooperative information exchange captures the essential nature of linguistic discourse as a whole. At the very least, this claim must be argued for rather than assumed. The linguistic phenomena considered above are ubiquitous and central to ordinary linguistic experience; they are of importance not only in relation to the purely theoretical project of understanding the mechanisms of communication, but in relation to the practical project of intervening in our linguistic practices in order to promote interests of social justice. Thus, the conception of language as cooperative information exchange fails as a minimalist idealization.

What about Galilean idealization? The cases above would be unproblematic from the standpoint of Galilean idealization if they could be accommodated by introducing minimal modifications to the framework. However, we have explored various strategies for doing so, and their failures suggest that the framework is in need of more drastic restructuring. While each of these strategies increased the framework's capacity to represent different kinds of speech acts, we saw that all of them fell short of providing the tools to make all the distinctions needed to represent our speech practices (at least, without rendering the basic tools of the framework explanatorily inert). An additional problem with each of these solutions is that while they increase the overall *number* of information sets available to represent speech phenomena, they continue to restrict the framework to *shared* information sets. Our consideration of deception and interpretation brought out that a model that is restricted to shared content is unable to represent the aims of certain kinds of speech production, and lacks the tools to adequately explain important interpretive phenomena.

Let me be very clear about the scope of the critique that I've been trying to articulate here: My claim is not that the Stalnakerian framework *under any conception* cannot be adapted to accommodate non-ideal speech. Rather, my claim is that we must start from the ground up *with respect to those aspects of the framework that are grounded in a conception of language as cooperative information exchange*. The conception of language as cooperative information exchange, I claim, fails both as a minimalist and a Galilean idealization. However, many other core features of the Stalnakerian framework remain intact when this idealization is rejected. In Chapter 8, I propose an alternative framework which makes use of what I take to be the most central and valuable insights of the traditional Stalnakerian account, while rejecting the five idealizations associated with the picture of language as cooperative information exchange. This framework represents each conversation—even each utterance—as potentially involving a variety of goals, which need not be shared. These goals aim

to update, and are interpreted against a variety of information sets, which need not be shared. Finally, these information sets are individuated by a variety of fine-grained mental states. In that chapter I revisit some of the problems discussed here to show how they can be overcome by rejecting the Five Stalnakerian Idealizations.

Notes

1 See Stalnaker (1974).
2 Stalnaker (1998, p. 5).
3 Stalnaker (2002, p. 703).
4 Stalnaker (1974, p. 51).
5 Stalnaker's use of the term "common ground" is inspired by Grice's work: Stalnaker references Grice (1989, pp. 65 and 274) in Stalnaker (2002, p. 701).
6 In simpler models, the iterated attitude is the same as the first-order attitude, but one might also adopt a more complex model which represents them as distinct. For instance, Stalnaker (2002) presents a simplified model according to which both attitudes are belief, then proposes a revision according to which the first order attitude is acceptance and the iterated attitude is belief.
7 Stalnaker models propositions as possible worlds, but I remain neutral on this point.
8 Stalnaker (2014, p. 42).
9 Though I use a definite description, I don't mean to suggest that this framework involves only five idealizations.
10 I'll be using non-literal speech and indirect speech interchangeably. I provide a definition of indirect speech in Chapter 5.
11 See Stalnaker (2002, p. 717) and Donnellan (1966, p. 55).
12 Stalnaker (2002, p. 718).
13 See Grice (1989, p. 4).
14 Following Stokke (2018, p. 72), a Stalnakerian could claim that only the meta-phorical content—*not* the literal content—of metaphorical speech gets added to the common ground. I argue in Keiser (2020) that this fix is unmotivated, however there is a further worry: While this move effectively separates the different types of content, it does so at the cost of removing the literal content from the representational scope of the model. This seems like the wrong result, given that the literal content is playing an important communicative role.
15 See, e.g., Davidson (1978), Camp (2006).
16 Abbott (2008) makes a related point in raising the concern that the Stalnakerian framework cannot adequately distinguish between informative presupposition and assertion.
17 Harris (2020) raises problems for this feature of the Stalnakerian framework, though his argument focuses on speech contexts where there is delayed timing, rather than deception. Moreover, most of Harris' examples place more pressure on the *public* aspect of the common ground (iterated shared attitudes), while I'm concerned primarily with the *shared* aspect of the common ground. Though I'm in broad agreement with Harris on this issue, we draw different conclusions from the data, which I discuss in Chapter 8.
18 Though focused on speech acts in general rather than assertion in particular, Harris (2020, p. 2714) makes a similar point in claiming that an update to a shared information set is "at best a secondary and inessential aim of communicative acts". Unlike Harris, I'm not committed to this broad claim about communicative acts generally, but I do think it applies to assertion in particular.

19 This phrase comes from a hymn titled "There is Power in the Blood" by Lewis Jones, https://hymnary.org/text/would_you_be_free_from_the_burden_jones
20 Stalnaker (2002, p. 15).
21 Dotson (2011, p. 238).
22 Anderson (2017a, p. 140).
23 In the next section I consider, and reject, the idea that individual presupposition sets could do this work.
24 "Close enough" is understood as applying to cases where the divergences do not affect the issues that actually arise in the course of the conversation. See Stalnaker (1978, p. 35).
25 Peet (forthcoming) notes that defective contexts may also be difficult to resolve in cases in which the interlocutors are misaligned on presuppositions that are central to their understanding of the social world (and so difficult to adjust). He also points out that it may be difficult to articulate problematic presuppositions, and attempts to do so may cause confusion that results in further defects in the context. Finally, he notes that defective contexts may go unresolved in cases where the avoidance of conflict takes priority over communicative success.
26 Anderson (2017b) offers a more detailed taxonomy of various kinds of miscommunication, which he calls *hermeneutical impasses*.
27 Anderson (2017a, p. 140).
28 Anderson (2017b) argues that the power relations ingrained in existing social structures make certain kinds of miscommunication inevitable.
29 Compare with Stalnaker's martini case.
30 Cf. Harris (2020).
31 Though sometimes factual information gets imported into the fictional story. See Lewis (1978) for discussion.
32 Stalnaker (2002, p. 704).
33 For a related criticism of Stokke's account, see Michaelson (2018).
34 Peet (forthcoming) also points out that acceptance comes in degrees and cannot be thought of as an all or nothing affair—I take this to support my claim that the notion of acceptance is not fine-grained enough to do the theoretical work that we need it to do.

References

Abbott, Barbara (2008). Presuppositions and common ground. *Linguistics and Philosophy*, 31(5):523–538.

Anderson, L. (2017a). Epistemic injustice and the philosophy of race. In J. Medina, I. J. Kidd, and G. Pohlhaus Jr. (eds.), *The Routledge Handbook of Epistemic Injustice* (pp. 139–148). New York: Routledge.

Anderson, Luvell (2017b). Hermeneutical impasses. *Philosophical Topics*, 45(2):1–19.

Camp, Elisabeth (2006). Metaphor and that certain 'je ne sais quoi'. *Philosophical Studies*, 129(1):1–25.

Camp, Elisabeth (2018). Insinuation, common ground, and the conversational record. In D. Harris, D. Fogal, and M. Moss (eds.), *New Work on Speech Acts* (pp. 40–66). Oxford: Oxford University Press.

Collins, Patricia Hill (2000). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. 2nd ed. New York: Routledge.

Davidson, Donald (1978). What metaphors mean. *Critical Inquiry*, 5(1):31–47.

Donnellan, Keith S. (1966). Reference and definite descriptions. *Philosophical Review*, 75(3):281–304.

Dotson, Kristie (2011). Tracking epistemic violence, tracking practices of silencing. *Hypatia*, 26(2):236–257.

Fricker, Miranda (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Harris, Daniel W. (2020). We talk to people, not contexts. *Philosophical Studies*, 177(9):2713–2733.

Harris, Sandra. (1995). Pragmatics and power. *Journal of Pragmatics*, 23:117–135.

Keiser, Jessica (2020). The Limits of Acceptance. *Inquiry: An Interdisciplinary Journal of Philosophy*. https://doi.org/10.1080/0020174X.2020.1775380

Lewis, David (1978). Truth in fiction. *American Philosophical Quarterly*, 15(1):37–46.

Lorde, Audre (1984). *Sister Outsider*. Berkeley, CA: Crossing Press.

Michaelson, Eliot (2018). The lies we tell each other together. In Eliot Michaelson and Andreas Stokke (eds.), *Lying: Language, Knowledge, Ethics, Politics* (pp. 183–205). Oxford: Oxford University Press.

Noah, T. (2004). Why Bush opposes Dred Scott. http://www.slate.com/articles/news_and_politics/chatterbox/2004/10/why_bush_opposes_dred_scott.2.html

Peet, Andrew (2021). Defective contexts. In Rachel Katharine Sterken and Justin Khoo (eds.), *Routledge Handbook of Social and Political Philosophy of Language*. New York: Routledge.

Saul, Jennifer (2018). Dogwhistles, political manipulation, and philosophy of language. In Daniel Fogal, Matt Moss, and Daniel Harris (eds.), *New Work on Speech Acts* (pp. 360–383). Oxford: Oxford University Press.

Stalnaker, Robert (1974). Pragmatic presuppositions. In *Context and Content* (pp. 47–62). Oxford: Oxford University Press.

Stalnaker, Robert (1978). Assertion. *Syntax and Semantics (New York Academic Press)*, 9:315–332.

Stalnaker, Robert (1998). On the representation of context. *Journal of Logic, Language and Information*, 7(1):3–19.

Stalnaker, Robert C. (2002). Common ground. *Linguistics and Philosophy*, 25(5–6):701–721.

Stalnaker, Robert (2014). *Context*. Oxford: Oxford University Press.

Stokke, Andreas (2018). *Lying and Insincerity*. Oxford: Oxford University Press.

Weisberg, Michael (2007). Three kinds of idealization. *Journal of Philosophy*, 104(12):639–659.

Witten, Kimberly (n.d.). Dogwhistle politics: The new pitch of an old narrative. (unpublished ms)

Yalcin, Seth (2007). Epistemic modals. *Mind*, 116(464):983–1026.

PART II

Non-ideal foundations

5

NON-IDEAL FOUNDATIONS OF LANGUAGE[1]

5.1 Introduction

This chapter develops a novel foundational theory of linguistic communication that breaks away from the idealizations that I've been critiquing throughout this book. The picture I present below conceives of linguistic communication as a complex social affair characterized by a multiplicity of goals and interests. Though linguistic communication serves many different functions, I argue that one in particular plays a primary explanatory role: this is the goal of attention-direction. What we do, first and foremost, when we communicate, is to draw one another's attention to contents—whether these be objects, such as cloud formations, or states of the world, such as the fact that the driveway is icy. We often have further communicative goals in performing speech acts; we want to draw our audience's attention to something for specific reasons, i.e., to influence their beliefs, attitudes, or behavior. These additional goals can vary according to the context, but what is common to each type of communicative act is the goal of attention-direction.

Though the notion of attention has long been of interest to philosophers and cognitive scientists, it has recently become a topic of central importance in the public sphere. Many warn that we are collectively experiencing an "attention crisis" that is generated by the current business model of big tech and exacerbated by pandemic lifestyles.[2] It is thought that this crowding (or hijacking) of attentional space has profound effects on agency and general wellbeing, both compounding the risk of depression and addiction, and curtailing human potential. My project here to is further magnify the role of attention in our lives. What I want to suggest is that attention is not

only a crucial component of human agency and wellbeing, but also of our practices of public language use. Directing one another's attention is the central thing that we use language to do.

This picture of linguistic communication retains from Grice, Lewis, and Stalnaker the idea that language is a fundamentally social affair involving interdependent reasoning, but rejects the idea that it is always characterized by common interests. Because our social lives are complex, our communicative goals will at times serve shared interests and at other times be adversarial. This social complexity should be familiar from attention-direction in the realm of technology, which may at times—but not at others—be driven by shared interests. Though the scope, speed, and economic structure of online communication make it an especially interesting case, it is nonetheless an instance of a more general phenomenon, from which it derives its basic features: linguistic communication.

The proposed theory of linguistic communication is non-ideal in the theoretical sense of being devoid of some relevant theoretical idealization—in this case, the idealization which represents language as essentially a cooperative affair aimed at information exchange. As a consequence of being non-ideal in this *theoretical* sense, the theory provides the tools for thinking about language use that is non-ideal in the *normative* sense of failing to conform to standards of justice or morality. This positions the theory to provide a bridge between foundational issues in the philosophy of language and contemporary work in social political philosophy of language, as well as work on attention in the philosophy of mind, ethics, and epistemology.

The broader theory can be divided into roughly three parts: the theory of meaning, the metasemantic theory, and the pragmatic theory. These parts are interrelated, and each will draw from what I take to be the most important and compelling insights of Grice, Lewis, and Stalnaker. The upshot of the criticisms developed in the foregoing chapters is not that we ought to abandon their projects wholesale, but rather that we abandon very specific idealizations generated from their common perspective regarding the function of communication. My project essentially looks at what is left of their theories when these idealizations are done away with, and uses the remaining tools to rebuild anew. The theory of meaning outlines an account of a foundational act of communication, and will draw primarily from Grice's work on speaker meaning. The metasemantic theory incorporates this theory of meaning into a proposal about the conventions of use that constitute the public language of a population, and will draw primarily from Lewis's theory of linguistic convention. The pragmatic theory is an extension to the theory of meaning in the sense that it explores the goals and effects of a broader range of speech acts that extend beyond the foundational notion of meaning that is used in the metasemantic theory. Though I won't present

it here, in Chapter 8 I demonstrate how we might represent these speech acts in a framework that draws primarily on Stalnaker's work in formal pragmatics.

My proposal is also influenced by the work of Grice, Lewis, and Stalnaker in relation to aspects of its broad methodological commitments. Much of this book has been devoted to critiquing their common methodology of using a conception of language as cooperative information exchange as the launching point for a theory of linguistic communication. However, as I mentioned in the introduction, they share other important methodological commitments which I take on board. Like theirs, my project conforms to the principle of the *Autonomy of Pragmatics*, according to which general facts about the nature and structure of linguistic communication (which I call *pragmatic facts*) can be theorized independently of conventional devices governing specific linguistic practices. I also follow these theorists in adopting a *Pragmatics First* methodology in treating pragmatic facts as more foundational than semantic facts in the sense that the latter are analyzed in terms of the former; semantic facts are ultimately reducible to more general and familiar facts about the mental states, behavior, and relations between language users. In the remainder of this chapter, I present the broad outlines of this theory. Though certain parts may receive only a cursory treatment here, remaining chapters will explore its details more fully as I show how the proposal overcomes problems faced by the Gricean, Stalnakerian, and Lewisian accounts.

5.2 Meaning

There are many of ways of meaning things. So, in order to develop a theory of meaning, we should first ask what we want this theory to do. I suggest that at least one role for a theory of meaning is to carve out a conception of language use that can be put to work in a metasemantic theory. That is, it should characterize the particular kind of use that plays a role in determining which public language(s) are used by a given linguistic community. I use this *desideratum* as my launching point; in this section, I give a rough characterization of a very basic kind of speech act that I call *meaning*, arguing that it is a good candidate for playing this theoretical role. In the following section, I present a fuller pragmatic picture that extends to more complex uses of language. Finally, I return to the metasemantic picture, explaining in more detail how the basic speech act of meaning can be used in providing the grounding conditions of public language use.

Language can be used in seemingly endless different ways, presenting a challenge for identifying the particular kinds of use that are relevant to metasemantic theorizing. While it is commonplace to claim—echoing Wittgenstein—that meaning is use, Quine notes that:

> the trouble is that [Wittgenstein] stops just short of the problem. I also think that the meaning [of an expression] is the use … I think this is one sufficiently broad characterization to cover the whole lot; the problem is, it's too broad … so it's a question of what use counts and what use is irrelevant and should be dismissed … and there's where the problem begins, not where it ends.[3]

Lewis' solution to this problem was to identify metasemantically relevant use with the kind of use that is involved in cooperative information exchange. As we saw in Chapter 2, however, this solution is problematic; it both ignores alternative uses of language that are relevant to determining linguistic meaning and fails to sift out ways of exchanging information that are *not*. But if the popular conception of language as cooperative information exchange fails to deliver a viable launching point for delineating metasemantically relevant use, where else might we look? I propose that a promising starting place would be to examine our practices of language such as they are—in all their diversity and failures to conform to various normative standards—and try to find a common ingredient. This strategy of searching for a common *ingredient* to various kinds of language use can be contrasted with that of searching for a common *characterization*. As I noted in my critique of Stalnaker's use of the notion of acceptance, it's difficult to find a unified characterization of all uses of language which preserves the fineness of grain needed to mark important theoretical distinctions. Rather than trying to formulate a characterization that is sufficiently broad to "cover the whole lot", to use Quine's phrase, we should instead look for an ingredient that is common to all of them.

In order to find this common ingredient, it helps to first think about how speech acts are composed. An idea put forward and developed at length by Austin (1975) is that speech acts are complex and layered in the sense that they can form parts of one another. Though I do not adopt the particulars of Austin's speech act theory, I take on board this general insight about their structure. In the picture that I am proposing, speech acts are characterized by goals. In some cases, there will be a multiplicity of goals which can form a series, like links in a chain. These links are related in that the accomplishment of one goal serves as a means to achieving the next. Speech acts may also form parts of one another. In keeping with our analogy, we can think of the chain of goals that represents any particular speech act as being composed of segments of one or more links. Such segments may also represent speech acts in themselves, which form parts of more complex speech acts. I suggest that a common segment to the chain of goals which characterizes any communicative speech act is the intention to direct attention to a content. This intention characterizes a speech act that I call *meaning*, which is basic in the sense that it is the minimal requirement for

successful communication. I define this speech act in more detail below, but let me first sketch the rough idea. Consider the following utterance:

> "North Korean archeologists have discovered an ancient unicorn lair."

A speaker may have any number of goals in performing this utterance. She may, for instance, intend to garner a laugh, spark imagination, or induce a belief, among other things; but a prerequisite to achieving any of these goals is to direct the audience's attention to a particular content. If successful, she has managed to communicate on some minimal level, regardless of the outcome of her further goal. Thus, while the act of drawing the audience's attention to a content is not basic with respect to *goal-directed action*—in that it characterizes a (normally) proximate, rather than ultimate, communicative goal—it is *metaphysically* basic in that achieving this proximate goal is a necessary condition for achieving the ultimate communicative goal, whatever that may be.

Let me say a few things about how I'm thinking about attention. I neither offer nor commit to any particular theory of attention here, as this would take me far beyond the scope of this project and of my expertise. Rather, I take it on faith that we have a strong enough pre-theoretical grasp of the nature of attention that it can be put to use as an unanalyzed concept in the theory. Nonetheless, I adopt the following assumptions about its characteristic features:

1. Attention is directed toward something.[4] For example, when driving in heavy traffic, a person may direct their attention toward the *activity* that they are performing—namely, driving. If their cell phone rings, their attention may be redirected toward an *object*, namely the cell phone. This, in turn, might lead to their attention being redirected toward a *proposition*—that it is illegal and unsafe to hold a phone while driving—or a *question*, such as whether they should answer it. As the phone continues to ring, the driver's attention may then be directed toward a *property*, such as the shrillness of the ringtone, or toward the unpleasant phenomenal *experience* of hearing it. In the theory I present below, the objects of attention are both the contents of speech acts, and (derivatively) the contents of conventionalized linguistic expressions.

2. Attention selects its objects as targets for action.[5] (Here we can understand "action" broadly to include psychological actions involving, e.g., attitudes and memory.) When one's attention is directed toward one's phone, for instance, the phone is selected as a target for action—e.g., silencing it, picking it up, or throwing it across the room, as the case may be. In the case of linguistic communication, we direct our

audience's attention to a content, usually in order that they will select this content as the object of yet another kind of action. In the case of assertion, for example, the speaker directs her audience's attention to a content in order that she may select it as a target for belief.

3. Attention need not be perceptual. Though the majority of empirical work on attention focuses on perceptual attention, I take it for granted that one can also attend to, e.g., questions, thoughts, memories, and emotions. It is worth noting, however, that perceptual attention may play a privileged role in language use, as early forms of communication in phylogeny (the development of the species) and ontogeny (the development of the individual) involve directing the perceptual attention of the audience to an object in a shared visual field.[6]

4. Attention is not an all-or-nothing affair, but comes in degrees.[7] The degree of attention that we direct toward an object may determine the degree to which we are able to select that object as a target for action. For instance, suppose you are in the audience at a philosophy talk and you focus your attention to a very high degree on the content of the talk. As a result, you will be able to select that content for a range of high-level actions. For instance, you may develop an objection or an idea for future avenues of research, and synthesize the ideas well enough to have a fruitful exchange with the speaker during the Q & A. On the other hand, maybe you didn't get enough sleep the night before and ended up focusing on the talk to a lesser degree, drifting in and out with tide of your waning energy. In this case, you might still be able to select the content of the talk as a target for action, but to a lesser degree. Perhaps, for instance, you get the gist well enough to form an opinion and give a summary to an absent colleague, but not well enough to offer an interesting or helpful contribution during the Q & A. Or perhaps you are the type of person who uses their laptop during a philosophy talk. In such a case, we can imagine, you direct your attention very minimally to the content of the talk, as it is being pulled elsewhere. You are still able to select that content for action, but to a very limited degree. Perhaps, for instance, you can identify the topic of the talk, but not its thesis or central argument.

5. Attention is structured. As Watzl (2017) notes, when we pay attention to multiple things simultaneously, we do not experience them as separate and disjointed, but rather as standing in structural relations of relative peripherality and centrality. When a subject dances, for instance, she may attend to different parts of her body, her breath, the music, and her partner. The activity of dancing consists, at least in part, of organizing, integrating, and coordinating her attention to these objects so that she can prioritize them as targets for action in the appropriate ways. Similarly, the contents of speech are structured in relations of

relative peripherality and centrality, with some being foregrounded and others backgrounded.

6. Perhaps most controversially, I assume that attention need not be conscious.[8] Given that attention is not an all-or-nothing affair, one may be able to attend to something to a degree that is significant enough to select for action—but nonetheless small enough that it does not rise to the level of consciousness. I suggest that speech acts may also draw an audience's attention to contents in ways that do not rise to the level of consciousness; nonetheless those contents may become targets for action involving, e.g., memory, attitudes, behavior, etc.

By characterizing the basic speech act of meaning in terms of the goal of attention-direction, my account unifies a growing narrative about the role of attention in communication. For instance, Campbell (2002), Dickie (2011) and Dickie (2015), building on ideas from Bertrand Russell (1913), argue that attention explains our ability to use and interpret demonstrative expressions.[9] Tomasello (2006) has argued that joint attention plays a crucial role in early language development and early forms of communication. Bittner (2009), Bittner (2011), and Murray (2014) have put the notion of attention to work in modeling the introduction of discourse referents. I propose that the role of attention in a theory of meaning is broader still; it is the mechanism by which conventional language is constructed and sustained. In this section I refine this account of meaning, before going on to show how it fits into broader pragmatic and metasemantic theories.

I define this basic speech act of meaning as follows:

> *Meaning*: In uttering *e*, a speaker *S* means *m* just in case for some audience *A*, *S* intends to direct *A*'s attention to *m* at least partially on the basis of her utterance of *e*.[10]

The kinds of things that can stand in for *m* above are those things that can be the objects of attention.[11] Though I won't take a stand on exactly what such objects are, I take it that they include propositions, properties, and ordinary objects.

This account of meaning is broadly Gricean in the sense that it appeals to the intentions of the speaker. This may be off-putting to those who find the Gricean approach too cognitively demanding, so let me attempt to allay these concerns. First, it is important to note that, though this fact is often ignored, Grice was thinking of intentions in a rather weak sense:

> I must disclaim any intention of peopling all our talking life with armies of complicated psychological occurrences … Explicitly formulated linguistic (or quasilinguistic) intentions are no doubt

comparatively rare. In their absence we would seem to rely on very much the same kinds of criteria as we do in the case of non-linguistic intentions … we are presumed to intend the normal consequences of our actions.[12]

I follow Grice in thinking of intentions in a cognitively non-demanding way. Intentions need not rise to the level of consciousness, nor be explicitly formulated. They are mental states that we regularly attribute to agents in order to explain their actions, even when agents are not aware of having such mental states.

Though my account is Gricean in its appeal to goal-directed action, it is weaker than Grice's own account in several respects. First, it does not involve overt communicative intentions; i.e., in order to perform an act of meaning, a speaker need not have an additional higher-order intention that her first-order intention be recognized. Though I make room for a class of speech acts with this feature (more on this below), it is not required for performing speech acts in general. Second, the account is weaker than Grice's in aiming at attention-direction rather than belief.[13] Attention is achieved by non-human primates, and in humans at an earlier age than belief; thus, this account of meaning requires a lower level of cognitive sophistication on the part of the audience.

Though weaker than the traditional Gricean approach in these respects, the account is nonetheless committed to a characterization of speech acts as intentional, goal-directed actions. Moreover, it characterizes these goals as audience-directed. This might raise concerns about whether the theory can account for non-intentional linguistic phenomena, as well as speech acts that are not audience-directed. I discuss such cases in Chapter 7, where I argue that both can be accommodated because they are *derivative* of intentional, audience-directed communication.

The conception of meaning defined above will form the basic theoretical building block of the foundational theory of language that I develop in the remainder of the book. In the following section, I explore distinctions that can be made between different subcategories of this basic act of meaning, and describe the role that it plays in more complex speech acts such as assertion.

5.3 Pragmatics

5.3.1 Illocutionary action

A fundamental difference between my approach and that of Grice, Lewis, and Stalnaker, is that I take the basic act of meaning described above—rather than assertion—as the foundational theoretical tool with which to

build a broader theory of linguistic communication. This is a direct result of my rejection of the conception of language as information exchange. On that conception, assertion is the most salient and important speech act; as such, it is sensible to use it as the core concept in a theory of linguistic communication. Once this assumption is rejected, however, there is no longer any motivation for using assertion to play this role.

This displacement of assertion from its privileged place in a theory of language may give some theorists a sense of unease. In particular, one might worry that the notion of meaning provided above is too weak to play various theoretical roles of assertion, such as explaining the transmission of knowledge through testimony and the interpersonal commitments we make through speech. The legitimacy of this worry, however, depends upon precisely which theoretical role is at issue. According to the picture of language outlined in this chapter, meaning replaces assertion in its role in characterizing the most basic and fundamental speech act, as well as its role in metasemantic theory. It is not intended, however, to replace the role of assertion altogether. An adequate theory of language will need to characterize a broad range of speech acts, including—but not limited to—what we often think of as the "primary" illocutionary acts of asserting, commanding, and querying. Though I won't give an analysis of any of these speech acts, in this section I provide a sketch of a general strategy for characterizing them within the proposed framework.

In order to unpack the idea that speech acts may involve a chain of goals, we can consider the example of assertion. For the sake of clarity, I adopt a toy theory of assertion according to which it is aimed at belief. The correct account of assertion is likely more nuanced, but this simplified account should do for our purposes here. I take for granted, then, that in performing an assertion, the speaker has the goal of getting their audience to believe the content of their assertion. Directing their audience's attention to this content, however, is a precondition for accomplishing this goal; one must attend to a content in order to select it for belief. Perhaps there are cases in which attention is not a precondition for belief. For instance, scientists may be able to alter beliefs by tinkering with brain states, and it is plausible that we believe logical consequences of our beliefs without attending to them. Assertion, however, is importantly different from these kinds of case in that it is *communicative*. Though I won't attempt an analysis of communication here, I take it that a characteristic feature of communicative acts that their contents are intended to be understood, and that their success conditions depend on this understanding. I am suggesting that we think of this pretheoretical notion of *understanding* in terms of attention. It follows, then, that attention-direction is characteristic of communicative action.

Because the asserter intends their audience not just to believe a content, but also to understand it, an act of assertion includes two goals: it aims to

draw the audience's attention to a content, and to elicit the audience's belief in that content. We can think of an act of assertion, then, as involving a chain of goals, with some links in the chain serving as a means to achieving others. A simple act of assertion can be thought of as a chain with two links; the first link is aimed at attention-direction, which serves as a means to achieving the second link, which is aimed at belief. I want to emphasize, however, that the chain analogy is a heuristic; the speaker need not explicitly represent these goals, or intend that they take place in a particular temporal order. Like Grice, we can think of this chain of goals—and the relation between its links—as a rational reconstruction.

We can define this toy theory of assertion as follows:

Assertion: In uttering e, S asserts m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e.
2. A to believe m.

This toy theory provides a framework for thinking about illocutionary action in relation to acts of meaning. Illocutionary actions are those speech acts which contain an act of meaning as a proper part. Other kinds of illocutionary acts can be defined in an analogous way to assertion, but with variations in the type of goal(s) which is featured beyond the first clause. Again, I want to stress that this account of assertion is a toy theory which may need to be amended in several respects; my aim here is not to provide an analysis of any particular illocutionary act, but rather to give a general recipe for how they can be characterized within the framework I am developing here. Illocutionary actions, on this general picture, will have meaning as a component, but will also include further goals, which will vary according to the type of illocutionary act under consideration. Recall that one of the characteristic features of attention is that it selects contents for action. When performing an illocutionary action, the speaker wants the audience to select a content for action; in the case of assertion, this action will be belief, while in other cases it will be a different type of action.

This framework raises the question of what kinds of goals belong in a characterization of illocutionary action vs. a characterization of its perlocutionary effects. The terms *illocutionary action* and *perlocutionary effect* were first coined by Austin (1975). Though different philosophers offer different stipulative precisifications of these concepts, it is common to characterize them in roughly the following way: The goals involved in illocutionary action are constitutive, in that they make it the kind of speech act that it is. In contrast, perlocutionary effects are downstream, non-constitutive consequences of illocutionary actions. It follows that though perlocutionary

effects may be intended by the speaker, her goal of producing them should not be part of the analysis of her speech act. They are intended consequences of her illocutionary action, but are not constitutive of it. For instance, suppose that I perform an assertion, and that one of my goals in doing so is that you think well of me. According to our toy theory, this goal is not constitutive of the illocutionary act of assertion. However, if I'm successful, your thinking well of me will be one of its perlocutionary effects. I adopt this broad characterization of illocutionary action and perlocutionary effect. However, since I do not offer an analysis of any particular speech act here, I won't take a stand on which goals characterize an illocutionary action or its perlocutionary effects in any particular case. Nor will I presuppose any particular constraints on which kinds of goals characterize illocutionary action generally. Harris (2020), for instance, suggests that illocutionary actions are always directed at changing the audience's mental states, and that any goal which does not have such an aim is related rather to perlocutionary effect. While I do not rule out that Harris is correct on this point, neither am I convinced that it is true. Thus, I do not commit to any constraints on the types of goals that characterize illocutionary action. Finally, I want to emphasize that though goals aimed at perlocutionary effects are not constitutive of illocutionary action, this does not imply that they have no place in pragmatic theory. A good theory should capture the phenemona that are important to our theoretical interests, some of which may relate to the downstream effects of speech.

5.3.2 Indirect speech

I argued in previous chapters that it is important for a pragmatic theory to be able to distinguish between direct and indirect speech acts. In keeping with the idea that speech acts involve a chain of goals, we can define indirect speech by appealing to a hierarchy of communicative intentions.

DIRECT MEANING: By uttering e, S directly means m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e.

   and there is no content k (distinct from m) such that S intends

2. A to attend to k at least partially on the basis of S' utterance of e.
3. (2) is a means to (1).

INDIRECT MEANING: By uttering e, S indirectly means m just in case for some audience A, S intends

1. *A* to attend to *m* at least partially on the basis of *S'* utterance of *e*.

    and there is some content *k* (distinct from *m*) such that *S* intends

2. *A* to attend to *k* at least partially on the basis of *S'* utterance of *e*.
3. (2) is a means to (1).

Here we see that meaning functions not only as a means for selecting a content in order to perform an illocutionary act such as assertion; it also functions as a means for directing attention to a different content. The way that we distinguish indirect meaning from direct meaning is by appealing to its location in the chain of speaker's goals. A content is directly meant just in case the speaker didn't intend to direct the audience's attention to that content *via* directing her attention to another content. That is, in the case of direct meaning, attention-direction is not mediated by another act of meaning. It is mediated simply by the utterance (together, perhaps with some background information).

   A chain of goals that constitutes an act of indirect meaning may involve more than two links. I have said above that it is rarely the speaker's ultimate aim simply to direct her attention to a content—she usually wants that content to be selected for an action of some sort. This holds true of acts of indirect meaning as much as it does for acts of direct meaning. For instance, the speaker may have the additional goal of getting the audience to believe the content. In this case she will have performed an act of indirect assertion. I define direct and indirect assertion as follows.

DIRECT ASSERTION: By uttering *e, S* directly asserts *m* just in case for some audience *A*, *S* intends

1. *A* to attend to *m* at least partially on the basis of *S'* utterance of *e*
2. *A* to believe *m*

    and there is no content *k* (distinct from *m*) such that *S* intends

3. *A* to attend to *k* at least partially on the basis of *S'* utterance of *e*
4. (3) is a means to (1).

INDIRECT ASSERTION: By uttering *e, S* directly asserts *m* just in case for some audience *A*, *S* intends

1. *A* to attend to *m* at least partially on the basis of *S'* utterance of *e*
2. *A* to believe *m*

    and there is some content *k* (distinct from *m*) such that *S* intends

3.   *A* to attend to *k* at least partially on the basis of *S*' utterance of *e*
4.   (3) is a means to (1)

Again, these definitions are not offered as analyses, but rather as toy theories that will serve to illustrate a general recipe for defining indirect illocutionary actions. Other kinds of indirect illocutionary actions can be defined using this template, by changing the type of goal(s) involved in clause (2).

There may be more than one layer of indirectness. Consider, for instance, a sarcastic metaphor. My mother often says of people she doesn't like: "So-and-so's a real gem". That so-and-so is a real gem is the direct content of the utterance, in that she attempts to direct her audience's attention to this content without that attention-direction being mediated through another act of meaning. She directs her audience's attention to this content in order to direct it to another—namely, that so-and-so is rare and valuable, etc.[14] Let's call this the metaphorical content. The metaphorical content is not, however, what we might think of as the ultimate point of her utterance. Because she is being sarcastic, this second act of attention-direction serves as a means to directing her audience's attention to the further content that so-and-so is *not* rare and valuable, etc. We can call this the sarcastic content. Because both the attention-direction to the metaphorical content and the sarcastic content are mediated by other acts of meaning (involving the direct content and the metaphorical content, respectively), this utterance involves not one, but *two* acts of indirect meaning. For this reason, implicature should not be understood as identical to indirect content; my mother did not *implicate* the metaphorical content, though she did indirectly mean it. Rather, implicature seems to be more like a privileged subset of indirect speech, namely indirect assertion. Again, I don't aim to offer an analysis of implicature here; perhaps a more nuanced account is in order. However, I suggest that the distinction between indirect speech that is merely meant (i.e., involves only the goal of attention-direction) vs. indirect speech that is illocuted (i.e., involves a goal beyond intention-direction) might ultimately be of use in an analysis of implicature.

The direct content is closely related to what we might think of as the *conventional literal content* of the utterance, but only as a contingent matter. Crucially, my proposal does not build the notion of conventional content into its analyses of meaning, because the former is analyzed in terms of the latter (more on this below). In order to avoid vicious circularity in the reductive metasemantic theory I offer below, it must be possible to mean things—even indirectly—without presupposing the existence of a public language. However, as a contingent matter, the content of most indirect speech will align with the conventional literal content of the utterance. This is because the only thing mediating attention-direction in cases of direct speech is the utterance itself, together with background information

available to the audience. Usually—but not as a matter of necessity—this background information will include knowledge of the conventional content of the utterance in the interlocutors' shared public language.

5.3.3 Overt and covert meaning

As we saw in Chapter 3, speech acts vary in the extent to which the speaker cooperates with their audience in making their communicative intentions transparent. In some cases, the speaker wants the audience to recognize what they are doing, and to have that recognition aid in achieving their goal. In order to account for these differences, we'll want to tease apart two different subcategories of the broader definition of meaning proposed above: overt and covert meaning. I suggest that we can characterize overt meaning as follows:

OVERT MEANING: In uttering e, S overtly means m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e
2. A to recognize that S intends (1).[15]

COVERT MEANING: In uttering e, S covertly means m just in case for some audience A, S intends

1. A to attend to m.
2. A to fail to recognize that S intends (1).[16]

Illocutionary actions such as assertion can also fall into subcategories, according to whether the speaker's goals are overt or covert. However, since illocutionary actions involve multiple goals, these subcategories will be complex; some, but not others of the speaker's goals in performing an assertion, e.g., may be covert. For illustration, let's consider a completely overt assertion, where the speaker intends the audience to recognize both of her goals:

OVERT ASSERTION: In uttering e, S overtly asserts m just in case S intends

1. A to attend to m at least partially on the basis of S' utterance of e
2. A to recognize that S intends (1)
3. A to believe m.
4. A to recognize that S intends (3).

On the other hand, the speaker may intend that her audience recognize her goal of attention-direction, but not her goal of eliciting belief. Another way

of thinking of it is that the speaker wants the content of her speech act to be recognized, but not its illocutionary force. Examples might include passive-aggressive criticisms that are disguised as questions, and insults that are disguised as jokes. In such a case, the assertion will be partially covert:

PARTIALLY COVERT ASSERTION: In uttering *e*, *S* partially covertly asserts *m* if *S* intends

1. *A* to attend to *m* at least partially on the basis of *S*' utterance of *e*
2. *A* to recognize that *S* intends (1)
3. *A* to believe *m*.
4. *A* to fail to recognize that *S* intends (3).

This suggests a general recipe for thinking about subcategories of illocutionary action: An illocutionary action is completely overt if all of its associated goals are overt. It is partially covert if some, but not all of its goals are covert. It is completely covert if all of its goals are covert.

There is an open question about whether it is characteristic of certain speech acts that at least some of their goals are necessarily overt. For instance, one might think that there is an essentially public element to assertion (perhaps involving commitment) even if, e.g., the goal of eliciting a belief can be covert. I want to remain neutral on this point. If true, it brings out the limitations of the toy theory of assertion that I'm working with here. However, a consideration of these limitations against the backdrop of the framework I'm proposing here might point us toward a path for better understanding the nature of assertion; in order to develop an analysis of assertion—or any other illocutionary action—we must be able to identify precisely which goals it involves, and whether there are any constraints on whether these goals are overt or covert.

There is a final question about whether any speech act can be completely covert. Again, I remain neutral on this point, but I suspect that the answer is *no*. Even paradigmatically covert speech acts tend to have at least one overt element. I explore these kinds of cases in more detail in Chapter 7. In that chapter I also discuss a second way in which speech acts may be partially covert; we'll see that this is not always a matter of their comprising a mixture of overt and overt goals, but also of their comprising goals that are covert to some degree. A goal may be covert to some degree in the case where the speaker intends that it be recognized, but also intends that this recognition fail to rise to the level of knowledge. I set this issue aside for now and pick it up again in Chapter 7.

A final feature of the structure of speech acts that I want to note is that the categories of direct/indirect speech acts and covert/overt speech acts may intersect. Like direct speech acts, indirect speech acts involve a chain

of goals. Thus, they can be divided into subcategories in the same way: An indirect speech act is completely overt if all of its associated goals are overt. It is partially covert if some, but not all of its goals are covert. It is completely covert if all of its goals are covert.

5.3.4 Not-at-issue content

Conversational content, like attentional space, is structured. The contents that make up discourses—and even those that make up single speech acts—are not always on a par; usually, some information is foregrounded while other information is backgrounded. The same is true of attentional space. There may be a body of content which is at the foreground of attentional space while other content remains in the background. This commonality between attention and conversational content may suggest a new way of thinking about not-at-issue content, which is usually characterized as the backgrounded content of speech acts.[17] As before, my aim here is not to give an analysis of not-at-issue speech, which would take me beyond the scope of my project and expertise. I do, however, want to note some ways in which thinking about speech acts in terms of attention-direction may provide tools for thinking about not-at-issue content.

Consider an utterance of the following sentence:

My roommate, who is a lawyer, might be able to advise you on this.

The content of the relative clause—namely, that the speaker's roommate is a lawyer—is not at issue. Among the full range of information communicated by the utterance, this particular information is backgrounded. At the foreground is the information that this roommate may be able to advise the audience on the matter at hand. This is the at-issue content. On my account, both the at-issue and the not-at-issue contents are both meant and asserted, given that the speaker intends the audience to attend to and to believe both of these contents. This raises the question of how we might distinguish between the two. I want to suggest that the notion of attention may be of some use here. At least in certain cases, the relation between at-issue and not-at-issue content maps reasonably well onto the relation between the level of attention that the speaker intends the audience to pay to different contents. Though the speaker draws her audience's attention to both contents, the information that the roommate may be able to advise the audience is intended to occupy a privileged position in the audience's attentional space relative to the information that the speaker's roommate is a lawyer.

Something similar may be noted about a presupposition of the sentence, namely that the speaker has a roommate. This information, too, qualifies as not-at-issue content; it is backgrounded relative to the information that the

roommate may be able to advise the audience. Like Stalnaker, I suspect that presupposition is a heterogeneous phenomenon, and there may be different ways of distinguishing it from the at-issue content, depending on the situation. For instance, in many cases, the presupposition may not be meant at all, in the sense that the speaker does not intend her audience to attend to that content. We can imagine, for instance, that in uttering the sentence above, the speaker does not intend that the audience attend to the presupposition; rather, the fact that the audience will attend to it is a foreseeable consequence of their attending to the at-issue content. That is, as a consequence of attending to the proposition that the speaker's roommate will be able to advise her, the audience will likely attend to the proposition that the speaker has a roommate. Foreseeable outcomes of our goals, however, are not always intended. In such cases, presuppositions can be distinguished from at-issue content in that the latter, but not the former, are meant by the speaker. In other contexts, presuppositions may be meant; this is the case with what are often called "informative presuppositions". For instance, the speaker may intend to use her utterance to inform the audience that they have a roommate. In such cases, both the at-issue content and the presupposition are meant (and asserted), so there is a question of how to distinguish between them *qua* speech acts. Again, a degree of attention might play a role in at least some cases; usually, the speaker intends for the at-issue content to occupy a privileged place in the audience's attentional field relative to the presupposed content.[18]

Let me briefly contrast this proposal with Murray (2014), who puts attention to use in a dynamic semantic theory in order to model the introduction of discourse referents and to differentiate the role of not-at-issue content from that of at-issue content. Murray suggests that at-issue content adds a discourse-referent, which is modeled as an update to an attention set. Then, there is a separate update corresponding to the sentential force of the sentence. The spirit of this proposal is similar to mine in that it posits a chain of events in response to a successful speech act: first, a content gets added to an attention set. Then, by means of this event, something else happens that reflects the speaker's further illocutionary goals. Thus, we both think that attention plays an important role, and that is in some sense primary to illocutionary force. A second similarity is that we both want to use the idea of attention to distinguish between at-issue and not-at-issue content. For Murray, the at-issue content is added to an attention set, while the not-at-issue content is not. This explains why the former, but not the latter, is available for anaphora, direct challenging, etc.[19] While I agree that attention can be used to distinguish between at-issue and not-at-issue content, I think Murray gives the wrong assessment regarding *how*. The not-at-issue content of a successful speech act must reach the audience's attentional space in some way, as it is available to be selected for action. For instance, the audience might come to believe that content. They might even

challenge that content (this may involve more of a disruption to the conversation than in the case of challenging at-issue content, but it can be done). Moreover, it's hard to explain what the role of not-at-issue content would be in communication if it completely escaped the audience's attentional radar and was not available to select for action of any kind. A better way to mark the distinction, which I've offered above, posits a difference in degree rather than a difference in kind: the audience directs their attention to not-at-issue content to lesser degree than the at-issue content, rather than not at all. On the assumption that degree of attention is tied to degree of availability for selection for action, this still leaves us with the resources for explaining why at-issue content is more easily available for anaphora and direct challenge.[20]

5.4 Metasemantics

5.4.1 Meaning conventionalized

In this section I propose a metasemantic theory which makes use of the basic speech act of meaning articulated above. This part of the theory adopts Lewis' general framework for thinking about linguistic convention, which can be summarized with the following claim: a language L is the language of a community C just in case C participates in a convention of use in L. Of course, Lewis had a particular analysis of what use in L amounts to—namely, truthfulness and trust in L. I adopt his general metasemantic framework, but will analyse use in L in terms of overt direct meaning. This substitution results in the following metasemantic theory:

METASEMANTIC THEORY: A language L is a language of a community C just in case C participates in a convention of overt direct meaning in L.[21]

where to perform an act of overt direct meaning in L is to overtly directly mean $L(e)$ in uttering an expression e of L.

While I won't offer an analysis of convention, I adopt as a working theory a simplified version of Lewis' account, according to which a convention is a regularity in actions and mental states that is sustained by expectations of conformity. While this working theory will ultimately need to be strengthened in some way, those details will not matter for our purposes here so we may set them aside. Lewis' own account incorporated a number of additional conditions, including the requirement that conventions be characterized by a predominance of common interests. While I won't adopt a stance on the rest of Lewis' constraints on convention, I reject this one in particular, allowing that conventions may be solutions to strategy problems located anywhere on a scale between (and including) pure collaboration games and zero-sum games. (I discuss this further in Chapter 6.)

Overt direct meaning is a very specific subcategory of the more basic speech act of meaning that I introduced above. So why is this subcategory in particular—rather than the category of meaning more generally—put to use in the metasemantic theory? As I demonstrate in Chapter 6, this is partly justified by the theory's predictions. The directness and overtness contraints, however, are also independently motivated.

Speakers can communicate a multitude of contents in performing a speech act, many of which may not be semantically encoded by their utterance. By restricting metasemantically relevant use to *direct* meaning, we are able to home in on the communicated content that bears a suitably tight relation to the utterance itself; it is the content that communicated on the basis of the *utterance* (together perhaps with some background information), without being mediated by *other contents* that are meant by the speaker. This close connection between direct meaning and the utterance gives us some *prima facie* reason to suppose that it plays a privileged role in determining linguistic meaning. This is the content that the speaker meant *by her words*, so to speak, rather than content that she indirectly communicates.

The restriction to *overt* meaning is motivated by the fact that only overt speech acts can be plausible candidates for conventionalization. Recall that conventions require that community members believe that the behavior in question is in fact a regularity among them—this belief is what generates expectations of future behavior and motivates conformity. Covert meaning cannot play this role for several reasons. The first is that it is not easily recognizable; when a speaker performs an act of covert meaning, she intends to direct her audience's attention without this intention being recognized. Though on occasion an audience member may succeed in recognizing the speaker's intentions in spite of her plans to the contrary, it is unlikely that this would happen often enough to produce the expectations needed to support conventionalization. Second, a regularity in covert meaning would be self-undermining. When her act of overt meaning is recognized as such, the speaker's goals are thwarted; thus, if a certain strategy for performing an act of covert meaning is regularly recognized as such, speakers will be unlikely to stick with that strategy. Any regular strategy for performing acts of covert meaning, then, will be unstable. Though neutral acts of meaning—i.e., cases in which the speaker neither tries to make her intentions manifest to the audience nor tries to hide them—are not self-undermining in this way, they share the problem of failing to be sufficiently recognizable for conventionalizion. One is normally unable to recognize others' intentions reliably and consistently unless they intentionally make it easy to do so—that is, unless these intentions are *overt*. Both because it is more easily recognizable, and because its recognition will incentivize repeated use of a strategy, *overt* direct meaning is the best candidate for conventionalization.

It is important to note that the metasemantic theory requires only that overt intentions be conventional within a linguistic community—that they form a certain kind of robust regularity. It does *not* require that such intentions characterize every single communicative exchange. Once expressions have acquired their conventionalized intentional properties, we may in certain contexts choose to communicate by exploiting common knowledge of those properties rather than relying on mind-reading (though the rampant context-sensitivity of natural languages suggests that such contexts may be rather few and far-between.)[22] However, we must explain how a language comes to have its intentional properties to begin with; this is what a Gricean metasemantics aims to do, and it does so by appeal to the role of intentions and intention-recognition. Once a linguistic convention is up and running, there is plenty of room for bootstrapping—but getting there requires a certain level of mind-reading.[23] This is the Gricean insight. Thus, though I have argued above that a viable pragmatic theory must acknowledge the existence of covert speech acts, I do think that Grice was right in thinking that overt intentions play an important role in language use; they are the metasemantic base of public languages.

One may also worry that pre-linguistic infants could lack the theory of mind necessary for performing and interpreting acts of overt direct meaning; learning a language helps them develop this capacity, the worry goes, which appears to render the account circular. A similar concern could be raised with regard to explaining language development in phylogeny; if our nearest ancestors lack the mind-reading skills that undergird public language, this poses a problem for explaining the emergence of linguistic communication.[24] But a growing body of evidence suggests that both non-human animals and pre-verbal infants possess the ability to reason about others' goals and engage in joint attention—the crucial cognitive capacities implicated in overt direct meaning.[25] Research suggests that these abilities play a central role in early language development in human children as well as primitive forms of communication in non-human animals.[26] Moreover, there is plenty of room for bootstrapping here as well. It is compatible with this account that language development in ontogeny involves scaffolding through interaction with linguistically competent adults. It is also compatible with a gradualist evolutionary picture according to which a natural signaling system could evolve into a language whose complexity and sophistication developed in tandem with the cognitive capacities of its users.[27]

5.4.2 Subsentential constituents

In the remainder of this section, I clarify some of the implications of the account. First, meanings are determined in the same way for all contentful expression types. This detail sets the theory apart from that of Grice and

Lewis, who held that conventions of use directly determine only *sentence* meanings, from which the meanings of subsentential consituents can be derived.[28] This feature is a byproduct of their conception of language as cooperative information exchange, which is standardly achieved by performing assertions which express propositions using indicative sentences. My account, in contrast, characterizes the metasemantically relevant action as overt direct meaning. The content of an act of overt direct meaning may be anything that can serve as an object of attention; this is not limited to propositions, but may also include sub-propositional contents such as objects and properties, contents of phrases, and contents of entire discourses.

Because of this, conventions of overt direct meaning determine contents for subsentential expressions in the same manner as they do for sentences: Our conventions of uttering "Elizabeth Warren" to direct an audience's attention to Elizabeth Warren and "was warned" to direct attention to the property of having been warned determine Elizabeth Warren as the meaning for the former expression and the property of having been warned as the meaning of the latter.[29] One might worry about double counting; that with conventional use of "Elizabeth Warren was warned and yet she persisted" speakers directly mean the conjunction expressed by this sentence, as well as each conjunct—in which case my proposal would wrongly predict that this sentence means at least three propositions in English. However, we typically do not intend audiences to attend to the thought that Elizabeth Warren was warned on the basis of our uttering this *complete* sentence, but rather on the basis of our uttering a *proper part* of it; namely, "Elizabeth Warren was warned". Though one utterance is part of the other, they are distinct; when we are careful about individuating utterances, double counting is not problematic.

An exception to this uniformity in meaning determination can be seen in syncategorematic expressions. Speakers presumably do not intend audiences to attend to any particular content in uttering, for example, quantifiers, connectives, and morphemes. As Szabó (2019) notes:

> We would find it hopeless to specify the meaning of the definite article or the copula in isolation. There doesn't seem to be anything we mean by uttering such expressions alone—what we mean to do with them can only be articulated in connection with our uttering other expressions as well.
>
> *(Szabó, 2019, pp. 26–27)*

Continuing a tradition in logical and linguistic theory typified by Bertrand Russell's work on definite descriptions, Szabó posits an explanatory divide in the lexicon between *content* and *function* expressions[30]—where the

meanings of the former are determined by the speech acts that they are used to perform, and the meanings of the latter are determined by their contributions to the meanings of the content expressions in which they occur.[31] I adopt this approach. On my account, content expressions are those whose meanings are determined by conventions of overt direct meaning; they are expressions like nouns, predicates, and sentences, whose meanings can serve as objects of attention. Syncategoremata, in contrast, are function expressions: their meanings are just those functions which yield the meanings of the content expressions in which they occur, given the structure and meanings of their parts. Thus, conventions of direct meaning determine meanings for the entire language, though indirectly so for syncategoremata. As Szabó points out, the latter are obtained by reverse-engineering the meanings of content expressions in which they occur to find the complex function that gets the right results with respect to its structure and the meanings of its parts.[32]

5.4.3 Context sensitivity

Finally, let me say how the account will handle context sensitivity. This requires a slight modification to the account. I've said so far that a regularity of meaning in a language L is a regularity of using e to overtly directly mean $L(e)$. We'll need to revise this in order to accommodate context sensitivity. We will take a language L to be a partial function such that, when given an expression type E, it will return a partial function $L(E)$ from tokens of E to contents. We can call this the *character* of E relative to L. We'll say, as before, that a community speaks a language L just in case its members participate in conventions of use in L. However, now we will understand use in L as follows (where E is an expression type and e is an expression token): To use L is to overtly directly mean what $L(E)$ assigns to e by uttering e.[33] Though this way of dealing with context sensitivity is non-standard, it straightforwardly maps onto the traditional conception of character as a function from contexts of utterances to contents, given that each token determines a unique context.

Because of the way I am defining conventional linguistic communication, it will be a consequence of the theory that any character function of any actual language in conventional use can be characterized by the same rule: it assigns to the utterance what the speaker meant by it, on the assumption that she is conforming to the convention. On the plausible assumption that an act of meaning must be intentional, this may introduce a worry that the account predicts that the correct semantics for any natural language will be an extreme version of what Stokke (2010) calls Intention-Sensitive Semantics. According to this controversial view, the semantic content of at least some expressions in a context are at least partially determined by

the intentions of the speaker.[34] However, this worry is ill-founded; in the account I am proposing here, a speaker's intentions on a single occasion of use play no role in metaphysically determining the content of the expression used. An expression has a certain meaning relative to a language and a context for no other reason than this is what the character function given by that language assigns to it. Speakers' intentions play a role in determining which language is in use, but not what expressions mean relative to any given language—moreover, in a conventionalist framework, what is relevant is not a single speaker's intention on any given occasion, but rather stable regularities in use across the community. So the observation that any character function of any actual language in conventional use can be characterized by the function that assigns to tokens what was meant by them is trivial—it simply notes that when a speaker uses a language in accordance with a convention of use in a language L, her communicative intentions and the character function of L will correspond.[35] This correlation is also uninformative in the sense that it doesn't help us to understand properties of natural languages or make useful predictions. A more theoretically useful representation of a set of character functions will track patterns among them which reveal as much systematicity as possible—this is done in the familiar way using various kinds of parameters such as speaker, location, standards of evaluation, etc. Though these parameters do not *determine* meaning in any deep metaphysical sense, they allow us to make classifications and predictions which serve to deepen our understanding of natural languages and language users.

5.4.4 Sentence mood

Some readers may have noticed that I've given a metasemantics that's devoid of sentence mood. This implies that on my account, sentence mood is not part of the conventional semantic content of an expression, but rather a pragmatic phenomenon. Because this is a controversial position, I will briefly explain what is at issue, why I'm attracted to the pragmatic position, and what would be required from my account in order to accommodate the semantic position.

Sentences can universally be divided into roughly three moods: declarative, interrogative, and imperative. These three sentence moods tend to correspond with three basic kinds of speech acts: assertion-like acts, query-like acts, and command-like acts, respectively. There is an active and lively debate over how to explain this correspondence; some are attracted to a semantic explanation, while others are attracted to a pragmatic explanation. According to the semantic explanation, the correspondence is conventional. The idea goes roughly like this: though there is variation among assertion-like speech acts (compare, e.g., suggestions, predictions, guesses,

assertions, etc.), they all serve the same more general function. How to characterize this more general function may vary according to the theory, but on the popular Stalnakerian model, it is to update the common ground. Similarly, question-like acts serve a common function (e.g., to update a list of Questions Under Discussion), and command-like acts serve a common function (e.g., to update a To-Do List). The semantic picture posits semantic rules that link sentence type (e.g., declarative) to general function (e.g., update the common ground), thus explaining the correspondence between sentence types and speech act categories: it is conventional. In contrast, my proposal claims that all expression types have the same conventional function: to direct attention. (The word *conventional* is key here; though we use language for other functions, this is the only one that determines semantic content.) This puts me squarely in the pragmatic camp, in the sense that I need to explain the correspondence between sentence type and speech act category without positing conventional rules.

I follow other theorists in the pragmatic camp by appealing to content types to explain this correspondence. The idea goes roughly like this: declaratives, interrogatives, and imperatives have different types of contents, e.g., propositions, sets of propositions, and properties, respectively. (Again, these contents may be characterized differently depending on the theory; e.g., by using preferences rather than properties for imperatives.) These types of contents are naturally correlated with speech act types. E.g., the contents of assertion-like speech acts are propositions, while the contents of query-like speech acts are sets of propositions, and the contents of command-like speech acts are properties. Sentence types are correlated with speech act types not because of conventional rules, but because they have the same types of content. As Roberts puts it, "The correlation … is natural. The semantic types of the root sentences lend themselves naturally to the tasks they canonically serve."[36]

Murray and Starr (2021), proponents of the semantic approach, have complained that the difference in content between sentence types is superficial, pointing out that that sets of worlds, e.g., have been be used to model both declaratives and imperatives. This complaint, however, seems to confuse the model with the real thing; the fact that we can model different entities with the same tool doesn't show that the entities in question are identical. For the objection to have bite, one would need to show that declaratives, and interrogatives have the same contents—not that they can be modeled with the same tools. One might respond here that the way we model entities can provide some evidence about their nature. While this may be true, the kind of evidence needed to establish that the contents of sentence types are identical is that they are standardly modeled identically by our best theories, for principled reasons—not simply that they *can* be modeled identically.

A major motivation for the pragmatic approach is that a sentence's type underdetermines the function that it plays in the conversation.[37] You can't, for instance, read off the speech act that a speaker performs from the fact that she used a declarative sentence (hence: *Is that a promise, or a threat?*). The pragmatic approach can account for the looseness of the correlation between sentence type and speech act type, while the semantic approach wrongly predicts that the correlation will be one-to-one. On the pragmatic approach, the correlation is, as Roberts (2018) puts it, "a function of the rules of the game"—which can be thwarted—rather than "determined by the semantics". Murray and Starr respond that this line of argument confuses a more general category of conversational functions that sentences are used to perform (which they call *sentence force*) with the specific way that a sentence is used to perform that function on a particular occasion (which they call *utterance force*). On any particular occasion, the thought goes, a declarative may be used to do a variety of things, such as to make an assertion, a supposition, or a threat; this is the utterance force. But all of those possibilities can be characterized by the same general function, which is to update the common ground in some way; this is the sentence force. They claim that while the relationship between sentence type and utterance force is one-to-many, the relationship between sentence type and sentence force is one-to-one; in order to motivate the pragmatic approach, "one must find an utterance of a declarative sentence which does not even make the relevant proposition common ground in this weaker sense [of common acceptance]. No such examples have been actually produced".[38] Because I take myself to have produced such examples in Chapter 4, where I argued that many speech acts do not aim to update a shared content set, I am not convinced by this reply. It is also worth noting that Murray's (2014) strategy for preserving the one-to-one mapping between sentence type and sentence force can appear strained; declaratives that, on the surface of things, don't seem to update the common ground with their at-issue contents, are described as updating it with the current context set (which is empirically equivalent to no update).[39] While this may be a good approach to modeling, it's questionable whether it is a matter of limning natural boundaries.

There is much more to say on behalf of both the semantic and the pragmatic approaches to sentence mood, which I can't do justice to here. While the metasemantic theory I've offered above commits me to taking sides on this controversial issue, it places me on the side that I find most convincing. Moreover, the debate tends to take place independently from metasemantic considerations (instead, citing defeasible methodological principles), so the fact that the pragmatic approach fits with a viable metasemantic account would be one more consideration in its favor. If, as the debate unfolds, it turns out that considerations in support of the semantic approach turn out to be so strong as to constrain metasemantic theorizing, the account I've

offered above can be reconceived as a partial theory—giving the grounding conditions for semantic content, but needing supplementation to account for force.[40]

5.5 Conclusion

In conclusion, let me say a bit about where we've been and where we are going. In this chapter, I've introduced a basic speech act type called *meaning*, whose characteristic aim involves directing the audience's attention to a content. I've claimed that this aim is the common ingredient to all communicative speech acts, which may vary widely in their ultimate goals and levels of cooperation. This basic act of meaning serves as the fundamental building block for a broader foundational theory of linguistic communication that rejects the idealized conception of language as cooperative information exchange.

In particular, it can be used in a pragmatic theory to define a broader range of speech act types. Illocutionary actions, for instance, can be characterized as speech acts which have an act of meaning as a proper part. In performing such actions, a speaker not only attempts to direct her audience's attention to a content, but also intends her audience to select that content as a target for action. Illocutionary actions can be distinguished by the kind of action the speaker intends the audience to perform (beyond that of attention-direction). In presenting this broader pragmatic theory, I outlined two important distinctions between speech act types: the direct/indirect distinction, and the overt/covert distinction. The first distinction appeals to a hierarchy of communicative intentions; when a speaker means a content indirectly, attention-direction is mediated by another content that she means. When she means a content directly, attention-directed is mediated by the utterance alone (perhaps with contextual background information). The second distinction tracks whether or not the speaker intends her communicative goals to be recognized. In performing an overt speech act, the speaker intends that the audience recognize her aims, while in performing a covert speech act she intends that the audience fail to recognize them. Because speech acts may involve many goals, each of which may be overt/covert, the covertness of a speech act may be a matter of degree.

The conception of meaning as attention-direction can also be used as a building block in a metasemantic theory which aims to provide the grounding conditions for public language use. Above, I presented a metasemantic account according to which a language L is the language of a community C just in case C participates in a convention of overt direct meaning in L. In the remainder of the book, I spell out the details of the pragmatic and the metasemantic theory in more detail, as I demonstrate how they are able to overcome problems faced by the idealized work of Grice, Lewis, and Stalnaker.

Notes

1 Some of these ideas were first presented in Keiser (2020).
2 See, e.g., Carr (2010), Gazzaley and Rosen (2016), Hari (2022), Newport (2016), Williams (2018), Wu (2016).
3 Quine (2013).
4 See Locke (1689), James (1890), Bradley (1902), Broadbent (1958), Campbell (2002), Chalmers (2004).
5 See James (1890), Broadbent (1958), Allport (1987), Prinz (2011), Smithies (2011), Wu (2011).
6 Though contemporary discussion has largely focused on perceptual attention, there is a rich philosophical literature concerned with non-perceptual attention as it relates to epistemic and moral activity. See, for example, Descartes (1641), Locke (1689), James (1890), Weil (1986, p. 214), and Bommarito (2013).
7 See Watzl (2017).
8 Evidence of unconscious attention comes primarily from work by Kentridge, Heywood, and Weiskrantz (1999) with the blindsighted patient GY. For similar experiments with neurologically intact subjects, see Jiang et al. (2006), Kentridge, Nijboer, and Heywood (2008), and Norman, Heywood, and Kentridge (2013). See Kentridge (2011) and Wu (2014) for discussion.
9 See Campbell (2002), Dickie (2011), and Russell (1913, p. 40).
10 The clause "at least partially on the basis of" comes from Grice (1989), and is meant to indicate that the attention-direction is a consequence of the utterance. The wording leaves it open that it may be a reasons-based consequence, not only a merely causal consequence.
11 I'm ignoring indeterminacy of meaning here, which is not a special problem for my account. I'm confident that my proposal can be adapted to accommodate whatever turns out to be the best solution. See Buchanan (2010) for discussion.
12 Grice (1989, p. 222).
13 Though Grice includes more actions than just belief, he seems to treat this as the central case.
14 This is a simplification. In using metaphor, the speaker does not usually intend to direct her audience's attention to a unique and determinate content, but rather to an open-ended and structured range of contents that map onto the source domain. I discuss this more in Chapter 7.
15 Like intention, I am thinking of recognition—which we'll treat as a kind of factive belief—in a cognitively undemanding way. I assume minimally that to recognize x one must attend to x and thus be able to select x for action. I also assume, however, that recognition need not be conscious.
16 Acts of meaning may be neither covert nor overt; these are cases in which the speaker neither intends her intention to be recognized, nor that it fail to be recognized.
17 For a helpful discussion of the distinction between at-issue and not-at-issue content, see Tonhauser (2012).
18 Granted, this may not always be the case. Sometimes this information is supposed to draw the audience's attention to a stronger degree even than the at-issue content.
19 Murray (2014) also marks this distinction by claiming that the not-at-issue content gets added automatically to the common ground, while the at-issue content must first be considered and accepted. (Stanley (2006) makes use of this idea in his analysis of propaganda.) While this is an interesting proposal, I would resist the idea that any kind of semantic content changes the mental states of the audience "automatically".
20 Cf. Sperber and Wilson (1986).

21 This definition is compatible with there being many languages belonging to a single community.
22 The literature on context-sensitivity is broad, but see Lakoff and Johnson (1980), Travis (1985), Bezuidenhout (2002), Récanati (2004), Scott-Phillips (2017).
23 Natural meaning plausibly preceded non-natural meaning in the bootstrapping chain. See Bar-On (2013) and Sterelny (2017) for discussion. This is compatible with the thesis that conventions of non-natural meaning ground linguistic meaning.
24 See Bar-On (2013) and Moore (2017a) and Moore (2017b) for discussion.
25 See Onishi and Baillargeon (2005), Surian, Caldi, and Sperber (2007), Buttelmann, Carpenter, and Tomasello (2009), and Rubio-Fernández, and Geurts (2015). See Moore (2017a) for discussion.
26 Tomasello and Farrar (1986).
27 See Bar-On (2013), Moore (2017a), Moore (2017b), Sterelny (2017), and Planer (2017).
28 This strategy leads to the problem of referential indeterminacy, which I discuss in Keiser (2020). There I argue that this problem can be resolved using the metasemantic theory I offer here.
29 There are likely both top-down and bottom-up *causal* explanations for why speakers have the conventions of locution that they do. Cf. Szabó (2019).
30 Russell (1905) calls the latter "incomplete".
31 You might think that it *is* possible to attend to these functions, in which case they would get the same metasemantic treatment as the content expressions.
32 I provide a more detailed discussion in Keiser (2020).
33 I go into more detail about how I'm thinking of the metaphysics of languages in Keiser (n.d.).
34 Borg (2006) calls this "B-Style Intention Based Semantics".
35 A speaker may instead use an expression unconventionally; she may use *e* to mean something other than what is assigned to it by *L(E)*.
36 Proponents of the pragmatic approach include Zanuttini et al. (2012), Kaufmann & Poschmann (2013), and Portner (2016), and Roberts (2018).
37 See Roberts (2018, Section 4.2).
38 Murray and Starr (2021, p. 441).
39 See Murray's (2014) approach to reportatives on pp. 12–13.
40 For approaches to such grounding conditions, see García-Carpintero (2021) and Reiland (n.d.).

References

Allport, A. (1987). Selection for action: Some behavioural and neurophysiological considerations of attention and action. In A. Sanders and H. Heuer (eds.), *Perspectives on Perception and Action* (pp. 395–419). Hillsdale, NJ: Lawrence Erlbaum Associates.

Austin, J. L. (1975). *How to Do Things with Words*. Oxford: Clarendon Press.

Bar-On, D. (2013). Origins of communication: Must we "go Gricean?", *Mind and Language*, 28(3):342–75.

Bittner, Maria. (2009). *Tense, Mood, and Centering*. Rutgers University. https://festschriften.illc.uva.nl/Festschrift-JMF/papers/03_Bittner.pdf

Bittner, Maria. (2011). Time and modality without tenses or modals. In Renate Musan and Monika Rathert (eds.), *Tense across Languages* (pp. 147–188). Tübingen: de Gruyter.

Bezuidenhout, A. (2002). Truth-conditional pragmatics. *Philosophical Perspectives* 16:105–134.

Bommarito, N. (2013). Modesty as a virtue of attention. *The Philosophical Review*, 122(1):93–117.

Borg, Emma (2006). Intention-based semantics. In Ernest Lepore and Barry C. Smith (eds.), *The Oxford Handbook to the Philosophy of Language* (pp. 250–266). Oxford: Oxford University Press.

Bradley, F. H. (1902). On active attention. *Mind* 11(41):1–30.

Broadbent, D. E. (1958). *Perception and Communication*. Oxford: Pergamon Press.

Buchanan, Ray (2010). A puzzle about meaning and communication. *Noûs*, 44(2):340–371.

Buttelmann, D., Carpenter, M., and Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112(2):337–342.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Carr, Nicholas (2010). *The Shallows: What the Internet is Doing to Our Brains*. New York: W.W. Norton & Company.

Chalmers, D. J. (2004). The representational character of experience. In B. Leiter (ed.), *The Future for Philosophy* (pp. 153–181). Oxford/New York: Oxford University Press.

Descartes, R. (1641/1988). Replies to objections. In J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny (trans.), *The Philosophical Writings of Descartes* (3 volumes). Cambridge: Cambridge University Press.

Dickie, I. (2011). Visual attention fixes demonstrative reference by eliminating referential luck. In C. Mole, D. Smithies, and W. Wu (eds.), *Attention: Philosophical and Psychological Essays* (pp. 323–341). Oxford: Oxford University Press.

Dickie, Imogen (2015). *Fixing Reference*. Oxford: Oxford University Press.

Gazzaley, Adam, and Rosen, Larry (2016). *The Distracted Mind: Ancient Brains in a High Tech World*. Cambridge, MA: MIT Press.

García-Carpintero, Manuel (2021). How to understand rule-constituted kinds. *Review of Philosophy and Psychology*, 13(1):7–27.

Grice, Herbert Paul (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Harris, Daniel W. (2020). We talk to people, not contexts. *Philosophical Studies*, 177(9):2713–2733.

Hari, Johann (2022). *Stolen Focus: Why You Can't Pay Attention*. London: Bloomsbury.

James, W. (1890). *The Principles of Psychology*. New York: Dover.

Jiang, Yi, Costello, Patricia, Fang, Fang, Huang, Miner, and He, Sheng (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Pnas*, 103(45):17048–17052.

Kaufmann, M., and Poschmann, C. (2013). Embedded imperatives: Empirical evidence from colloquial German. *Language*, 89(3):619–637.

Keiser, Jessica (2020). Language without information exchange. *Mind and Language*, 37(1):22–37.

Keiser, Jessica (forthcoming). Languages and language use. *Philosophy and Phenomenological Research*.

Kentridge, Robert (2011). Attention without awareness: A brief review. In Christopher Mole, Declan Smithies, and Wayne Wu (eds.), *Attention: Philosophical and Psychological Essays* (pp. 228–246). New York: Oxford University Press.

Kentridge, R. W., Heywood, C. A., and Weiskrantz, L..(1999). Attention without awareness in blindsight. *Proceedings: Biological Sciences*, 266(1430):1805–1811.

Kentridge, R. W., Nijboer, T. C. W., and Heywood, C. A. (2008). Attended but unseen : Visual attention is not sufficient for visual awareness. *Neuropsychologia*, 46(3):864–869.

Lakoff, G., and Johnson, M. (1980). Metaphors we live by. *Journal of Aesthetics and Art Criticism*, 40(2):208–210.

Locke, J. (1689/1979). *An Essay Concerning Human Understanding*, ed. by P. Nidditch. Oxford: Oxford University Press.

Moore, R. (2017a). Social cognition, stag hunts, and the evolution of language. *Biology and Philosophy*, 32(6):797–818.

Moore, R. (2017b). Gricean communication and cognitive development. *Philosophical Quarterly*, 67(267):303–326.

Murray, Sarah E. (2014). Varieties of update. *Semantics and Pragmatics*, 7(2):1–53.

Murray, Sarah. E, and Starr, William B. (2021). The structure of communicative acts. *Linguistics and Philosophy*, 44(2):425–474.

Newport, Cal (2016). *Deep Work: Rules for Focused Success in a Distracted World*. New York: Grand Central Publishing.

Norman, L. J., Heywood, C. A., and Kentridge, R. W. (2013). Object-based attention without awareness. *Psychological Science*, 24(6):836–843.

Onishi, K. H. and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719):255–258.

Planer, R. (2017). Protolanguage might have evolved before ostensive communication. *Biological Theory*, 12(2):72–84.

Portner, P. (2016). Imperatives. In M. Aloni and P. Dekker (eds.), *The Cambridge Handbook of Semantics* (pp. 593–626). Cambridge: Cambridge University Press.

Prinz, J. (2011). Visual attention fixes demonstrative reference by eliminating referential luck. In C. Mole, D. Smithies, and W. Wu (eds.), *Attention: Philosophical and Psychological Essays* (pp. 174–204). Oxford: Oxford University Press.

Quine, W. V. (2013). In Conversation with W.V. Quine—The Boolos Panel. https://www.youtube.com/watch?v=2NmVVu7gF9c

Récanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.

Reiland, I. (n.d.). Meanings, conventions, and rules. (unpublished ms)

Roberts, C. (2018). Speech acts in discourse context. In D. Fogal, D. Harris, and M. Moss (eds.), *New Work Speech Acts* (pp. 317–359). New York: Oxford University Press.

Rubio-Fernández, P., and Geurts, B. (2015). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, 7(4):835–850.

Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.

Russell, B. (1913/1987). *Theory of Knowledge*, ed. by E. Eames. Oxford: Routledge.

Scott-Phillips, T. (2017). Pragmatics and the aims of language evolution. *Psychonomic Bulletin & Review*, 24(1):186–189.

Smithies, D. (2011). Attention is rational access consciousness. In C. Mole, D. Smithies, and W. Wu (eds.), *Attention: Philosophical and Psychological Essays* (pp. 247–273). Oxford: Oxford University Press.

Sperber, Dan, and Wilson, Deirdre (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stanley, J. (2006). *How Propaganda Works*. Princeton: Princeton University Press.

Sterelny, K. (2017). From code to speaker meaning. *Biology and Philosophy*, 32(6):819–838.

Stokke, A. (2010). Intention-sensitive semantics. *Synthese*, 175(3):383–404.

Surian, L., Caldi, S., and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7):580–586.

Szabó, Z. (2019). Semantic explanations. In E. Lepore and D. Sosa (eds.), *The Oxford Studies in Philosophy of Language* (pp. 240–276). Oxford: Oxford University Press.

Tomasello, M. (2006). Why don't apes point? In N. Enfield and S. Levinson (eds.), *Roots of Human Sociality: Culture, Cognition and Interaction* (pp. 506–524). Oxford: Berg.

Tomasello, M., and Farrar, M. (1986). Joint attention and early language. *Child Development*, 57(6):1454–1463.

Tonhauser, Judith (2012). Diagnosing (not-) at-issue content. *Proceedings of Semantics of Under-represented Languages of the Americas (SULA)*, 6:239–254.

Travis, C. (1985). On what is strictly speaking true. *Canadian Journal of Philosophy*, 15(2):187–229.

Watzl, Sebastian (2017). *Structuring Mind. The Nature of Attention and How It Shapes Consciousness*. Oxford: Oxford University Press.

Weil, S. (1986). *Simone Weil: An Anthology*, ed. by S. Miles. New York: Grove Press.

Williams, James (2018). *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge: Cambridge University Press.

Wu, Tim (2016). *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. New York: Knopf.

Wu, Wayne (2011). Attention as selection for action. In C. Mole, D. Smithies, and W. Wu (eds.), *Attention: Philosophical and Psychological Essays* (pp. 97–116). Oxford: Oxford University Press.

Wu, Wayne (2014). *Attention*. London: Routledge.

Zanuttini, R., Pak, M., and Portner, P. (2012). A syntactic analysis of interpretive restrictions on imperative, promissive, and exhortative subjects. *Natural Language & Linguistic Theory*, 30(4):1231–1274.

6

NON-IDEAL METASEMANTICS

6.1 Overview

There are a number of ways that the metasemantic picture outlined in Chapter 5 leans heavily on Lewis' ideas about language. I take my proposal to broadly conform to the spirit of Lewis' metasemantic theory—adopting its basic structure, while differing in what turn out to be theoretically significant details. Before launching into a discussion of those details, I want to briefly highlight the ways in which the proposal follows Lewis' general approach. I draw from Lewis in taking it to be the case that communities use the public languages they do in virtue of their participation in conventions of use in those languages. To my mind, this claim characterizes the basic structure of Lewis' metasemantic framework, and I take this broad picture on board. I also take on board Lewis' idea that conventions serve particular functions, and that facilitating the ease and efficiency of communication is at least one function of public language. People sometimes misread Lewis' metasemantic theory as being committed to the claim that the function of public language is to make communication *possible*, rather than merely more *efficient*; this is to misunderstand the project, which crucially depends on communication being possible prior to conventionalization.[1] Like Grice and Stalnaker, Lewis is committed to the principle of the *Autonomy of Pragmatics*, according to which it is possible to describe communication in terms of the actions and mental states of language users independently of conventions. Conventions are then understood as certain kinds of regularities in those actions and mental states, which facilitate ease of communication across a group of agents. I follow Lewis in adopting

this basic metaphysical structure, which I've been calling a *Pragmatics First* picture of language.

6.2 The theory

Most of the differences between Lewis' metasemantic theory and the one I'm proposing here can be traced back to differences in the pre-theoretical conceptions of language that furnish our respective launching points. Like Grice and Stalnaker, Lewis adopts an idealized picture of language as a mechanism for cooperative information exchange. Because I do not share this picture, I reject the particular ways in which it is manifested in Lewis' metasemantic theory—namely, in what I've called the *three Lewisian idealizations*. Most of this section will be dedicated to exploring the aspects of my proposal that differ from Lewis' as a result of rejecting these idealizations. Before I begin that discussion, however, I want to briefly note another difference between the two theories, which those familiar with Lewis' work may notice.

Because the regularity in behavior described by Lewis' theory involves the roles of both speaker and audience, it captures a direct coordination between the present speaker and audience of any particular exchange, as well as an indirect coordination along four additional dimensions: between present speakers and past speakers, present speakers and past audiences, present audiences and past speakers, and present audiences and past audiences. While this is the version of the theory presented in Lewis (1975), Lewis (1969) had originally conceived of linguistic conventions as involving only indirect coordination across time, and along one direction; between present speakers and past speakers. I prefer Lewis' original strategy for its simplicity, and adopt it here. Nothing much rides on this decision, however, as the account can easily be adapted to include the role of the audience and thus mirror Lewis' five-way coordination. This could be done as follows: Recall that for Lewis, the role of the speaker is more theoretically basic, given that the role of the audience is cashed out in terms of expecting speakers to conform to their part. Given that on my account the speaker's role is analyzed in terms of overt direct meaning in a language L, the audience's role would be analyzed in terms of expecting speakers to conform to a regularity of overt direct meaning in L. We could call this *interpretation in L*. In this way, the account can be adapted to mirror Lewis', for those who prefer his approach. Nonetheless, let me briefly explain how the one-way coordination simplifies the account without, to my mind, losing much. Though the account doesn't directly capture the audience's role in communication by including it in the regularity in behavior that undergirds linguistic meaning, it does so indirectly. Recall that for a regularity in behavior to qualify as conventional, there must be expectations of conformity. So the very fact

that the speaker's side is conventionalized within a community entails that audiences are doing their part; it entails that members of the community expect speakers to conform to a regularity of direct meaning in L—which, we just said, is the audience's role. The mechanism by which these expectations remain stable involves indirect coordination between speakers across time; by participating in a regularity of overt direct meaning in the same language L, they generate expectations that they will do so in the future. So, the fact that speakers coordinate across conversations diachronically allows for coordination between speakers and audience members synchronically. Thus, by including only the speaker's part in our analysis of the conventional regularity that constitutes public language use, we indirectly capture the audience's role while keeping the theory more streamlined. This difference between my approach and Lewis' concerns a technical detail—albeit an interesting one—and there is certainly more to be explored here.[2] However, given that the issue is peripheral to my main concerns in this book, and given that I've shown the account can be adapted to mirror Lewis', I won't spend more time discussing my preferred approach here.

A task that is more relevant to the present project is that of showing how the proposal departs from Lewis' account as a result of rejecting the three Lewisian idealizations outlined above. The first idealization, recall, is the identification of the regularity in behavior constituting linguistic convention with information exchange in a language L. In contrast, my account identifies the relevant regularity as overt direct meaning in a language L—a way of directing attention, rather than exchanging information. The use of language to exchange information is not, on my account, *metasemantically relevant*. That is, though members of a community may engage in a practice of information exchange, this practice does not play a role in determining which public language they share. Rather, this role is played by overt direct meaning, a more basic communicative act that is a component of a variety of linguistic practices, including—but not limited to—information exchange.

According to the second Lewisian idealization, public language serves a single metasemantically relevant function: to facilitate communication. In contrast, I claim that linguistic conventions may be solutions to multiple strategy problems, including—but not limited to—communication. The general idea is that various subsets of the community coordinate for the purpose of achieving a variety of goals such as communication, establishing/maintaining social identity, and establishing/maintaining power. A public language for a given community is sometimes a result of coordination among its various subgroups on various problems, rather than coordination between all members on a single problem.

The third Lewisian idealization claims that public language arises and is sustained by a predominance of common interests among the members of

a linguistic community. I deny that this is the case, arguing that linguistic conventions are sometimes determined by competing interests. As Lewis notes, situations of interdependent decision making can fall anywhere along a scale from pure conflict to pure cooperation. While Lewis assumes that conventions are solutions to games that fall along the cooperative end of that scale, I suggest that this need not be the case; the games of interdependent decision making that constitute public language use can creep toward the end of the scale that is characterized by conflict. There is nonetheless still a role for common interests to play in the story I am telling here. Above, I suggested that linguistic conventions are solutions to multiple strategy problems faced by subgroups of the broader linguistic community. I claim that at least in some cases these subcommunities share common interests in Lewis' sense of having roughly the same preferences concerning the different potential solutions to the strategy problem(s) that they face. However, this need not be true of the community as a whole. We'll explore the proposal in more detail below as we look at how it is able to accommodate facts about language change that posed problems for Lewis' view.

6.3 Possible linguistic conventions

We saw in Chapter 2 that because Lewis' theory identifies metasemantically relevant communication with information exchange (the first Lewisian idealization), it is unable to accommodate the possibility of a public language L being used predominately for non-literal or non-serious speech. In both cases, the community will fail to exhibit a regularity (and hence a convention) of information exchange in L. In the former case, this is because they aren't exchanging information at all, while in the latter case it's because they are not exchanging information *in L*. Because such scenarios do seem possible—and are to some extent borne out in our actual linguistic practices—they pose *prima facie* counterexamples to the theory. Deliberately restricting the scope of the theory to contexts of information exchange in order to avoid these sorts of counterexamples, I argued, is unmotivated. On my proposal, metasemantically relevant use is identified with overt direct meaning, a more basic and liberal communicative act than that of information exchange. Here, again, is the definition. (Since it is analyzed in terms of overt meaning and direct meaning, I include those definitions for reference.)

> *Overt direct meaning*: In uttering e, S overtly directly means m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e.
2. A to recognize that S intends (1).

and there is no content k (distinct from m) such that S intends

3. A to attend to k at least partially on the basis of S' utterance of e.
4. (2) is a means to (1).

Overt meaning: In uttering e, S overtly means m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e.
2. A to recognize that S intends (1).

Direct Meaning: In uttering e, S directly means m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e.

and there is no content k (distinct from m) such that S intends

2. A to attend to k at least partially on the basis of S' utterance of e.
3. (2) is a means to (1).

There are three main components to this definition that are relevant to its metasemantic role; the analysis of meaning in terms of attention-direction, the directness restriction, and the overtness restriction. While attention-direction is rarely the ultimate aim of any particular speech act, I have argued that it is a common component of all communicative speech acts. This commonality provides some initial support for treating the act of meaning as a suitable candidate for grounding linguistic conventions. The restriction to acts of *direct* meaning is motivated by the fact that it tracks the part of the communicative exchange where content is identified on the basis of the utterance, rather than on the basis of other things that were meant. Direct meaning, then, is the element of the communicative act in which we see the tightest connection between an utterance and a content. This tight connection suggests that direct meaning, in particular, plays an important role in determining conventionally encoded content. Finally, the restriction to *overt* meaning is motivated by the fact that it is the only category of meaning that is likely to be sufficiently recognizable to form the basis of a stable social convention. Recall that for a behavior to be conventionalized, it must be recognizable, else we could never develop the right kinds of expectations needed make the regularity stable.[3] Acts of overt meaning are designed to be recognized, while acts of covert meaning are designed to be hidden—thus, the former, but not the latter can plausibly form the basis of a convention.[4] Both because it is more easily recognizable, and because its

recognition will incentivize repeated use of a strategy, *overt* direct meaning is the best candidate for conventionalization. As I argued in Chapter 5, these considerations provide independent motivation for taking overt direct meaning to be the type of use underlying linguistic conventions. I won't further rehearse those arguments here, but will instead turn my attention to the predictions of the account, which offer additional support.

The account is at an advantage to Lewis' information-exchange theory with respect to its modal predictions. In particular, it is better able to accommodate the possibility of rampant non-literal and non-serious speech. Even when using a language L to perform indirect speech acts, language users are nonetheless performing acts of direct meaning in L. In such cases, they will be uttering expressions of L to direct the audience's attention to the contents assigned to them by L. When speaking non-literally, such attention-direction will not be the main goal of the utterance, but rather the first link in a chain of goals. For instance, in performing an indirect assertion with an expression e, the speaker intends to draw the audience's attention to L(e) in order that the audience attend to—and believe—a separate content L*(e). This account captures the fact that L is playing an important communicative role even in cases of indirect speech; it determines the initial content *on the basis of which* the audience is intended to identify the indirectly asserted content. While Lewis was forced to say that such speakers are really using L* (which bears some dubious relation to L) we preserve the explanatory function of L in non-literal speech—and correctly predict that even when used non-literally, L rather than L*, is the public language of the community.

For similar reasons, the account can accommodate the possibility of using public language primarily for the purpose of non-serious speech, which we are understanding as a broad category encompassing any type of speech act which is not aimed at information exchange. Though information exchange and non-serious speech are aimed at different goals, they nonetheless share the common component of direct meaning. In serious and non-serious speech alike, the speaker's first in a series of goals is to direct the audience's attention to a content. In contexts of information exchange, she will have the further goal of eliciting a belief in that content, while in the case of joking she will have the further goal of eliciting amusement; both cases, however, share the common element of direct meaning. Because a practice of direct meaning will be present even in cases of rampant non-serious speech, we preserve the possibility that a community of non-serious speakers could still share a public language without participating in information exchange. In contrast to Lewis' approach, there is no need for an *ad hoc* restriction to contexts of information exchange; because direct meaning is common to serious and non-serious speech alike, we aren't forced to arbitrarily commit to the idea that only one of these contexts is meaning-determining.

You might have noticed that in talking about the theory's predictions for possible speech communities, I've only mentioned *direct meaning*; the overtness condition was playing no role in generating those predictions. I'd like to argue, however, this condition is supported by other types of predictions—specifically, it seems to draw the semantics/pragmatics boundary in the right place. There is reason to think that the content of covert speech acts is not semantically encoded. Consider covert dogwhistles, discussed in Chapter 3. Using diagnostic tools from linguistics, Khoo (2017) argues—persuasively, to my mind—that the content of such speech acts is not semantically encoded. This is just what our theory predicts; by restricting metasemantically relevant language use to overt speech acts, we rule out the possibility that covert meaning gets conventionally encoded in the public language. As I argued in Chapter 5, I believe that there are independently motivated reasons for this restriction, but it is worth noting that these predictions appear to provide further support. There is a tricky issue regarding neutral meaning, however; that is, acts of meaning that are neither covert nor overt, where the speaker intends to direct her audience's attention toward a content but is disinterested as to whether or not the audience recognize this intention. Our restriction to overt meaning rules out covert meaning—which, I've argued, generates the right predictions—but it also rules out neutral cases. Is this the right result? Here I am less sure. I argued in Chapter 5 that neutral meaning cannot play this metasemantic role on its own; overt meaning will be necessary to get a language off the ground, because absent a previously established linguistic convention, acts of neutral meaning will not be sufficiently recognizable to undergird a convention. But after a linguistic convention is up and running, can neutral acts of meaning play a part in sustaining it? I'm inclined to think that they can't, but remain open to that possibility. For now, I think there is *prima facie* support for the restriction to overt meaning and so I will proceed with it; however, if future considerations reveal this restriction to be too strong, the account can easily be remedied to accommodate that.

6.4 Language change

In this section, we'll explore the implications of rejecting the second and third idealizations of the Lewisian account, according to which public language serves the single metasemantically relevant function of facilitating communication, and is undergirded by common interests.

Recall from Chapter 2 that real world language change involves three phenomena that are contrary to what would be predicted by the Lewisian account. First, linguistic communities commonly exhibit divergence without distance, in the sense that linguistic divisions develop within those communities without decreased contact between members. In the Lewisian

picture, public language use is a solution to a single strategy problem—that of communication, which is achieved when the speaker and the hearer encode/decode a message using the same semantic system. The phenomenon of divergence without distance challenges this assumption. It comes at a cost to communicative efficiency, suggesting that there are other, crucial interests at play; that public language is serving another function which competes with that of communication.

Second, linguistic communities commonly exhibit non-neutral convergence—that is, convergence between linguistic communities that pattern along socioeconomic axes. The phenomenon similarly challenges the Lewisian picture. If the sole (or predominant) problem that linguistic conventions function to solve were that of communication, we should not expect to see stable patterns of convergence in favor of the language of the powerful; given that any language is as good as any other for solving that problem, we would expect rather to see neutrality regarding the choice of solution. (Granted, we might still expect to see preferences toward certain solutions, but these would be informed not by socioeconomic factors, but by factors relevant to communicative efficiency.) Moreover, if group members' interests in solving this single problem were roughly aligned, we would not expect any particular group to have the incentive or the bargaining power to establish their language of preference. The phenomenon of non-neutral convergence, then, suggests that public language use involves a variety of incentives, and that the payoff structures of community members aren't fully aligned.

Finally, linguistic conventions commonly fail to be coordination equilibria. When a strategy problem is characterized by a predominance of common interests, its solution is a *coordination equilibrium* in the sense that no agent is better off if any agent (including herself) changes their strategy unilaterally. In the case of language use, this implies that everyone in the community prefers that everyone in the community conforms to a regularity of use in the same language. Cases of public language use in which there is some preference for nonconformity, then, challenge the Lewisian assumption that linguistic conventions are necessarily grounded in common interests.

I propose, instead, that public language may serve as a solution to multiple strategy problems, including—but not limited to—the problem of communication. Second, I propose that such strategy problems may fall anywhere on a scale between (and including) games of pure cooperation and games of pure conflict. I argue below that this broader range of background conditions provides the resources for explaining divergence without distance, non-neutral convergence, and the absence of coordination equilibria. In doing so, I offer a few how-possibly stories for examples of these phenomena discussed in Chapter 2. My argument doesn't depend on these

stories being true, and the details of the actual cases will certainly be far more complex than what I suggest. The main point I aim to make in this section is that when we drop the second and third Lewisian idealizations, we gain the resources to explain these patterns of change. The potentially contentious assumptions involved in the how-possibly stories I present below are made for the sake of simplification and clarity, and are in favor of the Lewisian account. For example, in each case I only consider a few strategy problems, while there are likely many. I also assume that the preferences of each subgroup are aligned, while that may not be the case. If these assumptions turn out to be inaccurate, that would only be further grist for my mill, as it would speak against second and third Lewisian idealizations. But again, my main concern in discussing these case studies is not to give an account of any *particular instance* of language change, but rather to argue that my metasemantic proposal has resources—unavailable to Lewis' idealized theory—to explain them.

6.4.1 Northern Cities Shift

Recall that during the great northern migration, white residents of various northern American cities changed their speech patterns through vowel shifting in order to differentiate themselves linguistically from new Black residents. For concreteness, let's call the original language of those white residents *L*, and their shifted language *L1*.[5] L1 is similar to L except for certain vowel patterns. For simplicity we'll suppose that the new Black residents also spoke L—likely, it was different to some degree, but close enough to motivate linguistic white flight. Let's call the resulting language of the broader, merged community after the vowel shift *L2*. There are different ways to think about L2. For instance, we could think of it as a coarse-grained language—something like the union of L and L1—which could each be considered dialects of the broader language. Alternatively, we might think of L2 as relativizing its rules to users, perhaps including racialization as an element of context.[6] While there are interesting theoretical issues pertaining to how best to understand L2, such details won't relevant to the discussion that follows. We'll set them aside, then, and simply think of L, L1, L2 as distinct but related languages. Under this description, the Northern Cities Shift involves at least two changes, each relating to a different way of delineating linguistic communities; the broader community shifted from L to L2, while the white subcommunity shifted from L to L1. These changes are related in the sense that the former occurred in virtue of the latter occurring. In this case, recall, we see divergence without distance and an absence of a coordination equilibrium. Thus, it runs contrary to the predictions that we'd expect from Lewis' account, given its commitments to idealizations 2 and 3. I'll now offer a how-possibly story about how this change could have

occurred. Again, my aim is merely to show that by dropping these idealizations, my proposal has resources—unavailable to Lewis's account—to explain patterns of change such as divergence without distance.

As I said above, there are at least two identifiable changes in the Northern Cities Shift, relative to different ways of delineating linguistic communities. Let's first focus on the white subcommunity, whose language changed from L to L1. Though there are likely many strategy problems involved in this change, our how-possibly story will focus on two in particular: that of communication and that of establishing social identity (in this case, establishing social identity in opposition to new Black residents). We will also assume that, at least when we restrict our attention to the white community, the payoff structures of group members regarding each of these strategy problems is roughly aligned; i.e., both of these strategy problems fall toward the cooperative end of the scale.[7] In this case, the two strategy problems—rather than the preferences of group members—are at odds. Communication will proceed most effectively if the community sticks with L; however, given that L is also the language of the new Black residents, this solution will not allow the white community members to linguistically differentiate themselves in opposition to those residents. Adopting L1 serves as a compromise. Though it involves some communicative cost, this cost will be relatively low given that L and L1 are mutually intelligible. Thus, by moving from L to L1, white residents adopt a suboptimal—but nonetheless viable—solution to the problem of communication which allows them to simultaneously solve the problem of establishing a linguistic identity in opposition to Black residents.

Now let's turn our attention to the broader speech community, which includes both white and Black residents. Here again, we'll focus on the same two strategy problems: that of communication, and that of establishing social identity. With respect to communication, we can suppose for simplicity that their payoff structures are aligned, and that the strategy problem is characterized by a predominance of common interests. (This is likely not the case, but we can work with the idealization in this particular example. In later examples we'll consider scenarios in which there are misaligned payoff structures with respect to this problem.) We can suppose that payoff structures with respect to establishing social identity, however, are misaligned—that while the white subcommunity has an interest in using language to signal racial identity, the Black subcommunity does not share this interest. Again, I want to highlight that I am not making claims about the *actual* preference structures of people involved in these historical cases—rather I am supposing that their preferences structures were a certain way in order to illustrate how divergence without distance and a lack of a coordination equilibrium could arise in such conditions. It's not unrealistic, however, to suppose that at this time in American history, the newly freed Black

residents did not share an interest in using language to signal racial identity, as this would have likely added to the already insurmountable socioeconomic obstacles in the way of establishing lives for themselves. (Granted, the advantages of solidarity that such linguistic signaling affords oppressed groups can sometimes counteract its socioeconomic disadvantages, but we will consider this type of scenario below.)

The situation faced by the broader community, on this picture, would be an example of a mixed-motive game that involves both elements of cooperation and elements of conflict. There is a certain amount of cooperation among members of the larger community in achieving communication, and a certain amount of cooperation among members of the white subcommunity in using language to signal racial identity—but competing interests among members of the broader community with respect to this latter problem. Under this scenario, the fact that there is divergence without distance and a lack of a coordination equilibrium is not surprising. For the white subcommunity, the moderate amount of divergence resulting from the change from L to L2 achieves the appropriate balance between communication and social distinction—a balance that would be upset by complete convergence. Note that given the difference in payoff structures between the two subcommunities, L2 may be an unstable solution. If the Black community has an interest in complete convergence, as we are artificially supposing, then we will expect them to eventually adopt the vowel shift. This, in turn, may cause the white community to "flee" even farther, and so forth. We'll return to this issue below.

6.4.2 Martha's Vineyard

The case of language change in Martha's Vineyard, recall, is structurally similar to the Northern Cities Shift. In this case, islanders working within the previously autonomous island fishing and farming industries started to speak more like their grandparents in response to a growing tourist industry. This change in their speech worked to undo the convergence that had occurred between the speech of islanders and mainlanders in the previous decades. Again, we can identify two changes, depending on how we delineate the linguistic community. Zooming in for the moment on the islanders, we can refer their language before the change as *L*. After the influx of mainlanders started to threaten aspects of the local economy, the speech patterns of a subset of community members—specifically those whose identity was connected to traditional island life, and who resented the changes effected by the newcomers—started to shift from L to L1. Again, L and L1 are mutually intelligible, and differ most notably in that L1 exhibits more centralized diphthongs. L1 is closer to the dialect spoken on the island during a time when it had a thriving and independent fishing economy, years before

the incursion of the mainland tourist industry.[8] As in the case of linguistic white flight, the change in the language of the broader community is a function of the change in the subcommunity. Before the change, the language of the broader community—like that of the islanders—was L. (This is a simplification; likely there were differences, but we are focusing on a point at which the speech of the islanders had lost its traditional patterns and was more or less assimilated into the speech of the mainlanders.) We can refer to the language spoken after the change as L2. Again, there are different ways of thinking of L2 and its relation to both L and L1, but we'll set that aside. Like in the case of the Northern Cities Shift, sociolinguists hypothesize that this change was driven by the desire to establish social identity among one group in opposition to another. The islanders were using language as a symbol of solidarity with one another *as islanders* in opposition to mainlanders, who posed a threat to their traditional economic system and way of life. Like the example of linguistic white flight, this case is problematic for Lewisian idealizations 2 and 3 because it exhibits divergence without distance and a lack of a coordination equilibrium. Here again, a subgroup of the broader community has changed their speech patterns in order to establish an identity in opposition to another subcommunity. This change involves a communicative cost, calling into question the idea that communication is the only strategy problem grounding public language use. And again, there is a subgroup that *disprefers* general conformity, suggesting a lack of common interests across the broader group.

The how-possibly story I want to tell here is similar to the previous one. Again, we can identify at least two strategy problems driving the language change. Focusing on the subcommunity of islanders for the moment, there is the problem of communication as well as the problem of establishing a social identity distinct from that of the mainlanders. With respect to both of these problems, we'll assume that the payoff structures of the group members are roughly aligned.[9] As before, however, the strategy problems *themselves* are at odds. Communication will proceed most effectively if the community sticks with L; however, given that L is also the language of the mainlanders, this solution will not allow the islanders to linguistically differentiate themselves in opposition to that group. Adopting L1 serves as a compromise—it imposes a small but manageable cost to communication while allowing for a linguistic identity that can be distinguished from that of the mainlanders.

When we turn our attention to the broader speech community that includes both islanders and mainlanders, we see the same strategy problems, but involving more complex payoff structures. With respect to the strategy problem of communication, we can suppose for simplicity that there is a predominance of common interests—that both the islanders and the mainlanders must be able to communicate with one another for economic

reasons connected to the growing tourist industry. Though the islanders leading the sound change are those most likely to express resentment about this industry, many of their livelihoods are nonetheless dependent on it. We can suppose that payoff structures with respect to establishing social identity, however, are misaligned in the sense that only the islanders have something to gain from signaling their identity through speech. This is merely a simplifying assumption, but it is nonetheless plausible. In the competition for the control of resources between the islanders and the mainlanders, the mainlanders have a safe lead. It is only the islanders who have an incentive to resist the *status quo*. They must be able to build connections and solidarity in order to do so, and language is one tool for achieving this. Again, divergence without distance and the lack of a coordination equilibrium are not surprising under these conditions. The moderate amount of divergence resulting from the change from L to L2 achieves the appropriate balance between communication and social distinction for the islanders—a balance that would be upset by complete convergence.

6.4.3 Upward mobility

Language change is often driven by the upwardly mobile, who take on the speech patterns of higher socioeconomic groups—for instance, change tends to be led by the lower middle class. Like in previous cases, we can describe the change in different ways, depending on how we choose to delineate the linguistic community. Focusing on the lower middle-class community, we'll say that before their convergence with the language of the upper class, their language is L. After the convergence, their language is L1—which is also the language of the upper class. Now thinking about the broader community, we can think of their language before the convergence as L2, which is a course-grained language that combines L and L1. After the convergence, the language of this broader community is L1. This phenomenon is a case of non-neutral convergence, which would be surprising under Lewisian idealizations 2 and 3. If public language is functioning to solve the single strategy problem of communication, and if agents' payoff structures regarding possible solutions to this problem were roughly aligned, we should expect convergence to be arbitrary with respect to socioeconomic factors. Systematic preference for the language of the more powerful group would only make sense in the case that this language is systematically the most communicatively efficient. Though this is often offered up as a *post-hoc* rationalization, it is implausible.[10]

To illustrate how this phenomenon is explainable once we drop the Lewisian idealizations, consider the following how-possibly story. Focusing for the moment on the change occurring within the lower middle-class community, we'll suppose that there are at least two strategy problems involved:

that of communication and that of establishing social identity. We'll suppose that everyone in the community has roughly the same payoff structure with regard to communication. The second strategy problem is more complex. While in the previous cases we considered groups that had an interest in establishing an identity in opposition to another group, here we have a case where the interest is in *adopting* the identity of another group. Because they are socially mobile, members of the lower middle class will have some incentive to signal identification with a more prestigious group; this can help them to gain access to economic resources, which can aid upward movement. Even the signaling itself can constitute upward movement: symbolic displays of power (rather than, e.g., economic resources) can still buy status.[11] While we can suppose that this incentive is shared by all group members, the strategy problem likely involves an element of conflict; while each member may prefer to signal identity with the upper class, she may not prefer the same behavior from others in her group. After all, the advantages of aligning with the upper class depend on the existence of lower classes.

Now let's consider these same strategy problems in relation to the broader linguistic community. While it is plausible that all members of this broader group have at least some interest in being able to communicate with every other member, their payoff structures are misaligned. The lower socioeconomic group will likely have a greater incentive to communicate with the higher socioeconomic group, since they have more to gain from being able to do so. The upper class may benefit economically from employing lower-class workers, however costs that result from miscommunication between them are likely to be borne by the latter.[12] Moreover, often the upper class will be able to outsource communication with such workers to upper middle-class managers who are communicatively competent with both groups. Thus, there will not be as strong an incentive for members of the upper class to develop fluidity in the speech of the lower middle classes. Because members of the lower middle class aren't in the same position to outsource communication, they will experience more pressure to acquire the communicative skills needed to participate in economic exchanges with the upper class. The better they are able to communicate with the upper class, the more economic opportunities will be available to them (including managerial positions which require communicative competence with both groups). There are costs to adopting a different language; the upper class have less incentive to bear this cost, as they have less to gain from communicative efficiency with the lower middle class.

The members of the broader linguistic community also have misaligned payoff structures with respect to the problem of establishing social identity. Because they are socially mobile, members of the lower middle class will have an incentive to signal identification with the upper class, as this

can help them to gain access to economic resources, aiding upward movement. Members of the upper class, in contrast, will have little incentive to signal identity with the lower middle class, as there are few social benefits to doing so. On this picture, then, non-neutral convergence is not surprising. We should expect the middle classes to tend to accommodate the speech patterns of the upper classes, rather than *vice versa*, given that both strategy problems provide them with more incentive to do so.

One might wonder, then, why in such a situation there would not eventually be complete convergence on the language of the upper class. Again, the how-possibly story I've given here is incredibly simplified. There will likely be a myriad of identities, strategy problems and incentives that interact in complex ways. Some of these may counteract convergence. For instance, we saw in the case of the Martha's Vineyard islanders that in cases of competing economic forces, the less advantaged socioeconomic group may be incentivized to signal solidarity by linguistic means (this is sometimes referred to as "covert prestige"). So, while there will be pressure on socially mobile groups to adopt the language of the upper class, there will also be a competing incentive to exhibit covert prestige forms of the lower classes— at least in some contexts. When we add such additional factors into the how-possibly story, it no longer predicts convergence, but rather that the upwardly mobile will learn to code-switch.

Futhermore, it is plausible to suppose that not only does the upper class lack the incentive to identify with the lower classes, they have an active interest in differentiating themselves from those groups. After all, as Bourdieu notes, the symbolic value of their speech is dependent on its scarcity.[13] The result will be a lack of equilibrium—as the middle class moves toward the upper class, the upper class flees in response. As long as the lower class (who are less upwardly mobile and thus more incentivized by signaling solidarity with the lower class than identity with the upper class) do not move along with the middle classes, and as long as the upper class keeps fleeing the middle classes, there will not be convergence. And this is in fact what we do see. Language is always in flux, and the speech patterns of groups occupying distinct social locations tend *not* to converge, at least not completely or for very long. Again, the actual dynamics of class-driven language change will be far more complex than the how-possibly story I've outlined here, but given the incentive structures of the upwardly mobile with respect to communication and social signaling, we should expect to see non-neutral convergence.

6.4.4 Linguistic gender divide

In the case of the linguistic gender divide, we talked about changes in the speech of gender groups as they age. In just about every culture, there is a

marked difference between men's and women's speech forms. Studies show that early in life, boys' and girls' speech patterns are more or less identical, beginning to diverge sharply around adolescence. In considering this case, we'll focus only on the change that occurs in the language of the broader group of mixed genders. Let's call the language shared by boys and girls in early childhood L, and the one they adopt in early adulthood L2. L patterns after the speech of adult women, while L2 incorporates both L and L1—the languages of adult women and men, respectively. Again, there are various ways of describing L2, but it is useful to think of use in L2 in this case as a complementary convention: *if a woman, use L, if a man, use L1*. Such a complementary convention would be surprising under the second Lewisian idealization, given that adopting gendered speech styles within a community will impede—rather than improve—the communicative efficiency of the group as a whole

By dropping this idealization, we gain the resources to explain gendered speech styles. Our how-possibly story will suppose that this case involves both the strategy problem of communication and that of establishing social identity—in this case, gender identity. We'll suppose that with respect to both of these strategy problems, the payoff structures of group members are aligned in the sense that they share a common interest not only in being able to communicate with all other members of the group, but also in being able to signal gender linguistically.[14] As in some of the other cases we've considered, however, though community members' *interests* are aligned, the strategy problems they face are at odds; the most efficient solution to the problem of communication is for both genders to adopt the same speech styles—however, this would impede the goal of linguistic gender signaling. The adoption of L2 serves as a compromise; L and L1 are different enough to signal distinct social identities, but similar enough so that these differences impose limited barriers to communication.

I want to stress that this how-possibly story is greatly simplified. Social identities are intersectional, so the explanation for the differences between men's and women's speech in any particular speech community will likely involve intersecting social categories, such as class. Sociolinguists have noted, for instance, that women tend to make heavier use of standard speech patterns (associated with the upper class), while men make more use of covert prestige forms (associated with local, and often lower-class groups). The fact that men and women occupy different social positions might help to explain why they would have different incentives regarding identity-signaling: because of their more powerful social position, men may be able to adopt covert prestige forms with less pushback. Moreover, Trudgill (1972) and Eckert (2000) hypothesize that because women often lack concrete power, they are more incentivized to make use of the purely symbolic as a means of gaining status. That is, they have more incentive

than men to signal group membership linguistically, given they are less able to control their status using other means.[15] Thus, gendered speech divisions are likely to be a complicated mix of functions and incentive structures which involve intersecting social dynamics.

6.4.5 Linguistic colonialism

The final case we'll consider is linguistic colonialism. When distinct linguistic communities converge in a context of colonization, the colonized group tends to adopt the language of the powerful. For concreteness, let's call the language of the colonialized group before the convergence L1, and the language of the colonizers L2. After the convergence, the language of the broader group is L2.[16] The convergence on L2 is contrary to what we'd expect in the Lewisian picture, given its commitments to idealizations 2 and 3. Under these idealizations, systematic preference for the language of the more powerful group would only make sense in the case this language were systematically the most communicatively efficient. But again—though such an explanation is often offered up as a *post-hoc* rationalization, it is implausible.

Once we give up these idealizations, however, we gain the resources to explain the phenomenon of linguistic colonialism. Consider a how-possibly story according to which the broader community faces two strategy problems: that of communication and that of establishing social control. With respect to the problem of communication, we can again suppose that while all members of the broader group will have some interest in communicating with one another, they do not have equal incentives for doing so. The colonizers will have some economic interest in communication insofar as they benefit from the services of the colonized group, but such communication need not be perfectly efficient. Moreover, just as in the case of the upper class, they may be able to make use of translators competent in the languages of both groups. Finally, the colonizers will usually have the option of opting out of communication with the colonized group; while they may stand to benefit from economic interactions that require communication, they are less likely to be completely dependent on them. In contrast, the colonized group will have a stronger incentive to communicate, as they may be more economically dependent on the colonizers. They will not have the option of opting out of economic arrangements with that group, so they will be more strongly incentivized to solve the communication problem required for such arrangements. Moreover, the better they are able to communicate with the colonized groups, the better their prospects for socioeconomic success (e.g., moving to a more highly paid managerial position, understanding contracts, negotiating pay, etc.).

We'll suppose that a second strategy problem informing language change in this scenario is that of establishing and maintaining social control. In

contrast to communication—where all members of the group have an incentive to solve the problem, but to greater or lesser degrees—achieving social control is more like a zero-sum game in which one group's loss is another's gain. Because the choice of language affects a group's ability to establish social control, different subgroups will receive competing payoffs with respect to the choice of the language adopted by the broader group. Various factors contribute to the social control resulting from having one's language adopted as the *lingua franca*; e.g. the symbolic power that tends to attach to the dominant language as it becomes associated with prestige, the economic power that comes from meeting linguistic requirements for various jobs and opportunities for advancement, and the strategic power that comes from being in a greater position to understand laws, negotiations, and contracts (and thus better positioned to manipulate others). In such a situation it would be unsurprising that the powerful will be more consistently successful in getting the broader group to adopt their language. Because they have less of an incentive to achieve communication, they have less to lose if the group fails to converge on a language. The colonized group, in contrast, will not be in a position to make the economic tradeoff resulting from failure to achieve a level of convergence sufficient for communication. Moreover, the more powerful group may sometimes have the means to impose their language on the colonized group by force, given that they have control of laws governing education, etc. As I've been trying to highlight, however, the situation is likely to be more complex than the story I've presented here, in that there may be myriad other strategy problems at play. In some cases, maintaining their social identity may provide a strong enough incentive for the colonized group that they will be willing to take the economic hit resulting from refusal to adopt the language of the powerful. For instance, there are many regions—notably South Africa and Montreal—where there is a sustained effort by the less powerful group to resist linguistic colonization.[17]

It is interesting to note that in this scenario not only are there two strategy problems informing linguistic change, but they are to some degree at odds.[18] The social control that results from one group's language being adopted as the *lingua franca* depends on there *failing to be* the kind of complete convergence that would fully maximize communication. The prestige that a language offers depends on its scarcity; the job opportunities and the strategic power it affords in negotiations depends on the existence of others less competent in that language. As Bourdieu notes,

> Quite apart from the literary (and especially poetic) uses of language, it is rare in everyday life for language to function as a pure instrument of communication. The pursuit of maximum informative efficiency is only exceptionally the exclusive goal of linguistic production and

the distinctly instrumental use of language which it implies generally clashes with the often unconscious pursuit of symbolic profit.[19]

So, while I've simplified by describing the phenomenon of linguistic colonialism as one of convergence on the language of the colonizers, the optimal solution for the colonizers will fall short of complete convergence; it will be close enough to enable communication while distinct enough to allow for social control.

6.5 Conclusion

In conclusion, let me reiterate the picture I've presented above and clarify some important points about what it commits us to. My proposal draws from Lewis in taking public language to be determined by conventions of use, while rejecting the following three idealizations:

The three Lewisian idealizations:

1. The regularity in action constituting public language use is taken to be information exchange.
2. The sole (or, at least, the predominant) function of public language is taken to be that of solving the strategy problem of communication.
3. Public language is taken to be rooted in common interests.

Against the first idealization, I take the metasemantically relevant component of communication to be overt direct meaning. This is a more liberal kind of use, whose goal is characterized by attention-direction, rather than information exchange (though in some cases this attention-direction may be in service of information exchange). Not only is this move independently motivated, but it delivers the right predictions with respect to the possibility of using public language for indirect and non-serious speech, without recourse to *ad hoc* restrictions.

Against the second idealization, the proposal allows that language has many functions. Lewis' metasemantic account includes a line that specifies the type of problem that the convention is used to solve; he claims that L is the language of C just in case they participate in conventions of use *as a solution to the problem of communication*. My account, in contrast, claims that L is the language of C just in case they participate in a convention of overt direct meaning in L, without specifying the particular strategy problem(s) this convention is functioning to solve. This way, we can allow that there may be many strategy problems which determine the public language of a community, and that these may vary from case to case. One might worry that although some of these functions are variable and contingent, others—such as communication—are not. That is, one might

have the intuition that it is impossible for a community to have a public language without using it to communicate, even if—as I have argued—there are additional metasemantically relevant functions. I share this intuition, and believe that it is entailed by my proposal, even if not explicitly packed into it. If language users have a convention of overt direct meaning in L, then it follows that they are communicating on some level; this is because it is implausible that a community would participate in such a convention without systematically achieving communicative success. Granted, whether or not one accepts this claim depends on how they are understanding *communication*. While I won't attempt to give an analysis here, it is noteworthy that in most accounts, communication involves the transmission of a thought from one mind to another—something that occurs when an act of overt direct meaning is successful. A convention of overt direct meaning is unlikely to be sustained if these acts are not successful. Thus, by declining to explicitly specify its function in our metasemantic theory, we nonetheless preserve the result that public language is necessarily used for communication while allowing for flexibility with regard to its other functions. Such functions may include those discussed above—e.g., establishing/maintaining social control and social identity, as well as others not considered here—and vary depending on the circumstances. Given that these other functions don't seem to be entailed by the theory in the same way as communication, we preserve the intuition that the latter plays a privileged role in determining the language that we use without committing ourselves to the idea that it is the *only* role.

Before turning to the third idealization, let me briefly address a way that the friend of Lewis might be tempted to respond to my criticism of the second idealization. The ostensibly alternative functions of public language that I've identified, the response goes, are in fact *kinds* of communication; establishing social identity is a communicative act, and establishing social control involves the communication of dominance. There are two problems with this response. First, while I grant that there is communication involved in establishing both social identity and social control, neither are *reducible* to communication. Establishing social identity through one's linguistic choices is not simply a matter of communicating the identity that one already has, but an active process of developing that identity.[20] Gender identities, e.g., are not just *signaled* by speech styles, but partially *constituted* by them. Similarly, while communicating dominance might be an aspect of establishing social control, there are other aspects—such as the unequal distribution of resources—that may function without a communicative component. The second problem is that in the examples considered above, the information communicated by practices of establishing social control and social identity is, at least for the most part, pragmatically communicated rather than semantically encoded. Thus, these acts of communication

can't simply be folded into the Lewisian theory without drastically—and adversely—affecting its predictions, given that Lewis identifies the conventionally communicated content with the semantic content of the language.

Against the third idealization, my proposal preserves Lewis' idea that language use involves a game of interdependent decision making, while dropping the assumption that this game is fully cooperative. Rather, language use may involve many strategy games—each of which may fall anywhere along a scale between full cooperation and full conflict. In other words, what Lewis took to be the basic case, we are taking as a limiting case. We've explained and justified this idea by showing that language serves a range of functions relative to which members of a linguistic community may have competing payoff structures. While the theory doesn't rule out common interests, neither does it insist on them. It is somewhat surprising that Lewis relies so heavily on Schelling (1960) while missing his warnings about idealization. Schelling criticized the orthodoxy for taking the limiting case of cooperation as the starting point for theorizing about strategy games, pointing out that many—perhaps most—strategy games involve at least some element of cooperation. Lewis went to the opposite extreme and assumed that the strategy games involved in conventions generally—and linguistic conventions in particular—were characterized by cooperation. But he thereby seemed to miss a major point that Schelling was trying to make, which is that by taking the limiting case as "the general case rather than a special one, game theory may have overshot the level at which the most fruitful work could be done and may have defined away some of the essential ingredients of typical non-zero sum games".[21] My account draws from Lewis' insights by taking language to be a solution to a repeated strategy problem, but takes on more of Schelling's insights about the dangers of idealization, and about taking edge cases to be the norm. The result, I hope, will allow us to explore the level at which the most fruitful theorizing about the nature of public language can be done

The rejection of the second and third Lewisian idealizations provides the resources to explain how divergence without distance and non-neutral convergence might arise, and why our linguistic practices sometimes fail to exhibit a coordination equilibrium. When public language is being used for establishing distinct identities between subgroups of the linguistic population, a moderate amount of divergence allows them to achieve this without completely sacrificing communication. In such cases it will not be surprising to see divergence without distance and a preference for at least some level of nonconformity. When public language is being used as a means for establishing social control in addition to communication, socioeconomically powerful groups will have less of an incentive to sacrifice the former for the sake of the latter. In such cases it will not be surprising to see convergence in favor of the language of the powerful. Moreover, the fact that

language serves so many functions—and that our payoffs are so misaligned and complex—may at least partially explain why language is constantly changing. Given Lewisian idealizations, you'd expect that groups who are in regular contact would eventually converge on the same language.[22] But this is not what we see; dialectical differences remain strong, and this is partly explained by the fact that language is serving many, often competing functions. Moreover, given the payoff structures of some of the strategy problems involved, we should expect that it would be difficult to ever arrive at a stable equilibrium point.

While the metasemantic theory offered above should be viewed as an adaptation of Lewis' account, it should not be viewed as the second step in the process of Galilean idealization. It does not build on the three idealizations associated with his conception of language as cooperation exchange, but rather drops these idealizations altogether and builds upon what is left standing. The result is a picture of our linguistic conventions as complex practices that serve diverse functions for social groups with broad and sometimes competing interests.

Notes

1 See e.g., Davidson (1986), Lepore and Stone (2014), Lepore and Stone (2017).
2 On both Lewis' proposal and my own, the audience's role involves belief. (In Lewis' account, this is belief in the content of the utterance, while in mine it is belief about the speaker's intentions.) So, one issue that comes up in the question of whether to include the audience's role in the metasemantic account is whether belief can be conventionalized. See Lewis (1975, pp. 11–12) for discussion.
3 Cf. Leckie and Williams (2019).
4 I argued in Chapter 5 that neutral acts of meaning (neither overt nor covert) are also generally not sufficiently recognizable to form the basis of a stable social convention. More on this below.
5 For the purposes of this discussion, I'll be treating dialects as separate languages, though nothing rides on this. Moreover, many linguists question whether there is a principled distinction to be made between them. We can suppose, however, that one language is a dialect of another if it bears some suitable relation (involving, e.g., mutual intelligibility) to it.
6 In this case, the convention of use in L2 would be considered a *complementary convention*.
7 Again, if this assumption is false, it is just further grist for my mill.
8 There are some differences. In the traditional speech of the islanders, only /aw/ exhibited centralization. After the change, islanders started to centralize both /aw/ and /ay/. So, the islanders' speech returned to centralized diphthongs, but to an even stronger degree, perhaps to further emphasize the distinction between island and mainland speech.
9 Again, if this assumption is false, it is just further grist for my mill.
10 Linguists, who don't think that any languages are inherently more suited for communication than any other, call this the "Myth of the Standard".
11 See Bourdieu (1991) for discussion.
12 Thanks to Thomas Brouwer for this point.

13 See Bourdieu (1977, 1991).
14 Again, this is an idealization that works in Lewis' favor.
15 Tellingly, exceptions to the correlation between women's speech and prestige forms occur in communities where women are not upwardly mobile. See Labov (1966) and Nichols (1976).
16 The colonized group will often continue to use L1 in addition to L2, but not as a language that is *shared* with the colonizers. Though the issue of forced code-switching is interesting, I don't have the space to discuss it here. I'll limit my discussion to the language that is shared by colonized and colonizers, which is L2.
17 See https://www.culturalsurvival.org/publications/cultural-survival-quarterly/language-policy-and-oppression-south-africa and Levine (1990).
18 Thanks to Thomas Brouwer for this point.
19 Bourdieu (1991, p. 67).
20 C.f. Eckert (2000) and Coupland (2007).
21 Schelling (1960, p. 119).
22 This is not to say that we don't see this at all. Certainly, we do see languages dying, and people converging on English. But we don't see, for example, the elimination of English dialects.

References

Bourdieu, Pierre (1977). The economics of linguistic exchanges. *Social Science Information*, 16(6):645–668.

Bourdieu, Pierre (1991). *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.

Coupland, Nikolas (2007). *Style: Language Variation and Identity. (Key Topics in Linguistics)*. Cambridge, MA: Cambridge University Press.

Davidson, Donald (1986). A nice derangement of epitaphs. In Ernest Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 433—446). Malden, MA: Blackwell.

Eckert, Penelope (2000). *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Malden, MA: Wiley-Blackwell.

Labov, William (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Khoo, Justin (2017). Code words in political discourse. *Philosophical Topics*, 45(2):33–64.

Leckie, Gail, and Williams, Robert (2019). Words by convention. In David Sosa and Ernie Lepore (eds.), *Oxford Studies in Philosophy of Language Volume 1* (pp. 73–98). Oxford: Oxford University Press.

Lepore, Ernie, and Stone, Matthew (2014). *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press.

Lepore, Ernie, and Stone, Matthew (2017). Convention before communication. *Philosophical Perspectives*, 31(1):245–265.

Levine, Marc (1990). *The Reconquest of Montreal: Language Policy and Social Change in a Bilingual City*. Philadelphia, PA: Temple University Press.

Lewis, David K. (1975). Languages and language. In Keith Gunderson (ed.), *Minnesota Studies in the Philosophy of Science* (pp. 3–35). Minneapolis: University of Minnesota Press.

Lewis, David Kellogg (1969). *Convention: A Philosophical Study*. Cambridge, MA: Wiley-Blackwell.

Nichols, Patricia (1976). *Linguistic Change in Gullah: Sex, Age, and Mobility*. Redwood City, CA: Stanford University Press.

Schelling, Thomas (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Trudgill, Peter (1972). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

7

NON-IDEAL THEORY OF MEANING

7.1 Overview

There are many ways in which the picture of linguistic communication that I proposed in Chapter 5 is influenced by the work of Paul Grice. First, it takes conventional expression meaning to be reducible to facts about speaker meaning, which are in turn reducible to facts about actions and mental states, and relations between language users. In this way, my proposal conforms to what I have called the *Pragmatics-First* picture of the grounding conditions of semantic facts, which is also shared by both Lewis and Stalnaker. Because Grice's account of the relation between speaker meaning and expression meaning is left as a rough sketch, the metasemantic aspect of my proposal leans more heavily on Lewis, who explored these issues in greater depth. I have found Grice's work, however, to be more sophisticated than Lewis' in articulating the kind of use that serves as the basic building block for conventional linguistic meaning. Here I lean more heavily on Grice's ideas by explicating the act of meaning in terms of the communicative intentions of speakers, and by giving a special role to overt communicative intentions. I depart from Grice, however, in a number of other ways.

The first departure concerns the scope of the theory. For Grice, recall, the target phenomenon was rational communication. Instead, I am concerned with linguistic communication more generally. While the phenomena I discuss may qualify as rational in some accounts of rationality, I don't to commit to any particular account here. Rather, I consider my target to be linguistic communication in general, and leave it as an open question to what extent its scope extends beyond that of rational communication.

The second, and most important departure from Grice concerns my rejection of his conception of language as cooperative information exchange. In Chapter 3, I discussed particular aspects of Grice's work that relied on the assumption that communication is a cooperative enterprise. I argued that this assumption functions as a minimalist idealization, in that it is motivated by the idea that cooperation is an essential feature of rational linguistic communication. In rejecting this idealization, the details of my proposal differ from Grice's in two main ways. The first concerns its commitments regarding the process of interpretation. For Grice, audiences are able to successfully interpret speakers only by adopting the assumption that they are involved in a mutually cooperative enterprise, such as information exchange. On this account, audiences represent (even if only at a subconscious level) the means-end reasoning of the speaker, assuming that the speaker's primary end is to cooperate in this joint endeavor. My account preserves the general idea that interpretation often involves representing the speaker's means-end reasoning, but rejects the idea that the audience must assume that the speakers' goals are cooperative in order to properly interpret them. I also consider cases that do not involve reasoning about the speaker's goals at all, but instead involve automatic inferences based on implicit associations. Though Grice's Maxims of Conversation are still relevant on my proposal, they play a less central explanatory role. They are not the norms that guide rational communication in general, but rather the norms that govern a particular and limited range of communicative contexts.

This rejection of the idealization of cooperation also leads to a different understanding of the speaker side of linguistic communication. On Grice's account, acts of meaning are essentially cooperative, at least to the degree that the speaker intends to reveal her communicative intentions to her audience. That is, Grice characterizes meaning as essentially overt. He also assumes that speakers are cooperative beyond this minimal sense—that acts of meaning are performed in service to a broader shared goal (such as information exchange). In contrast, I define a more general category of meaning, which has two subcategories: overt meaning and covert meaning. The subcategory of overt meaning is very close to Grice's characterization of meaning; it entails at least some minimal level of cooperation in the sense that the speaker is trying to help the audience identify her communicative intentions. However, acts of overt meaning need not be cooperative beyond this point; a speaker, e.g., might merely want the audience to recognize that she is insulting her or blackmailing her. The subcategory of covert speaker meaning, in contrast, involves deception; in such cases the speaker intends that her audience fail to recognize what she is up to. On my account, overtness comes in degrees, with purely overt acts of meaning and purely covert acts of meaning serving as the limiting cases. In making room for covert communication, my proposal employs a more expansive account of

speaker meaning than that used by Grice; however, it nonetheless reserves a special role for overt meaning, which is the metasemantic base.

7.2 Communicative intentions

Recall that on Grice's account, acts of meaning are essentially cooperative, at least to the degree that the speaker intends to reveal her communicative intentions to her audience. In Chapter 3, we looked at a number of examples—including insinuation, dogwhistles, propaganda, and advertising—which did not fit this characterization. Each of these phenomena involve the intention to hide, or at least partially obscure one's communicative goals. My proposal departs from Grice's in allowing for covert communicative intentions, providing the resources for characterizing the types of speech acts that were problematic for his account. Before taking a closer look at these cases, I briefly remind the reader of the basic outlines of the account developed in Chapter 5. This account defines a very basic and general speech act, which I call *meaning*:

MEANING: In uttering *e*, a speaker *S* means *m* just in case for some audience *A*, *S* intends to direct *A*'s attention to *m* at least partially on the basis of her utterance of *e*.

This definition is neutral on the subject of overtness, in the sense that a speech act that fits this description qualifies an act of meaning regardless of whether the speaker intends her audience to recognize her communicative intentions. Because the distinction between overt and covert meaning is nonetheless important, however, I define the following two subcategories of meaning:

OVERT MEANING: In uttering *e*, *S* overtly means *m* just in case for some audience *A*, *S* intends

1. *A* to attend to *m* at least partially on the basis of *S*' utterance of *e*.
2. *A* to recognize that *S* intends (1).

COVERT MEANING: In uttering *e*, *S* covertly means *m* just in case for some audience *A*, *S* intends

1. *A* to attend to *m*.
2. *A* to fail to recognize that *S* intends (1).

Because overtness comes in degrees, these definitions only serve as rough characterizations of the limiting cases. The degree to which an act of

meaning is overt corresponds to the level of credence to which the speaker intends the audience's recognition of her intentions to rise. If the speaker intends that the audience retain some doubt about her intentions, then her act of meaning is covert, at least to some degree.

An act of meaning is usually not an end in itself, but a means by which to accomplish another communicative goal. Most communicative goals extend beyond that of merely calling an audience's attention to a content. For instance, an assertion might involve the additional goal of producing a belief in the audience, a question might involve the additional goal of getting the audience to produce an answer, and a request might involve the additional goal of getting the audience to produce an action. Each of these additional goals may also be overt, covert, or somewhere in between. Because some speech acts involve a chain of goals, they may be partially overt along two different dimensions. The first dimension concerns *how many* covert goals are involved in the speech act. The second dimension concerns the extent to which any of these individual goals are covert—that is, the extent to which the speaker wants the audience to harbor doubt about her intentions regarding that particular goal. Some speech acts, of course, may be partially covert along both of these dimensions. Let's consider the latter case first.

7.2.1 Insinuation

My account will treat cases of insinuation as partially covert. In order to see the details, let's consider a version of Camp's (2018) bribing case discussed in Chapter 3:

BRIBE: Person stopped while driving says to police officer "I'm in a bit of a hurry. Is there any way we can settle this right now?"[1]

On my account, this utterance qualifies as an indirect speech act, featuring both overt and covert elements. The speaker directly means the literal content of the utterance in its context, in the sense that she wants the audience to attend to the proposition that she is in a bit of a hurry, and the question of whether they can settle the issue right now. Though the case is under-specified, we can reasonably stipulate that this act of direct meaning is overt—i.e., the speaker intends for her audience to recognize this particular communicative goal. But her utterance also involves an indirect speech act, which is partially covert. This speech act is a bribe proposal, which involves a chain of two goals: that the audience attend to the proposal and respond to it in some way. (How to characterize the intended response depends on whether we want to understand the proposal as question, command, assertion, or something else. I gloss over such details.) The speaker intends for

the audience to recognize these goals—otherwise, they will not be achieved. She does not, however, want the audience's degree of certainty about these goals to rise to the level required for knowledge; rather, it is in the speaker's interests that the audience's beliefs exhibit enough instability to preserve plausible deniability. As long as the audience retains an adequate level of doubt about the speaker's goals, then—in the case that the audience's reaction to the bribe is unfavorable—the speaker will be able to provide a denial that further undermines those beliefs. Though it features overt elements, the bribe is partially covert in the sense that it involves communicative goals that are covert to some degree.

Next, we'll consider speech acts that are at least partially covert in the sense of involving a mixture of covert and overt goals. There is a question of whether there are any speech acts which are completely covert. While I see no *prima facie* reason to rule this out, most paradigmatic cases of covert speech involve at least some overt element. Thus, I will restrict my attention to partially covert cases.

7.2.2 Dogwhistles

Dogwhistles tend to involve a mixture of both overt and covert communicative goals. Consider, for example, an adaptation of an example discussed in Chapter 3:

INNER CITY: "Programs like food stamps and Medicaid represent important safety-nets for many inner city families."

Like the veiled bribe, my account classifies this utterance as an indirect assertion, which involves a chain of goals. The speaker directly asserts the literal content of the utterance, in the sense that she intends both for the audience to attend to, and to believe this content. Again, the case is underspecified, but we can reasonably stipulate that these goals are overt in the sense that they are intended to be recognized. However, there is also an indirect element to her utterance. In addition to these goals, the speaker also wants her audience to attend to various pernicious racial stereotypes, and produce certain responses (e.g., form the belief that these programs are likely to benefit Black families in particular, and to develop an opposition to them). These latter goals are covert; if the audience were to recognize the speaker's goals, as Saul, Khoo, and others have discussed, these goals would be undermined. This example differs from the case of insinuation in that the indirect aspect of the speech act is not overt to any degree; the speaker is intending to completely mask her intention to affect the audience's behavior. In this sense it is a fundamentally manipulative act.[2]

7.2.3 Propaganda

Propaganda is another type of speech that often involves both covert and overt communicative goals. Consider again the following example from Tirrell (2012):

RWANDA: The Tutsi are *inyenzi*.

Because this utterance makes use of metaphor, it qualifies on my account as an indirect speech act, involving a chain of goals. First, the speaker intends her audience to attend to the proposition that the Tutsi are inyenzi (cockroaches). This component of her broader speech act is an act of direct meaning, as she intends this goal to be accomplished on the basis of her utterance, without being mediated by other things that she means. Again, we can reasonably stipulate that this act of direct meaning is overt, in the sense that the speaker intends her audience to recognize the particular communicative goal that characterizes it. A second component of the broader speech act is the intention that the audience have a certain response to the attention-direction accomplished by the act of direct meaning; after directing their attention to the content that the Tutsi are inyenzi (cockroaches), they are intended to create a mapping between the source domain (cockroaches) and the target domain (the Tutsi). This is an act of indirect meaning, since the accomplishment of its characteristic goal is mediated by attention to the directly meant content.[3] This goal is likely to be partially covert in that the speaker intends that the audience recognize the goal of getting the audience to create a mapping between the source domain and the target domain; however, she will likely intend that the audience retain some level of doubt with regard to *which* mapping is intended. This response, in turn, is intended to reconfigure the information that the audience associates with the Tutsi people, in order to better map onto the set of information associated with cockroaches; namely, as dirty and disease-ridden pests that ought to be killed. A further chain of responses are also intended: as a result of reconfiguring their information about Tutsis in this way, the audience will be more willing and motivated to perform acts of violence toward them. The speaker's intentions with respect to these further goals are likely to be completely covert, because—as in the case of covert dogwhistles—their recognition would undermine the speaker's goals. As Stanley (2015) points out, there are often norms against explicitly derogatory speech. He suggests that in masking her speech as a public health warning, the propagandist is able to present herself as participating in reasonable discourse rather than in dehumanization and incitement of violence. Nonetheless, her speech will have the effect of the latter, through subtle inferences made at the subconscious level, rather than by explicit persuasion.

7.2.4 Advertising

A mixture of overt and covert communicative intentions are also involved in advertising. Consider, for example the following quote used in an advertisement for Pond's "Age Miracle" skin care products:

PONDS: "A thing of beauty is a joy for ever: its loveliness increases; it will never pass into nothingness." – John Keates[4]

This advertisement involves an act of direct meaning; the advertisers intend to draw the audience's attention to the literal content of the Keates quote as a direct result of the utterance, without this response being mediated by attention to other contents meant by the ad. Moreover, we can reasonably suppose that this act is overt in that the audience is intended to recognize this particular communicative goal. However, this is not the only goal that the text is meant to serve. The advertiser also intends that the audience's attention to this quote selects for a range of responses. The advertiser is exploiting the fact that women have internalized the message that their worth lies in their perceived sexual attractiveness, which decreases with age. When a woman ceases to be perceived as a "thing of beauty", she will lose her worth, and "pass into nothingness". In drawing the audience's attention to this quote, the ad-makers are making this message more salient, with the goal of causing emotional responses such as shame, unworthiness, and fear—which, in turn, are intended to affect her purchasing behavior. These further goals, however, are covert in the sense that the audience is intended not to recognize them; such recognition would likely undermine their aims, bringing to salience the sexist and manipulative nature of the ad, which may introduce resistance to the product. Advertising is similar to dogwhistling and propaganda, then, in involving at least some goals that are not cooperative in the minimal sense of revealing to the audience the nature of the speaker's intentions toward her.

7.2.5 Unintended speech

One might worry that the theory of meaning that I've proposed ignores an important linguistic category; namely, unintended speech acts. Saul (2018), e.g., notes that various effects of dogwhistles may be unintended. Stanley (2015) notes that characteristic effects of propaganda, too, may be unintended. Other examples of unintended effects of speech are documented by Alim and Smitherman (2012). They note that certain utterances tend to reinforce racist ideologies even when this is not intended. For example, during Obama's presidency, white commentators and politicians frequently remarked that Obama was articulate. While intended as a compliment, such remarks served both to reveal and to reinforce pernicious racial stereotypes.

People only remark on things they find remarkable. Pointing out that someone is articulate, then, suggests that this is somehow outside the norm, that the recipient of this "compliment" belongs to a group in which this trait would be surprising. We tend to point out that very small children are articulate, e.g., because this is a surprising feature of children in that age group. But we don't tend to point this out about adults (at least not white adults) because they are assumed to be competent speakers of their language. That Obama, a Black man, was consistently called "articulate", while the same remark was seldomly made of his white counterparts, reveals an underlying assumption that this trait is rare or surprising among Black people.[5] Similar observations may be made of the frequent "compliment" that so-and-so is a "strong, independent woman". Such descriptions are seldomly applied to adult men; when they are, we tend to assume that the man falls into a category for which these traits would be remarkable. Perhaps, e.g., he has overcome a debilitating injury, or a life of poverty and addiction. In contrast, the description is often applied to women simply because they are unmarried.[6]

The theory of meaning and speech acts that I've provided does not account for such speech, as its scope is limited to intentional communicative goals. However, I agree with Beaver and Stanley (2019) that it is important for a theory of language to be able to account for unintentional linguistic phenomena. I'm optimistic that this can be done by augmenting, rather than restructuring, the theory I've been outlining here. The account, for instance, would need to be supplemented with a theory of presupposition, which is commonly characterized as the information that speakers take for granted in performing their utterances. Presuppositions could then explain the racist ideology revealed and reinforced by the patterns of calling Obama "articulate". Though often no racist content is intentionally meant by such utterances, it is reasonable to suppose that the speaker is taking racist ideas for granted in some way. As a result, they serve to draw attention to, and reinforce racist ideology, regardless of whether this effect was intended.[7] Other kinds of cases might be explained by Khoo's (2017) theory of coded speech, according to which the mechanism that produces the response is not routed *via* reasoning about the speaker (neither about her goals, nor what she is taking for granted), but rather by automatic inferences that operate on stereotypical beliefs. Because this mechanism may be triggered when unintended by the speaker, the explanation may apply to certain cases of unintentional speech acts.

Such theories are compatible with the theory of meaning and speech acts that I've proposed here. While I hope that my proposal provides a helpful starting point for thinking about foundational issues in language, it will need to be supplemented with accounts of various kinds of unintentional speech in order to cover the full range of linguistic phenomena that we'd want a theory of language to explain.

7.2.6 Speech Acts with no Audience

My theory of meaning is limited not only to intentional actions, but also to audience-directed actions. In Chapter 5, I provided a recipe for defining more complex illocutionary actions using this basic notion of meaning—but this recipe, too, relies on the presence of intentional, audience-directed intentions. This raises the question of how to account for speech acts such as reciting poetry to oneself in the privacy of one's own room. Such actions do not fall within the scope of the picture I've been outlining, though some readers may have the intuition that they nonetheless qualify as speech acts.[8] We can call this category of speech acts *non-communicative speech acts*; they involve no communication, as there is no intended audience to communicate with.

I suggest that non-communicative speech acts—at least, when performed using a public language—are derivative of intentional, communicative speech acts.[9] We can't use language to perform such acts without already having a language at our disposal, and I've argued that intentional, overt, audience-directed acts are necessary for getting a public language up and running. We could never develop the kinds of expectations needed to sustain a linguistic convention in a particular language if people were only using that language to recite poetry in the privacy of their rooms. Once a public language is up and running, however, its users will be able to exploit it in non-communicative ways. In other words, communicative speech acts are the metasemantic base, and the primary type of speech acts needed to account for the bulk of our linguistic practices. Once our theory of communicative speech acts is in place, we can define a derivative category of non-communicative speech acts. I leave this project to others—however, I do think that such a project is both compatible with, and dependent on, a foundational story that focuses on communicative speech acts, such as the one I've offered.

While I claim that non-communicative speech acts performed in a public language are derivative of communicative acts, I take it that purely expressive acts may have preceded audience-directed acts in the evolution of public language. The same may be true of *unintentional* audience-directed acts. For this reason, the foundational account of human public language outlined here will also need to be supplemented by (or worked into) a more general theory of communication and signaling systems.

7.2.7 Too Broad?

In concluding this section, I want to address the question of what is lost when we expand our notion of meaning to include covert acts. Grice's apparent motivation for including a restriction to overtness, recall, was to rule out the following kind of case as an instance of meaning:

> I might leave B's handkerchief near the scene of a murder in order to induce the detective to believe that B was the murderer; but we should not want to say that the handkerchief (or my leaving it there) meant$_{NN}$ anything or that I had meant$_{NN}$[10] by leaving it that B was the murderer, Clearly we must add that, for x to have meant$_{NN}$ anything, not merely must it have been "uttered" with the intention of inducing a certain belief but also the utterer must have intended an "audience" to recognise the intention behind the utterance.[11]

I argued in Chapter 3 that restricting the account of meaning to include only overt communicative intentions is the wrong fix; while it may deliver the right result in the handkerchief case, it does so at the cost of getting the wrong results in cases of insinuation, propaganda, dogwhistles, and advertising. This is not a trade I'm willing to make, as I consider the latter phenomena to be considerably more important to a theory of linguistic communication than the handkerchief case.

Perhaps, however, there is a way to accommodate them both. My theory of meaning includes the clause that the audience's response is intended to occur at least partially on the basis of the utterance. Here again is the definition, for reference:

MEANING: In uttering e, a speaker S means m just in case for some audience A, S intends to direct A's attention to m at least partially on the basis of her utterance of e.

This can be contrasted with Grice's account, according to which the audience's response is intended to occur at least partially on the basis of their recognition of the speaker's intentions. Changing the intended basis of the audience's response from the recognition of the speaker's intentions to the utterance itself allows for covert speech acts, while ruling out the handkerchief case. In order to see why, consider that utterances are *actions*, not the physical results of those actions.[12] So the candidate for being the speaker's utterance in the handkerchief case is their act of *placing* the handkerchief, not the handkerchief itself (or its existence at the scene of the crime). This action fails to qualify as an act of meaning on my account because the speaker doesn't intend for the audience's response to occur on the basis of the former, but rather on the basis of the latter. If the ploy works, it is not the act of *placing* the handkerchief, but rather the existence of the handkerchief at a certain location that is doing the explanatory work; it would have had the same effect if it showed up at the crime scene not as a result of the speaker's action but by some other means.

Perhaps creative types will be able to come up with examples in which the response is intended to be caused by the action itself, but which is still not

intuitively a case of meaning. However, I tend to find little value in obscure counterexamples unless they are accompanied by an alternative proposal which is able to accommodate them, and does at least as well as the original theory in other respects. Furthermore, whether counterexamples present cause for concern depends on the aim of the theory in question. The project that I'm engaged in here is not one of ordinary language philosophy; I'm not trying to articulate a definition of meaning that conforms precisely to intuitions about the ordinary usage of this term (indeed, I doubt that such a definition could be provided, given that intuitions may conflict). Rather I'm trying to articulate a definition of meaning that can be put to theoretical work in the foundations of language by serving as a basic building block in a metasemantic and a pragmatic theory. I think that the proposal I've offered is up to this task.[13]

7.3 Interpretation

In demonstrating the limited and contingent role that cooperation plays on my account, I've been focusing up to this point on the speaker's side of a communicative exchange; now we'll turn our attention to its role on the side of the audience. Before saying more about how I'm thinking of the process of interpretation, it will be helpful to get a little clearer about what is meant by this term, which is sometimes used differently by different theorists. Very roughly, interpretation refers to the audience-side of a communicative exchange. This doesn't tell us very much, however. An audience may do many things during such an exchange. She may, e.g., laugh, clear her throat, form beliefs, etc. Which, if any, of these, qualifies as interpretation? Communicative exchanges are asymmetrical in the sense that the audience responds to the speaker (though, of course, participants usually take turns in these roles). This suggests that if we want to identify the audience's side of an exchange we might first look to the speaker's side. In most accounts of linguistic communication—including both Grice's and my own—the speaker's side involves the intention to produce a response of some sort in the audience. This suggests two different options for thinking about interpretation: 1) we can identify the audience's interpretation with her beliefs about the speaker's communicative intentions, or 2) we can identify the audience's interpretation with her production of the intended response (or some other reaction where the speaker is not successful in producing it). I'll stipulatively reserve the term *interpretation* for the former category, and use the term *response* for the second, but I think it's important to say a bit about both.[14]

We saw in Chapter 3 that on Grice's account, interpretation is tightly connected to cooperation; the audience relies on the assumption that the speaker is being cooperative in order to identify her communicative intentions. Roughly, if what the speaker says conforms to the Cooperative Principle, the audience will interpret her as meaning *that*. If it doesn't

conform, she will interpret the speaker as having performed an implicature which *does*.[15] Grice also seems to envision a tight connection between cooperation and response. He claims that—at least in successful cases—the audience's interpretation of the speaker provides her with a *reason* to perform the response. For instance, in the case of assertion, the recognition of the speaker's intention to elicit a certain belief will provide the audience with a reason to adopt that belief. Though this reason is purely epistemic—rather than say, a practical reason for performing an action in service of a joint goal—its force relies on background assumptions about the trustworthiness (and so cooperation) of the speaker.

On my proposal, neither interpretation nor response bear an especially tight connection to cooperation. Interpretation, I suggest, involves reasoning about the speaker's goals. Such reasoning may in some cases, but not in others, involve the assumption that the speaker is being cooperative. The audience's response may or may not involve reasoning about the speaker's goals; when it does, this may or may not involve the assumption that the speaker is being cooperative. We'll consider some examples momentarily, but first a reminder of how I am conceiving of the basic structure of speech acts, which involves more complexity than Grice's account. On that picture, speech acts are simple in that they involve the intention to produce a single response, and they are always and completely overt.[16] In contrast, my proposal treats speech acts as being complex in the sense that they may involve a chain of intentions, with the fulfillment of one serving as a means to accomplishing another. They are also complex in that they may contain other speech acts as proper parts. Moreover, speech acts may be overt, covert, or somewhere in between. There are two ways in which a speech act can be partially overt. First, it may involve both covert and overt goals. Second, it may involve goals which are only covert to a degree, in the sense that the speaker wants the audience to recognize what she is up to, but does not want this recognition to rise to the level of knowledge.

Let's consider a few examples which will illustrate the contingent and variable role of cooperation in interpretation and response. First, we'll consider an overt direct assertion of "She's arrived". On my account, direct assertions involve a chain of at least two goals. One intention is that the audience attend to the literal content of *She's arrived*—this part of the chain is the speaker's act of meaning. Another intention is that the audience believe that the intended referent arrived. The entire chain constitutes her assertion—so the act of meaning forms a part of the assertion. Because this example of assertion is completely overt, the speaker intends the audience to recognize the intentions it comprises. Interpretation, I've said, is a matter of reasoning about the speaker's goals. If the audience recognizes the utterance to be overt, some minimal presumption of cooperation may figure in her interpretation; she may assume that the speaker is cooperating

at least to the extent of revealing her communicative goals, and will use this information to identify them. This may lead her to, for instance, look for the most salient referent of "she", on the assumption that the speaker will exploit mutually apparent features of the context in order to reveal her intentions. And she may reason that, absent evidence to the contrary, the utterance was an assertion (because the speaker would have provided clues (tone of voice, etc.) if she had instead intended to make a joke). On the other hand, overt speech acts may not necessarily involve cooperation; the speaker may be trying to reveal their communicative intentions to an audience that has no interest in recognizing them (harassment of any kind tends to fall into this category).[17] In such cases, the audience's recognition of the utterance as overt may not carry with it the presumption that the speaker is being cooperative, even to a minimal extent.

What about response? Because there are at least two goals involved in a direct assertion, there are at least two responses to consider: attending to the content of "She's arrived", and believing it. Both of these may, but need not, involve cooperation on the part of the audience, or their assumption that the speaker is cooperating. If the audience has correctly concluded that the speaker has asserted that she's arrived, then they are already to some extent attending to the content that she's arrived, given that this content is part of their belief about the speaker's goals. Therefore, the presumption of cooperation on the part of the speaker is not doing any extra work in producing the response, over and above the work it did in producing her interpretation. We might wonder, however, whether there is cooperation on the part of the audience. Again, this may—but need not—be the case. The way we direct our attention is under our control only up to a point, and many times we find ourselves attending to the content of speech regardless of whether we want to (e.g., being subject to catcalling on the street).[18] In some cases, however, the extent to which we attend to the speech directed at us is a matter of cooperation; "good listeners", e.g., make more of a concerted effort to cooperate in this matter than do "bad listeners".

The second response to consider is belief. Again, the assumption of cooperation on the part of the speaker may or may not play a significant role in whether the audience produces this response. Suppose that the audience assumes that the speaker is cooperating in an endeavor to exchange information. This may lead the audience to believe that the speaker believes that what she asserts is true, but this on its own is not enough to produce the intended response; the audience needs to also believe that the speaker is well-informed, and that her beliefs are likely to be true. If she does not believe this, then she will not have the intended response. What about when the audience assumes that the speaker is being *uncooperative*—does this lead to a failure of the intended response? Sometimes. This assumption may lead the audience to conclude that the speaker is lying, which—under

the assumption that the speaker's beliefs are reliable—will provide a reason for the audience to fail to have the intended response. However, if the audience considers the speaker's beliefs to be unreliable, they may have the intended response *in spite of* believing the speaker to be uncooperative.[19] So the success of the intended response neither stands nor falls with the presumption of cooperation.

What about cooperation on the part of the audience? I suspect that in most cases of assertion, where the intended response is belief, audiences produce this response for epistemic reasons rather than for the sake of being cooperative. If cooperation plays a role in the audience's response, it does so contingently, and its tendency to do so will vary according to the type of speech act they are responding to. One is usually cooperating when answering a question, for instance, though this need not be the case. In fulfilling requests, similarly, one may be cooperating, or one might be acting in one's own interests because of, e.g., fear of retaliation.

Now let's consider a case of covert meaning. Most cases of covert meaning involve at least some level of overtness in that they are often indirect speech acts. Consider the dogwhistle "inner city". In uttering this, the speaker has at least two goals. First, she wants the audience to attend to the content *inner city*. This is her act of direct meaning; she wants the audience to attend to this content on the basis of her utterance alone, rather than on the basis of something else that she means. Her second goal is as follows: on the basis of attending to the content *inner city*, she wants her audience to attend to various pernicious racial stereotypes. This is her act of indirect meaning. Her first goal will likely be overt, in the sense that she wants her audience to recognize her intention to get her to attend to the content *inner city*. Her second goal is covert in the sense that she doesn't want the audience to recognize her intention to get her to attend to racial stereotypes. We've already considered the role of cooperation in interpreting and responding to overt direct meaning above. The same general ideas apply here, so I'll not go over them again. Instead, let's focus on how the audience interprets the speaker's act of covert indirect meaning. If the speaker is successful, there will be no act of interpretation at all—the audience will not assign her any communicative intentions, and so the presumption that the speaker is being cooperative will play no role in identifying them. If the speaker is unsuccessful, however, the audience will recognize her intentions, and the assumption that the speaker is being uncooperative may either lead the audience to this realization, or be a byproduct of it. In neither case is the presumption of cooperation playing a role in the audience's process of interpretation. What about her response? Recall that when dogwhistles are successful, the desired response occurs as an automatic inference on the part of the audience rather than as a result of reasoning about the speaker's goals. Here again, then, the presumption that the speaker is being cooperative plays no role.[20]

Performing a speech act is a goal-directed action which involves trying to elicit a response in the audience. The goals of the speaker and audience need not be shared, or even presumed to be shared. Thus, in order to successfully interpret the speaker, the audience needs to represent her means-end reasoning. In order to do this, she will draw on background beliefs about both the speaker and the context. If she assumes that the speaker is being cooperative, then that fact will figure into her reasoning about the speaker's communicative goals. If it does not, it will not. My project here is to provide a foundational theory of language, and a so more detailed representation of the mechanics of conversation would take me beyond its scope. However, my proposal is a good fit with a game-theoretic approach that models conversations as (not necessarily cooperative) strategy games, where each player's move requires the application of general principles of rationality to order to understand others' goals, and tailor one's own moves accordingly.[21] Game-theoretic models of conversation, however, will have their limits, as they don't represent automatic inferences—such as those generated by dogwhistles—that are not generated by reasoning about the speaker's intentions. Thus, a more complete representation of conversational dynamics will ultimately need to adopt a cross-disciplinary approach, taking cues not only from game theory, but also from psychology and cognitive science.

7.4 Conclusion

While the theory of meaning and speech acts on offer should be viewed as an adaptation of Grice's account, it should not be viewed as the second step in the process of Galilean idealization, where the idealization concerns the conception of language as a cooperative enterprise. This is because it does not build on the tools that were developed from this conception— such as the characterization of meaning as essentially overt, or the characterization of interpretation as essentially involving the presumption of cooperation. Rather, it develops Grice's theory by exploring what we are left with when this idealization is completely stripped away. This, I've suggested, is an understanding of speech acts as complex actions that may involve multiple goals, each of which may fall anywhere along the scale between overt and covert. We are also left with an understanding of interpretation that preserves the fundamental insight that this process involves reasoning about the speaker's goals, and that this process is governed by general—rather than specifically linguistic—norms of reasoning. However, we lose the commitment that these goals must be—or be presumed to be— cooperative. The result, I hope, is a theory of meaning and speech acts that provides a broader range of fine-grained tools for understanding our varied and complex linguistic practices.

Notes

 1 Camp (2018, p. 43).
 2 See D'Ambrosio (n.d.) for a discussion of manipulative speech. He points out that such speech is deceptive in a second sense: not only is the speaker failing to cooperate in this regard, but she is presenting herself as though she is cooperating. So not only is she deceiving her audience with respect to her communicative intentions, but also with respect to her intention to cooperate. In this sense, manipulative speech is different from openly hostile communication, where it is common knowledge that conversational participants are mutually uncooperative.
 3 I am simplifying here by supposing that the response will be a result of attending to the literal content of the entire utterance; rather, it might be a function of attending to the content of its parts.
 4 Ponds Age Miracle, 2013. https://www.slideshare.net/BhallaAnurag/ponds-age-miracle-advertising-strategy
 5 In particular, it reveals a bias against Black American dialects. White ideology maintains that certain variations of English (belonging to privileged White people) are "standard" and naturally superior in various ways.
 6 For discussion, see Holroyd (2021).
 7 On this explanation, such speech would qualify as what Stanley (2015) calls *undermining propaganda*; it purports to embody a nonracist ideology by complimenting a Black person, while reinforcing that ideology with the presupposition that being articulate would qualify as a notable achievement for a member of that group.
 8 See, e.g, Reiland (n.d.).
 9 I take it, however, that purely expressive acts may have preceded audience-directed acts in the evolution of public language.
10 The subscript NN signifies non-natural. See section 3.4 for discussion.
11 Grice (1989, p. 217).
12 I say more about the metaphysics of utterances in Keiser (n.d.).
13 Moreover, in line with the process of Galilean idealization, the spirit of the proposal can be preserved by making minor adjustments in light of new theoretical interests.
14 Most people, following Austin (1962), distinguish between responses that qualify as illocutionary uptake in the sense that they are directly related to the speech act in some important way vs. perlocutionary effects, which are responses that occur further down the line. It's a controversial matter about where to draw this line, and since it doesn't matter for my purposes here, I gloss over it. However, we will return to this issue later.
15 Simons (2006) also applies Grice's approach to presupposition interpretation, arguing that the assumption of cooperation is needed to explain the audience's interpretation of presuppositions.
16 This is only a rough characterization; if we count higher order intentions, each speech act has infinitely many goals.
17 Thanks to Thomas Brouwer for this point.
18 For a discussion of controlled vs. automatic attention-direction, see Wu (2014, pp. 29–38).
19 Grice's countersuggestible man is a related case. See Grice (1989, p. 107).
20 Stanley (2015) Chapter 3 suggests that this is not a rational response at all. I'm not sure that I agree with this—though it is not a response that involves reasoning about the speaker's goals, or even reasoning explicitly at all, it may nonetheless be rational on a reasons-based conception of rationality. That is, the speaker will have (however unjustified) beliefs that lead her to draw certain connections. Again, however, I'll set aside the issue of whether this should be considered as falling under the scope of "rational communication".

21 Though we disagree about other details about conversational dynamics, examples of this general approach include Asher and Lascarides (2013), Asher et al. (2017), Asher and Paul (2018), and D'Ambrosio (n.d.).

References

Alim, H. Samy, and Smitherman, Geneva (2012). *Articulate While Black*. Oxford: Oxford University Press.

Austin, John Langshaw (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

Asher, Nicholas, and Lascarides, Alex (2013). Strategic conversation. *Semantics and Pragmatics*, 6(2):1–62.

Asher, Nicholas, Paul, Soumya, and Venant, Antoine (2017). Messagee exchange games in strategic contexts. *Journal of Philosophical Logic*, 46(4):355–404.

Asher, Nicholas, and Paul, Soumya (2018). Strategic conversations under imperfect information: Epistemic message exchange games. *Journal of Logic, Language, and Information*, 27:343–385.

Beaver, David, and Stanley, Jason (2019). Toward a non-ideal philosophy of language. *Graduate Faculty Philosophy Journal*, 39(2):501–545.

Camp, Elisabeth (2018). Insinuation, common ground, and the conversational record. In Daniel Fogal, Daniel W. Harris, and Matt Moss (eds.), *New Work on Speech Acts* (pp. 40–66). Oxford: Oxford University Press.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

D'Ambrosio, J. (n.d.). A theory of manipulative speech. (unpublished ms).

Holroyd, Jules (2021). Oppressive praise. *Feminist Philosophy Quarterly*, 7(4):2371–2570.

Keiser, Jessica (forthcoming). Languages and language use. *Philosophy and Phenomenological Research*.

Khoo, Justin (2017). Code words in political discourse. *Philosophical Topics*, 45(2):33–64.

Ponds Age Miracle, (2013). https://www.slideshare.net/BhallaAnurag/ponds-age-miracle-advertising-strategy

Reiland, I. (n.d.). Meaning, conventions, and rules. (unpublished ms).

Saul, Jennifer (2018). Dogwhistles, political manipulation, and philosophy of language. In Daniel Fogal, Daniel W. Harris, and Matt Moss (eds.), *New Work on Speech Acts* (pp. 360–383). Oxford: Oxford University Press.

Simons, M. (2006). Presupposition without common ground. https://www.cmu.edu/dietrich/philosophy/docs/simons/Presupposition%20without%20Common%20Ground.pdf

Stanley, Jason (2015). *How Propaganda Works*. Princeton, NJ: Princeton University Press.

Tirrell, Lynne (2012). Genocidal language games. In Ishani Maitra and Mary Kate McGowan (eds.), *Speech and Harm: Controversies over Free Speech* (pp. 174–221). Oxford: Oxford University Press.

Wu, Wayne (2014). *Attention*. London: Routledge.

8

NON-IDEAL DISCOURSE STRUCTURE

8.1 Overview

In the previous chapters, I presented a foundational theory of language that is non-ideal in the sense that it rejects the idealized conception of language as cooperative information exchange. In this chapter I don't *add* to this theory of linguistic communication, *per se*; rather, I develop an informal model that is able to represent the structure and process of communication that it describes. My hope, however, is that this model will not only be representational, but that it will also be generative—providing tools that lead to further developments in our understanding of linguistic communication.

Before presenting this model, I want to say a bit about how it both conforms to, and departs from, the classical Stalnakerian framework. Like Stalnaker, I take the goals of linguistic communication to furnish the starting point for developing a model of discourse structure. However, while Stalnaker assumes that the "central" or "essential" case of linguistic communication is cooperative information exchange, I begin with the assumption that linguistic exchange is characterized by a variety of goals, among which information exchange fails to hold a privileged place. Though I do away with those features of Stalnaker's framework that are grounded in the conception of language as information exchange, I preserve and build upon others. One way in which my model hews closely to this framework is in representing discourse goals using information sets, and characterizing these sets using mental states. These sets will act as the basic building blocks of the framework, representing the effects of speech acts as well as constraints on utterance production and interpretation.

More concretely, my proposal can be contrasted with the traditional Stalnakerian framework by the ways in which it moves away from what I have called the *Five Stalnakerian Idealizations*. Recall from Chapter 4 that the Stalnakerian model exhibits the following five problematic features, which are grounded in the assumption that language is essentially a tool for cooperative information exchange:

1. Conversations are taken to have a *unique* goal.
2. This goal is taken to be *shared*.
3. All speech acts are taken to be aimed at updating, and as being interpreted against a *unique* information set.
4. This information set is characterized in terms of *shared* mental states.
5. This information set is characterized using a *unique* and *coarse-grained* type of mental state.

In contrast, the model I propose below will represent each conversation—even each utterance—as potentially aimed at a variety of goals, which need not be shared. These goals are aimed at updating a variety of information sets, which need not be shared. Finally, these information sets are individuated by a variety of fine-grained mental states. In the following section I present a broad overview of the basic structure of this framework. The remainder of the chapter will explore its details as we look at some of the ways it is able to overcome problems faced by the Stalnakerian model.

Though I have been critiquing certain kinds of idealizations throughout this book, I want to reiterate that my claim is not that all manner of idealization should be avoided in theory building. Various epistemological and practical constraints necessitate that we simplify and even ignore aspects of the phenomenon we are trying to understand. It is not idealization in general that should be avoided, but rather pernicious idealization—the kind that distorts, rather than enhances our understanding of the target phenomenon. My claim is that idealizations grounded in the conception of language as cooperative information exchange tend to have this distorting effect, and should be removed from a theory of language—not that such a theory should be devoid of idealization altogether. So, while I advertise the framework outlined below as a "Non-Ideal Discourse Structure", this means that it is intended to be free from *pernicious* idealizations related to the function of language, rather than from any form of idealization. With that being said, let me outline ways in which the proposed framework will involve various kinds of simplifications and idealizations.

First, I present a zoomed-in, partial picture of a broader framework. The broader framework includes what we might loosely describe as two levels: the individual and the social. The framework will represent each individual using a set of sets of mental states, individuated by type. On the social level,

the framework will represent these individuals as situated within a broader network of social structures. This second part of the framework—the social level—is developed in joint work with Katherine Ritchie.[1] Because we articulate and defend this part of the framework elsewhere, my focus here will be to outline the details of the individual level.

Second, I use mental states to represent the relevant individual-level phenomena. Likely, there are other aspects of the individual, such as behaviors or physical states, that also play a role in the communicative process. In that case, the framework would need to be supplemented to represent them. Thus, the proposal involves Galilean idealization in the sense that its representational scope is purposefully restricted in order to render it theoretically tractable in its early stages. My hope is that further complexities can be systematically introduced without involving significant restructuring of the existing framework. If this hope is borne out, then while the initial restriction to mental states is an idealization, it is not a pernicious one.

Third, I characterize speakers' goals as aiming to update audiences' mental states with determinate contents, though this may not always be the case.[2] As I mentioned in Chapter 5, I anticipate that my proposal can be adapted to accommodate whatever turns out to be the best solution to the problem of indeterminacy of meaning, which is not a special problem for my account.

Finally, I simplify by modeling indirect speech as aiming to update an attitude. This way of modeling indirect speech ignores a potential intermediary goal of getting the audience to create a mapping between the source domain and the target domain. In order to accommodate this, the framework would need to represent not only the contents of attitudes, but the structured mappings between them.

8.2 The framework

According to the picture of language use that I presented in Chapter 5, speech acts involve a chain of goals, each in service to the next. I maintained that the goal occupying the last position in the chain—i.e., the ultimate goal of the speech act, insofar as one can be identified—can vary dramatically between speech acts. One can use language to impart information, to tease, to perform religious rituals, to socially bond, to socially exclude, to reinforce political ideologies, to manipulate purchasing behavior, etc. What such speech acts have in common, I suggested, involves the goal characterized by (roughly) the first link of the chain: the speaker aims to direct the attention of her audience toward a certain content.[3] The chain of goals that characterizes any given speech act will usually extend beyond attention-direction; a speaker rarely aims to draw her audience's attention to a content for its own sake, but rather because she wants the audience to select that content for some other type of action. For instance, in making an

assertion she may want to direct the attention of the audience to a content in order that the audience form a belief in that content. Alternatively, she may be making a joke—in which case she may want to direct the attention of the audience toward a content in order that the audience be entertained by that content. This reaction, in turn, may be in service to yet another goal; the chains of goals that characterize various speech acts may be arbitrarily long and complex.

Following Stalnaker, I represent these goals as aimed toward updating information sets that are characterized by mental states. Because the first goal in the chain is to direct the audience's attention toward a content, it will be represented by a set of contents at which the audience's attention is directed—we can all this the audience's *attention set* (abbreviated as A in Figure 8.1). We'll assume for now that the audience is an individual, but later we will look at cases in which the audience is better represented as a set of individuals. When a speaker means m, she performs an utterance with the immediate aim of adding m to the audience's attention. (For the moment, we'll ignore *overt* and *covert* communicative intentions.) See the diagram of the speaker's goals, with the arrow representing a means-end relation (Figure 8.1).

The speaker will usually be directing the audience's attention in service of a further goal. As we discussed above, if this goal does not involve a change in mental states, it will fall outside the scope of what the model is able to represent with the tools on hand. However, many speech acts do aim to change mental states, and so can be represented by the model with an information set corresponding to the relevant type of mental state. Consider, for example, a simplified characterization of assertion. When performing an assertion, the speaker aims to draw her audience's attention to a certain content in order that the audience select it as the object of belief. (I make this simplifying assumption throughout—if you don't think that assertions are always aimed at updating belief, think instead of the subset of assertions that are.) In this case the second goal will be represented by the set of contents believed by the audience: the audience's *belief set* (abbreviated as B in Figure 8.2). When a speaker asserts m, she performs an utterance with the immediate goal of adding the content m to A. This serves as a means of achieving her second goal of adding m to B (see Figure 8.2).

Utterance → (*m*)

A
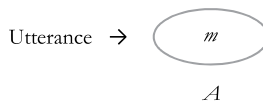
FIGURE 8.1 Diagram of meaning
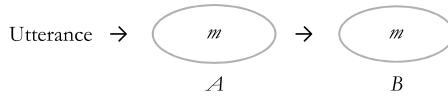
FIGURE 8.2 Diagram of assertion

FIGURE 8.3 Diagram of revised assertion

Now we can see a chain of goals beginning to form, beginning with the goal of directing the audience's attention. Note that this example involves a simplified account of assertion. More nuance is likely in order, as the fact that an audience *attends* to *m* is not, on its own, likely to result in her *believing m*. For this reason, many accounts of assertion involve more complexity in their characterization of the speaker's goals. Perhaps, for instance, the speaker intends the audience to recognize that the speaker herself believes *m*, or that she is putting *m* forward as true, or that she is undertaking a commitment to *m*, etc. Together with additional background beliefs about the trustworthiness of the speaker, such facts would provide the audience with a reason to believe *m*. Each of these options for characterizing the act of assertion can be represented using the theoretical tools of the framework, as each of them appeals to mental states. For illustration, consider the first option, in which the speaker intends the audience to recognize that she herself believes *m*.[4] This could be represented as per Figure 8.3, with the proposition that the speaker believes *m* is abbreviated as *sBm*.

This chain of goals is longer than the simplified version considered above; we have added a middle link which represents the speaker's goal of adding *sBm* to the audience's belief set. This goal, in turn, is in service to the further goal of adding *m* to that same set. You might think of this middle goal as representing the speaker's intention that the audience recognize the *force* of her utterance—i.e., that it qualifies as an *assertion*, rather than some other type of speech act. Relative to the alternative accounts mentioned above, the middle goal would involve a different update to the audience's belief set—i.e., adding the proposition that the speaker is putting *m* forward as true, or that she is undertaking a commitment to *m*, etc. My purpose here is not to defend any particular account of assertion, but rather to how show various proposals might be modeled in the framework. As long as they can be understood in terms of updates to mental states, they can be represented using its basic tools.

Though a viable account of assertion will likely involve a more nuanced chain of goals, for the sake of clarity I use the simplified account in the discussions below. One issue that the more nuanced alternatives bring out, however, is that conversational goals cannot be characterized simply by appeal to the information set that they add to; they must also be characterized relative to their position in the chain of goals. In the example above, for instance, the speaker has the goal of adding two contents to the audience's belief set: m, and the proposition that the speaker believes m. But each of these goals play a different role in the communicative process: adding m to the audience's belief set is what we might think of as the ultimate goal of the speech act, while adding the proposition that the speaker believes m is a means to that end. Thus, in order to represent the communicative role that various conversational goals play within a single utterance, we must individuate them along at least two dimensions: the information set that they aim to update, as well as their position in the broader chain of goals.

Because some conversational goals are intended to be recognized by the audience while others are not, the model must also be able to represent the distinction between overt and covert communicative intentions. Let's start with a case of overt meaning. Here, again, is the definition I presented in Chapter 5:

OVERT MEANING: In uttering e, S overtly means m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e.
2. A to recognize that S intends (1).

My formulation of overt meaning doesn't necessitate that the speaker's intentions stand in any particular means-end relationship. On the Gricean understanding, however, the audience's response is intended to occur at least partially on the basis of her recognition of the speaker's intention to elicit that response. For concreteness, let's consider that type of case. Here, overt meaning involves a chain of conversational goals which can be represented as shown in Figure 8.4, where $sI(aAm)$ abbreviates the proposition that the speaker intends the audience to attend to m.

Utterance $\rightarrow$ $sI(aAm)$ $\rightarrow$ m

B $\qquad\qquad A$
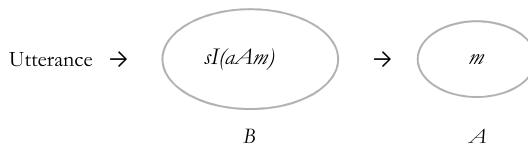
FIGURE 8.4 Diagram of overt meaning

The first goal of the speaker is to get the audience to recognize that the speaker intends the audience to attend to *m*: that is, for the audience to add *sI(aAm)* to her belief set. This goal, in turn, is in service of another goal— that the audience add *m* to her attention set.[5]

Covert communicative intentions, in contrast, will lack the middle link in this chain. The speaker, for instance, will intend that the audience attend to *m* without first recognizing that the speaker has this communicative intention. But we need a bit more to model covert communicative intentions. Here the speaker doesn't simply lack an overt communicative intention, as she does in the neutral case represented by Figure 8.1. Rather, she actively intends that the audience fail to recognize her intention—her goal is to *rule out* the update of *sI(aAm)* to the audience's belief set. We can represent this with Figure 8.5.

The line through the belief set represents the fact that the speaker intends that this update fail to occur. (Note that this can't be modeled using the negation of the proposition, because the speaker does not necessarily intend that the audience actively believe that the speaker does not intend her to attend to *m*. Likely, she does not want the idea to occur to her at all.) Here we see that in some cases we may want branching chains rather than purely linear chains. The belief set does not belong in the middle of the chain because it is not *on the basis of* failing to update her beliefs that the audience is intended to attend to *m*.[6] Rather, the speaker intends this to occur on the basis of her utterance. For the sake of simplicity, in the remainder of the chapter I default to the neutral case in explaining how we might model various types of speech. However, overt and covert versions of the same cases can be modeled by including the goal of updating the audience's belief set, or the goal that no such update occurs, respectively.

In the following sections we will explore the framework in more detail as we look at how it is able to overcome problems faced by the Stalnakerian model. In the final section I suggest ways the framework might be modified in order to address issues such as functional expressions and not-at issue content.
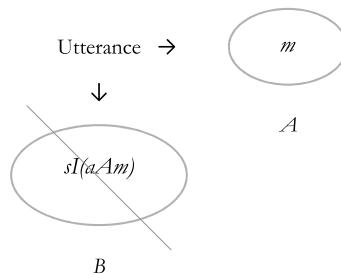
FIGURE 8.5 Diagram of covert meaning

8.3 Non-literal speech

In Chapter 4 I argued that the traditional Stalnakerian framework cannot adequately represent non-literal speech because it treats conversations as guided by the unique goal of updating a unique information set, which is characterized by a unique, coarse-grained mental state. As a result, it treats both the literal and non-literal content of indirect speech as contributing to the same information set—the set of information that is mutually accepted for the purpose of the conversation. This picture fails to distinguish between the different communicative roles of literal and non-literal content. In contrast, because the proposed framework incorporates multiple goals, information sets, and attitudes, it is able to represent the fact that in using non-literal speech, the literal content is the *mechanism* by which the speaker communicates the non-literal content. These resources also allow the framework to represent the fact that the communicative success of non-literal speech can come in degrees.

Before demonstrating how this works, it will be helpful to revisit the account of indirect speech that I offered in Chapter 5. My theory treats non-literal speech as a species of indirect meaning, which can be understood in terms of a hierarchy of conversational goals:

DIRECT MEANING: By uttering e, S directly means m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e.

   and there is no content k (distinct from m) such that S intends

2. A to attend to k at least partially on the basis of S' utterance of e.
3. (2) is a means to (1).

INDIRECT MEANING: By uttering e, S indirectly means m just in case for some audience A, S intends

1. A to attend to m at least partially on the basis of S' utterance of e.

   and there is some content k (distinct from m) such that S intends

2. A to attend to k at least partially on the basis of S' utterance of e.
3. (2) is a means to (1).

On this analysis, direct meaning and indirect meaning are distinguished by whether they are mediated by another act of meaning. In the case of

indirect meaning, the speaker directs her audience's attention to one content *by* directing her attention to yet another content. In contrast, in the case of direct meaning, the speaker directs her audience's attention to a content on the basis of her utterance alone (together with background information)—without directing her attention to another content as a means to this end.

Because the proposed model incorporates multiple goals, information sets, and mental states, it can more adequately represent the distinct communicative roles of literal and non-literal content. Each is associated with a different goal, which is characterized both by the information set it aims to update, and by its position in the chain of goals to which it belongs. We can see how this works by considering the case of indirect speech discussed in Chapter 4:

COFFEE: A speaker utters the sentence "you are the cream in my coffee" to her lover, with the intention of communicating that she is fond of her lover.

This utterance qualifies as an indirect speech act on the analysis proposed above because the speaker intends that:

1. The audience attend to the proposition that the speaker is fond of the audience at least partially on the basis of her utterance.
2. The audience attend to the proposition that the audience is the cream in the speaker's coffee at least partially on the basis of her utterance.
3. (2) is a means to (1).

It follows from our definition that the speaker indirectly means the proposition that she is fond of the audience. Though an utterance can involve more than one level of indirectness (think of a metaphor used sarcastically), for simplicity we can stipulate that this is not the case in *Coffee*; that is, we'll imagine that the speaker intends the audience to attend to the proposition that the audience is the cream in the speaker's coffee on the basis of her utterance alone, rather than on the basis of attending to another content that she meant by her utterance. It follows from our definition, then, that this proposition is directly meant. We can represent this speech act using Figure 8.6, where the proposition that the lover is the cream in the speaker's coffee is abbreviated as c and the proposition that the speaker is fond of her lover is abbreviated as f.
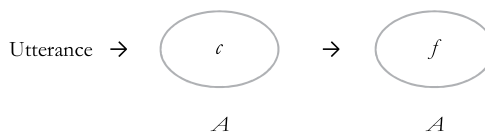
FIGURE 8.6 Diagram of indirect meaning

The distinction between the communicative function of the literal content can be distinguished from that of the non-literal content by their respective positions in the broader chain of goals that the speaker hopes to achieve with her utterance; the former is a mechanism by which the latter is communicated. In cases of multiple levels of indirectness, non-literal content can serve as a mechanism for communicating more non-literal content; however, literal content is special in that the speaker does not intend to draw the audience's attention to this content *by means of* drawing her attention to a different content. Thus, literal content occupies a privileged place in the chain in that it is the most immediate content that the speaker intends to add to the audience's attention set. When the speaker is using indirect speech, the next link in the chain will be occupied by the non-literal content, which represents the fact that the former is a mechanism by which the latter is communicated.

The success conditions of indirect speech acts come in degrees. In order to be fully successful, the audience must recognize both the literal and the non-literal content of the utterance. So, in the case of *Coffee*, she must attend to both the proposition that she is the cream in the speaker's coffee and that the speaker is fond of her. If she understands the literal content but misunderstands the metaphor—for instance, perhaps she believes that the speaker is lactose intolerant and takes her to be communicating that she (the audience) is making her sick—then the speech act is only partially successful. We can represent this in the framework with the addition of c—but not f—to the audience's attention set.

As I've mentioned, it is rarely the case that the speaker aims to get her audience to attend to a content just for its own sake; usually she has some further goal in mind. In such a case, the chain of goals representing the speaker's utterance will extend beyond that of adding the non-literal content to the audience's information set. For example, let's suppose that in the case of *Coffee*, the speaker is making an indirect assertion; that is, she not only wants the audience to *attend* to the proposition that she is fond of her lover, she wants her to *believe* it. We can represent the goals constituting her indirect assertion in Figure 8.7.

Here, again, the literal content occupies a privileged place in the chain in that it is the most immediate content that the speaker intends to add to the audience's attention set. It is followed by the non-literal content, because

FIGURE 8.7 Diagram of indirect assertion

the former is a mechanism by which the speaker intends the audience to attend to the latter. In this case, however, the speaker has a further goal; she wants the audience to attend to the proposition that she is fond of her lover in order that she believe it. Adding this proposition to the audience's belief set is the next link in the chain, following that of adding it to her attention set.

In this case, the speaker may succeed in communicating both the non-literal and the literal content of her utterance, but fail to achieve the ultimate purpose of her speech act, which is to get the audience to believe the non-literal content. That is, the audience may understand the metaphor but fail to believe it—assuming, perhaps, that the speaker is merely flattering her. In this case the partial success would be represented in the framework by the addition of c and f to the audience's attention set, but *no* addition of f to her belief set.

Because the proposed framework rejects the five Stalnakerian idealizations, it is better equipped to represent the distinct roles of literal and non-literal content in indirect speech. As we've seen, the proposed framework incorporates multiple goals, information sets, and attitudes. The various goals of the speaker can be distinguished both by the information set that they aim to update—which may be characterized by different types of attitudes—as well as their position in a chain of interlocking goals. In this way, we are able to represent the fact that the literal content serves as a *mechanism* by which the speaker communicates the non-literal content. In this section we've looked at how the proposed framework can represent various goals of the speaker within a single utterance, specifically with respect to indirect speech. In the next section we'll consider cases in which a speaker's goals can vary across a conversation by contrasting assertion with non-serious speech.

8.4 Non-serious speech

Conversational goals may vary not only within an utterance, as we saw above, but across a conversation. For instance, speakers are able to switch back and forth between assertion and non-serious speech within the course of a single conversation. In Chapter 4 we looked at ways in which this presented a problem for the Stalnakerian framework, which represents conversations as having a unique goal, and speech acts as aiming to update a unique, shared information set characterized by the coarse-grained attitude of acceptance. These idealizations robbed the framework of the tools needed for distinguishing between the different sorts of communicative goals served by assertion and non-serious speech. Assertions are standardly put forward as candidates for belief, while jokes are put forward as objects of amusement. In order to represent their different functions, a framework

needs to model each of these speech acts as updating a different information set, and to characterize these information sets using different mental states. In order to see how this can be done in the proposed framework, consider the following utterances:

WOLF JOKE: During her 2018 White House Correspondent's Dinner speech, Michelle Wolf joked that "Trump is so broke he looked for foreign oil in Don Jr.'s hair."

WOLF ASSERTION: At the end of her 2018 White House Correspondent's Dinner speech, Michelle Wolf asserted, "Flint still doesn't have clean water."

For the sake of simplicity, we can assume that the ultimate aim of the joke was to amuse (or perhaps infuriate) her audience with the content that Trump is so broke he looked for foreign oil in Don Jr.'s hair, which we'll abbreviate as t.[7] We can suppose that the ultimate aim of her assertion was to get her audience to believe the content that Flint still doesn't have clean water, which we can abbreviate as w. On the proposed framework, we can illustrate the different roles of assertion and jokes by appealing to the fact that they aim to update different information sets, which are characterized by different mental states. Again, we'll use B to indicate the audience's belief set, and E to indicate the set of contents that the audience is entertained by (see Figure 8.8).

Each of these speech acts involves a chain of goals. In both cases, the first link in the chain involves directing the audience's attention toward a content. However, each of these cases of attention-direction serves a different purpose. In the case of *Wolf joke*, directing the audience's attention toward t serves the goal of adding t to the set of contents that the audience is entertained by. In contrast, in the case of *Wolf assertion*, the speaker directs the audience's attention toward w in service of adding that content to the audience's belief set. In using a single information set characterized
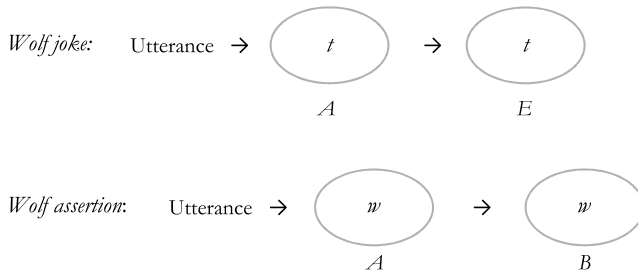
FIGURE 8.8 Diagram comparing joke and assertion

by the coarse-grained notion of acceptance, the Stalnakerian framework lacked the tools to represent the fact that assertions are standardly put forward as candidates for belief, while jokes are put forward as objects of amusement. The proposed framework is able to mark this distinction because it models each of these speech acts as ultimately aimed at different goals. These goals are distinguished within the framework using different information sets, associated with different attitudes. Rather than using a single information set characterized by acceptance for the purpose of the conversation—an attitude which is so broad that it encompasses the effects of jokes and assertions alike—we can distinguish between these different types of speech acts by appealing to finer-grained mental states such as belief and entertainment.

As before, the model is able to represent the various degrees of communicative success corresponding to these different speech acts. In the case of *Wolf assertion*, complete communicative failure will involve neither an update of *w* to the audience's attention set nor her belief set. But the assertion may be partially successful in the sense that the audience attends to the correct content, but fails to believe it. We would represent this as an update of *w* to the audience's attention set, with no corresponding update to her belief set. Degrees of success corresponding to *Wolf joke* would be modeled along the same lines; complete failure would involve no update at all, while partial success would involve an update to the audience's attention set but not the set of contents that she is entertained by.

Overt speech acts also introduce additional degrees of success. Recall that in such cases the speaker intends the audience to recognize her communicative goals. In the case of *Wolf assertion*, we can represent this with a middle link in the chain corresponding to the speaker's intention that the audience recognize her goal to add *w* to the audience's belief set (see Figure 8.9).

Because overt speech acts add another link in the chain of goals, they introduce yet another degree of success, corresponding to the success of getting the audience to recognize the force of one's utterance. If, for example, the audience recognizes the content of Wolf's assertion but misidentifies it as a joke, she has only been partially successful. This case would be represented in the framework by the update of the audience's attention set with *w*, and an update to her belief set with the proposition that the speaker

Utterance $\rightarrow$ w $\rightarrow$ $sI(aBw)$ $\rightarrow$ w

A B B
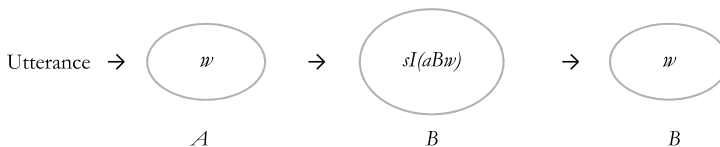
FIGURE 8.9 Diagram of overt assertion

intends the audience to be *entertained* by *w* rather than the proposition that the speaker intends the audience to *believe w*. The speaker will have been successful to a greater degree—but still only partially so—in the case that the audience recognizes her speech act as an assertion but fails to believe it. This case would be represented by the addition of *w* to the audience's attention set, and the addition of *sI(aBw)*—but not *w*—to her belief set. More complexity and degrees of success arise in cases that combine the issues we've been discussing—for instance, an overt, indirect assertion. Hopefully, however, the pattern is clear by now: because the model represents speech acts as potentially involving a chain of goals, it can represent partial success by representing success at some—but not other—links in the chain.

8.5 Deception

In Chapter 4 I argued that the Stalnakerian framework is unable to adequately model deceptive speech, given that it represents discourse goals using a single, shared information set characterized by acceptance. While acceptance may be a by-product of deceptive speech, such speech aims at the finer-grained mental state of belief. Moreover, it aims to create a *divergence*—rather than an alignment—between the belief sets of the speaker and the audience. Because the proposed framework incorporates a variety of information sets that are characterized by finer-grained mental states, and because these information sets need not be shared, it provides a more adequate model of deceptive speech.

In order to see the explanatory function of non-shared information sets, let's look again at a diagram representing a simplified case of assertion (see Figure 8.10).

As we discussed earlier in the chapter, the first goal in the chain is to get the audience to *attend to m*. In the case of assertion, this initial goal is in service of yet another goal, which is to is to get the audience to *believe m*. Assertions may be sincere, or they may be aimed at deception; in neither case, however, does a shared information set do any explanatory work.

Supposing that Figure 8.10 represents a case of deceptive assertion, the speaker ultimately intends to update the audience's belief with a content *m* which she does not believe herself. Given that the speaker does not believe *m* at the time of her utterance—and given that her assertion is aimed at

Utterance → (*m*)   →   (*m*)

                     A   *B*
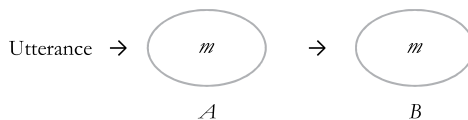
FIGURE 8.10 Diagram of assertion

convincing her *audience* rather than *herself*—her success will involve an update to the audience's belief set, rather than their shared belief set. On the other hand, if she is sincere, *m* is already part of the speaker's belief set; thus, an update to the audience's belief set with the content *m* will result in an update to their shared belief set with the content *m*. However, this update in their shared belief set is merely a *by-product* of the fact that the audience has updated her beliefs in a way that happens to bring it into alignment with that of the speaker. Even in the sincere case the shared information set is doing no explanatory work

Of course, there may be cases in which the speaker aims to update a shared information set, and so there is no principled reason to bar them from the framework—rather, the proposed model is not *restricted* to shared information sets. There are undoubtedly cases in which the speaker aims to update a shared information set, though these cases are more likely to involve shared *attention* than shared *belief*.[8] When I draw my partner's attention to a bird, or a shared memory, it is not merely that I intend for *him* to see the bird or experience the memory. My goal, rather, is that we watch the bird *together*, or that we reminisce about the past *together*. In this sort of case, the best way to model my speech act would appeal to an update to a shared attention set. Other times, shared attention may not be the goal. I might gesture to my partner's misplaced keys as I walk out of the room; here, my goal is to direct *his* attention to his keys. I have no interest in attending to his keys *together*. While assertion may sometimes be aimed at updating a shared belief set, this is not characteristically the case. When I assert a content *m*, my main concern is usually that *you* believe *m*, rather than that we believe *the same thing*. Deceptive assertions are the limiting case, in that they are *never* aimed at updating a shared belief set. Thus, while shared information sets may be helpful in characterizing some kinds of speech acts, non-shared information sets are needed to characterize others—most notably, deceptive speech.

One feature that offers the current proposal an advantage over the Stalnakerian framework in representing deception, then, is its use of non-shared information sets. Another is its appeal to mental states that are more varied and fine-grained than that of acceptance for the sake of conversation. This allows us to represent the fact that when the audience disbelieves a lie but fails to explicitly challenge it, the speaker has not fully succeeded in her aims. While she has managed to communicate the force and content of her utterance—and while the audience may accept the lie for the purpose of conversation—she has not achieved her ultimate goal of updating the audience's belief set in a way that diverges from her own. Fine-grained information sets such as belief allow us represent this case of partial success as an update of content of the lie to the audience's attention set, but not to her belief set. The Stalnakerian framework has trouble with such examples given its reliance on the attitude of acceptance, which is so coarse-grained

that its success conditions are almost vacuous. Because acceptance encompasses a broad range of attitudes including belief, supposition, etc., a characterization of its success conditions must be weak enough to apply to all of them. In light of this constraint, proponents of the Stalnakerian view suggest that the content of an utterance has been accepted if it has not been explicitly challenged. However, acceptance by silence is simply the wrong way to think about communicative success. Because different sorts of speech acts are aimed at updating different mental states, their success conditions will vary, and cannot all be measured by silence. On the proposed framework, we are able to be more specific about what constitutes the success conditions of particular speech acts, without having to use silence as a proxy for success. We can accommodate, for instance, the fact that an audience may fail to believe the assertion, but nonetheless be disincentivized to challenge it. This allows room for more nuanced explanations for the audience's silence: she fails to challenge the assertion because of, e.g., considerations of politeness or because of the incentive structures governing the power dynamics of the conversation, rather than because the speech act was successful in any substantive sense.

Thus far, I have been talking about the role of information sets in the proposed framework as functioning to represent various links in the chain of the speaker's conversational goals. However, the case of deception brings out the fact that a full characterization of discourse moves will also need to appeal to *other* information sets—i.e., information sets which play no part in characterizing the speaker's *goals*, but nonetheless serve in some way to characterize her speech act. As we've been discussing, the deceptive speaker aims to update her audience's belief set with a content that is not in her own belief set.[9] What differentiates deceptive cases of assertion from sincere cases of assertion is not the *goal* of the utterance, but rather its *background conditions*. In both cases, the speaker aims to update the belief set of the audience; deceptive cases are the subset of cases in which the speaker aims to update the audience's belief set with a content that she does not herself believe. This indicates that we will need to appeal to information sets that characterize the background conditions of the utterance, in addition to those that characterize the goals of the speaker. As we'll see below, such information sets have still further roles to play in the framework; they characterize the process of audience interpretation, as well as the information that the speaker exploits in order to achieve her goal.[10] They can also be used to characterize communicative goals that are directed toward multiple audience members.

8.6 Dogwhistles

The examples presented above were simplified in that they involved only one audience member, though of course this need not be the case. Conversations

involving multiple audience members introduce complexity because they raise the possibility that the speaker intends her utterance to achieve a different effect relative to different members of the audience. Dogwhistles serve to illustrate: When a politician utters the phrase "inner city", e.g., he may intend that some—but not other—audience members attend to and identify with a racist ideology. As we saw in Chapter 4, such cases posed difficulties for a Stalnakerian framework, which models speech as aimed at updating a shared information set. In order to represent this kind of case, our model will characterize the politician's speech act not with one chain of goals, as in the examples considered so far, but rather with separate chains of goals corresponding to sets of audience members (which may in some cases be singleton sets). Consider again, this excerpt from George W. Bush's (2003) State of the Union address:

BUSH DOGWHISTLE: "Yet there's power, wonder-working power, in the goodness and idealism and faith of the American people."

As Saul (2018) notes, it's plausible that with this utterance, Bush had different communicative goals relative to different subsets of the audience. Toward the subset of Christian fundamentalists, who would be familiar with the phrase "wonder-working power" as a reference to the power of Christ, the message will be that Christ's power is in the goodness and idealism and faith of the American people. We can abbreviate this first message as c, and the subset of the community it is directed toward as f. Toward the remainder of the audience, the message will correspond to the literal reading of the sentence—namely, that there is wonder-working power in the goodness and idealism and faith of the American people. We can abbreviate this second message as p, and the subset of the audience it is directed toward as g.[11] Because this case involves speech directed toward a large audience it will provide an illustration of how shared information sets retain a theoretical role within the theory. The shared attention set of the fundamentalist subset of the audience is abbreviated as A_f, while their shared belief set is abbreviated by B_f. For the remainder of the audience, we will use A_g and B_g for their shared attention and belief sets, respectively.[12] The double-messaging involved in this utterance can be represented by the framework using Figure 8.11, which maps out two separate chains of communicative goals, each relativized to a different subset of the audience. Without making any substantive assumptions about the complexity of Bush's communicative goals, I use the simplified model of assertion according to which it is aimed at updating the audience's attention set as a means of updating her belief set.

In this case, shared contents are needed to represent the fact that Bush likely had communicative intentions toward various subsets of the audience, rather than toward the individual members of these sets. Nonetheless,
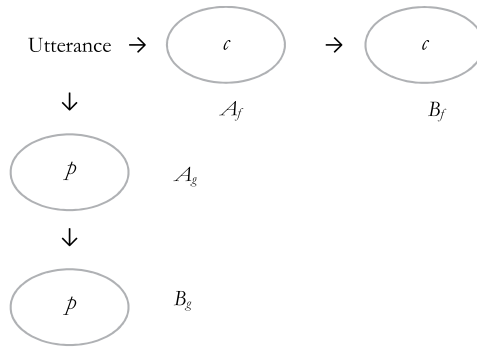
FIGURE 8.11 Diagram of dogwhistle

the case brings out the fact that in order to represent dogwhistles, we need information sets that are not shared by every member of the conversational exchange. Because the proposal incorporates multiple information sets that may be relativized to different audience members—or subsets of audience members—it has the tools to represent the kind of double-messaging that is involved in dogwhistling.

In order to keep the diagram (Figure 8.11) manageably simple, I've only provided a representation of the double-messaging involved in dogwhistling, ignoring any elements of covertness and indirectness. As we discussed in Chapter 3, however, dogwhistlers sometimes intend to keep at least some of their communicative intentions hidden from the very audience members that they are directed toward. These covert intentions can be represented similarly to Figure 8.5 above; for instance, we will represent that the speaker intends that A_g and B_g *fail to be* updated with the information the speaker intends to update A_f and B_f with c.

8.7 Miscommunication

In Chapter 4, I argued that because of its restriction to *shared* information sets, the classical Stalnakerian framework lacks the resources to provide an explanatory picture of the interpretive process involved in miscommunication. A *defective context*, recall, is one in which conversationalists diverge on what information they are accepting for the purpose of the conversation. In what Stalnaker assumes to be the "normal" case, speech acts are interpreted against the backdrop of a shared information set—i.e., the common ground. But in defective contexts, speech acts may be interpreted against information sets that are not shared; thus, conversational participants may have conflicting representations of how the conversation is unfolding, which may lead to a breakdown in communication. Because

the Stalnakerian framework models interpretation as occurring against the backdrop of a *shared* information set, it lacks the resources for representing interpretation processes which occur in defective contexts. Moreover, because the guiding background picture of language assumes that people have shared conversational goals, it wrongly predicts that defective contexts will tend to resolve themselves as a general rule. Because the proposed framework incorporates disjoint conversational goals and information sets, it has the tools for representing the interpretive process as it occurs in defective contexts.

Given the picture of language that I presented in Chapter 5, it is unsurprising that defective contexts are an ordinary feature of linguistic communication, and often exasperatingly difficult to resolve. On this picture, conversationalists often approach an exchange with misaligned goals, which explains the inexorability of some kinds of defective contexts. Of course, this is not to deny that in many cases our goals do align, leading us to engage in the kinds of repair strategies that Stalnaker predicts, and which are well documented by linguistic research.[13] The point is rather that this is not a general rule. We do not always readily adjust our own beliefs in order to repair a defective context, precisely because misaligned conversational goals disincentivize us to do so—thus, the communication breakdowns that occur in such contexts are not easily ameliorated.[14] The systematic misinterpretation of "Black lives matter" serves to illustrate the point. For years, All Lives Matter activists have been misinterpreting utterances of "Black lives matter" by wrongly assigning them the interpretation that *only Black lives matter*. This misinterpretation is at least partially a result of a defective context; as Anderson (2017a) notes, this "exclusive reading" is only available when ignoring background information which includes the mistreatment of Black Americans by the criminal justice system as well as the broader historical context of the history of racial violence in the US. It is unsurprising that this misinterpretation persists in spite of widespread and persistent attempts to clear up the intended message, given that ALM activists approach these conversations with incentives and goals that are misaligned with those of the BLM activists. While BLM activists are trying to draw attention to, and change an unjust social system, ALM activists often occupy social positions which allow them to benefit from that system. Thus, they are incentivized to direct attention *away* from facts about racial violence in the US, in order to evade confrontation with ways in which they participate in behaviors, structures, and institutions which reinforce the idea that Black lives do not matter. That is, they are disincentivized to adjust the information that they accept for the purpose of the conversation in a way that would line up with those of the BLM activists, and which would allow them to properly interpret their utterances. When we reject the idealized picture of language as cooperative information exchange, we

can more easily explain why defective contexts do not always easily resolve themselves; conversational goals sometimes conflict in ways that disincentivize participants to align their accepted background information with that of their interlocutor.

By incorporating non-shared information sets into our model of discourse structure, we are better able to represent the interpretive processes that occur within these kinds of contexts. Before demonstrating this, let me remind the reader of how this process works in the picture of language that I am defending. I've said that speech acts involve a chain of goals, which can be represented by various information sets. With each goal, the speaker aims to update some set of the audience's mental states (or in some cases, a set of their shared mental states) with a content. The audience's interpretation is what she takes to be the speaker's communicative goals. Here is where my picture diverges from the traditional Stalnakerian framework: in the process of interpretation—i.e., the process of forming beliefs about the speaker's communicative goals—the audience may draw on any information that she has on hand, whether or not it is shared. Suppose, for instance, that Anderson (2017a) is correct and that this misinterpretation is at least partially explained by the fact that ALM activists adopt a color-blind ideology—not shared by BLM activists—according to which we live in a society that is no longer characterized by racism and racist social structures. In interpreting utterances of "Black lives matter", ALM activists at least partially draw on information that is not shared between them and the BLM activists performing such utterances: namely, a set of beliefs that partially constitute a post-racial ideology. Not only the interpretive process, but also the conversational effect of such speech can be represented using information sets that are not shared by all conversational participants. In the proposed framework, the misinterpretation of "Black lives matter" will involve an update to the belief sets of ALM activists with the proposition that BLM protestors intend to communicate the exclusive reading (i.e., that *only* Black lives matter), but will involve a different update—namely that BLM protestors intend to communicate that Black lives matter—to the belief sets of others.

Non-shared information sets also provide the resources for representing the heterogeneity of communicative breakdowns. As I noted in Chapter 4, for example, there is an important difference between communication failures that result from the fact that interlocuters use different languages and those that result from the fact that interlocutors occupy different social positions. This difference can be marked in the framework by appealing to the *kind* of information that the audience draws upon in interpreting the speaker. In the former case, the communication failure is explained by differences in beliefs about language, while in the latter case the communication failure is explained by differences in beliefs about social facts. There

also seems to be an important distinction between one-off communication failures vs. those that are systematic and pattern with other important phenomena. Again, this is something that our framework can represent when we have access to non-shared information sets. One-off cases will be just that; cases that don't present an interesting pattern. In contrast, we can represent the systematicity of the ALM response to "Black lives matters" by looking at the sorts of beliefs that this response correlates with. Again, if Anderson's (2017a) hypothesis is correct, you'd see it occurring more often among interpreters who have internalized a color-blind ideology.

Again, I want to reiterate that the need to incorporate non-shared information sets in modeling interpretation does not imply that we thereby reject any role for shared information sets. In many cases, the audience will draw upon shared information in order to interpret the speaker, and in many cases the speaker will design her utterance with this fact in mind. Recall that in overt speech acts, for instance, the speaker intends that her audience recognize her communicative intentions. In such cases, she may rely on shared beliefs in order to make her intentions apparent. For instance, I may choose my words based on what I take to be our shared knowledge of the language in order that you more easily recognize my communicative intentions. (Note, however, that this *need* not be the case—I may purposefully use a malapropism if I have reason to believe that you will assign it a non-standard interpretation.) Or I may exploit our shared knowledge of the external environment; e.g. I anticipate that you will recognize my referential intention in uttering a demonstrative because I believe that we mutually perceive a certain object to be salient. Similarly, on the interpretive side, you may exploit what you take to be our shared beliefs in order to interpret me by, e.g., relying on what you take to be our shared knowledge of a public language, or mutually salient objects in our shared environment. (But again, this need not be the case; if you think I've used a malapropism, you'll appeal to my false beliefs about language—which you do not share—in order to identify my communicative intentions). Thus, while the proposed framework makes room for non-shared information sets within its representational framework, it nonetheless retains a role for shared information sets.

In the Stalnakerian framework, a single information set is used to do all of this work: the common ground represents the information that the speaker exploits in order to make her contribution apparent, the background against which the contribution is interpreted, and the outcome of that contribution. While its use of a single information set to play all of these theoretical roles lends the classical Stalnakerian framework a desirable level of elegance and simplicity, it also limits its representational and explanatory capacity. The added complexity that results from multiplying goals, information sets, and mental states is justified by the representational and explanatory power that this lends to the proposed framework.

8.8 Conclusion

Throughout this chapter I've shown how a pragmatic framework that rejects the five Stalnakerian idealizations can overcome problems faced by that theory. My goal in this chapter has not been to propose a complete model of discourse structure, but rather to provide a sketch of the minimal tools needed to represent our actual linguistic practices. I've also shown how this framework could be used to represent speech acts and their interpretation according to the picture of linguistic communication I presented in Chapter 5. Throughout this chapter I've been using examples and illustrations that are simplified in a number of respects. In this final section I discuss various ways that linguistic communication may be more complex, and suggest a few ways that this complexity may be represented within the framework.

Though the speech acts I considered above involved fairly straightforward goals, we might also want to be able to represent conditionalized goals. The goals involved in insinuation, for example, seem to be conditionalized. In attempting to bribe a policeman, for instance, my goal may be for him to recognize this goal conditional on his accepting the bribe—and to fail to recognize this goal conditional on his failing to accept the bribe.[15] In this case, we may want to represent the chains of goals corresponding to speech acts to be relativized not only to sets of audience members, but also possible worlds.

Another additional complexity is that that though I've presented the bodies of contents corresponding to links in the chain of goals as *sets*, an adequately complex account will need to introduce *structure* to these bodies. When performing utterances, speakers are not always aiming to add an element to a body of contents—rather, their aim may be, for instance, to structure that body of information or to provide background information about the evidential status of that information. In such cases, the desired update will involve imposing structure on a body of content rather than adding to it. In order to accommodate these sorts of cases our pragmatic framework will need to incorporate both internal and external structure. Functional expressions, for instance, may add internal structure to a body of contents by placing them in temporal or causal relations. Evidential markers, on the other hand, introduce external structure by relating one set of contents to a second set of contents that indicates the evidential status of the first.

The need for external structure also arises when we consider utterances involving non-at-issue content. While the examples above were restricted to at-issue content, consideration of non-at-issue content shows that a speaker's goals cannot be individuated solely by their position in a chain and the information set that they aim to update. Consider, for instance, an utterance

of "Variegated monstera plants, which are rare, can be propagated by rooting a cutting in soil or water." Suppose that the speaker aims to add this content to the audience's attention set. In this case, we want to mark a distinction between the information that variegated monstera plants are rare, which is not at issue, and the information that these plants can be propagated by rooting a cutting in soil or water, which is at issue. But given that the speaker intends the audience to attend to both of these contents, we cannot individuate them by the information set that they are intended to update. Nor can we individuate them by their position in a chain of goals—the speaker does not intend to update the audience's attention set with either content *on the basis of* the other. Rather, in this kind of case, we will need to represent the audience's attention set as being divided into subsets—one containing at-issue content and the other not-at-issue content. Again, we can appeal to external structure in order to model the different communicative roles that each type of content plays. As I suggested in Chapter 5, perhaps these information sets could be arranged hierarchically; at-issue content would occupy a higher position in the hierarchy in order to represent the fact that the speaker intends this content to occupy the audience's attention to a stronger degree than not-at-issue content.

The simplified framework presented above is also amenable to reinforcement by more sophisticated theoretical machinery associated with the dynamic semantics tradition that has grown out of the work of Stalnaker and others,[16] as well as other traditions in formal pragmatics. For instance, following Roberts (2012), the framework could make use of Questions Under Discussion and To-Do Lists to represent certain types of conversational goals. Following Heim, Kamp (1981), etc. the framework could also be fortified with tools serving to organize discourse referents. While my focus has been on the minimal conditions the framework needs to exhibit in order to represent certain features of non-ideal speech, a great deal more complexity would be needed in order to adequately model other important linguistic phenomena.

Finally, as I mentioned at the outset, the framework I've presented here is overly simplified in that it can only represent discourse phenomena relating to content-bearing mental states. But there are additional factors that play a significant role in the production, interpretation, and effects of speech. For instance, oftentimes the primary aim of speech is to produce an emotion or behavior, rather than to produce an attitude toward a particular content. Wider social structures also play a significant role in the production, interpretation, and effects of speech. I've suggested ways in which we might be able to represent patterns that correlate misinterpretation of "Black lives matter" with certain kinds of beliefs—but we might also want to represent the way that the ALM response is both a product of, and has a tendency to reinforce, racist social structures. In order to model this sort of

phenomena, the tools of the framework will need to be expanded in order to represent not only mental attitudes, but also emotions, behaviors, and social structures.

Notes

1 Keiser and Ritchie (n.d.).
2 See Buchanan (2010) for discussion.
3 As we'll see, depending on how finely-grained we model the links in the chain, attention-direction may occur *near* the first position of the chain without, strictly-speaking constituting the *first* link. However, for ease of exposition I drop the caveat "roughly" in what follows.
4 For an example of this type of account of assertion, see Bach and Harnish (1979). I defend a similar account in Keiser (2016).
5 Where there is no means-end relationship between the speaker's intention, the diagram will branch rather than form a chain, as in Figure 8.5.
6 Though in some cases, perhaps failure to recognize the speaker's intentions could be a precondition for other goals (as in the case of codewords). Thanks for Thomas Brouwer for this point.
7 I consider goals that are relativized to specific audience members in the final section.
8 Communicative goals directed towards shared attention are thought to play an especially crucial role in early language development and early forms of communication. See Tomasello and Farrar (1986).
9 Note that we can complicate the story to fit a more nuanced account of assertion. For instance, maybe the speaker also wants to update the audience's belief set with the content that the speaker believes m.
10 Cf. Stalnaker's (1998) discussion of the "dual role" of context.
11 This is a simplification; I do not assume that there were only two intended messages, nor that f and c are a precise characterization of them. But I follow Saul in thinking that something like this is roughly correct.
12 I am thinking of a shared attention set of a group p (and shared belief set, *mutatis mutandis*) as being the intersection of the individual attention sets of each member of p. I am not including any publicity condition in the sense of requiring the community members to believe that they have shared attitudes toward this information.
13 See Hayashi, Raymond, and Sidnell (2013).
14 See Anderson (2017b) and Peet (forthcoming) for more in-depth discussion of why defective contexts can be difficult to resolve.
15 For this reason, as Camp (2018) has noted, she will likely design her utterance in a way that preserves plausible deniability about what she meant.
16 Relevant others include Kamp (1981), Heim (1983), and Groenendijk and Stokhof (1991).

References

Anderson, L. (2017a). Epistemic injustice and the philosophy of race. In J. Medina, I. J. Kidd, and G. Pohlhaus Jr. (eds.), *The Routledge Handbook of Epistemic Injustice* (pp. 139–148). New York: Routledge.

Anderson, Luvell (2017b). Hermeneutical impasses. *Philosophical Topics*, 45(2):1–19.

Bach, Kent and Harnish, Robert M. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Buchanan, Ray (2010). A puzzle about meaning and communication. *Noûs*, 44(2):340–371.

Camp, Elisabeth (2018). Insinuation, common ground, and the conversational record. In D. Harris, D. Fogal, and M. Moss (eds.), *New Work on Speech Acts* (pp. 40–66). Oxford: Oxford University Press.

Groenendijk, Jeroen, and Stokhof, Martin (1991). Dynamic predicate logic. *Linguistics and Philosophy*, 14(1): 39–100.

Hayashi, M., Raymond, G., and Sidnell, J. (eds.). (2013). *Conversational Repair and Human Understanding (Studies in Interactional Sociolinguistics)*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511757464.

Heim, Irene (1983). On the projection problem for presuppositions. In P. Portner and B. H. Partee (eds.), *Formal Semantics - the Essential Readings* (pp. 249–260). Malden, MA: Blackwell.

Kamp, Hans (1981). A theory of truth and semantic representation. In P. Portner & B. H. Partee (eds.), *Formal Semantics – The Essential Readings* (pp. 189–222). Malden, MA: Blackwell.

Keiser, Jessica (2016). Bald-faced lies: How to make a move in a language game without making a move in a conversation. *Philosophical Studies*, 173(2):461–477.

Keiser, Jessica, and Ritchie, Katherine (n.d.). Social structures in context. (unpublished ms).

Peet, Andrew (forthcoming). Defective contexts. In Rachel Katharine Sterken and Justin Khoo (eds.), *Routledge Handbook of Social and Political Philosophy of Language*. New York: Routledge.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Saul, Jennifer (2018). Dogwhistles, political manipulation, and philosophy of language, in Daniel Fogal, Daniel W. Harris, and Matt Moss (eds.), *New Work on Speech Acts* (pp. 360–383). Oxford: Oxford University Press.

Stalnaker, Robert (1998). On the representation of context. *Journal of Logic, Language and Information*, 7(1):3–19.

Tomasello, M., and Farrar, M. (1986). Joint attention and early language. *Child Development*, 57(6):1454–1463.

INDEX

Pages followed by n refer notes.