Communication and content

Prashant Parikh

Topics at the Grammar-Discourse Interface 4

language science press

Topics at the Grammar-Discourse Interface

Editors: Philippa Cook (University of Göttingen), Anke Holler (University of Göttingen), Cathrine Fabricius-Hansen (University of Oslo)

In this series:

1. Song, Sanghoun. Modeling information structure in a cross-linguistic perspective.
2. Müller, Sonja. Distribution und Interpretation von Modalpartikel-Kombinationen.
3. Bueno Holle, Juan José. Information structure in Isthmus Zapotec narrative and conversation.
4. Parikh, Prashant. Communication and content.

ISSN: 2567-3335

Communication and content

Prashant Parikh

language
science
press

Parikh, Prashant. 2019. *Communication and content* (Topics at the Grammar-Discourse Interface 4). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/248

ISBN: 978-3-96110-198-6 (Digital)
978-3-96110-199-3 (Hardcover)

ISSN: 2567-3335
DOI:10.5281/zenodo.3243924
Source code available from www.github.com/langsci/248
Collaborative reading: paperhive.org/documents/remote?type=langsci&id=248

Cover and concept of design: Ulrike Harbort
Typesetting: Felix Kopecky
Proofreading: Ahmet Bilal Özdemir, Brett Reynolds, Catherine Rudin, Christian Döhler, Eran Asoulin, Gerald Delahunty, Jeroen van de Weijer, Joshua Phillips, Ludger Paschen, Prisca Jerono, Sebastian Nordhoff, Trinka D'Cunha, Tom Bossuyt, Umesh Patil, Vadim Kimmelman
Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono
Typesetting software: XƎLATEX

Language Science Press
Unter den Linden 6
10099 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

for john perry and ennio stacchetti

and

for avani and neal

‘The time has come,’ the Walrus said,
‘To talk of many things:
Of shoes – and ships – and sealing-wax –
Of cabbages – and kings –
And why the sea is boiling hot –
And whether pigs have wings.’

— Lewis Carroll, *The Walrus and the Carpenter*

Praise for *Communication and content*

This book is the culmination of Prashant Parikh's long and deep work on fundamental questions of language and how they can be illuminated by game-theoretic analysis.

— Roger Myerson, 2007 Nobel Laureate in Economics, University of Chicago

Prashant Parikh has, over the years, accumulated a substantial and impressive body of work on the nature of language, deploying the resources of game theory. *Communication and content* is a vastly ambitious culmination of this lifelong pursuit. It covers a tremendously wide range of themes and critically discusses an enormous range of writing on those themes from diverse intellectual traditions, as it systematically develops a game-theoretic account of content in the communicative contexts in which human linguistic capacities are employed, eschewing standard distinctions between semantics and pragmatics, and offering instead a highly integrated elaboration of the slogan "meaning is use". It is a work that is at once creative yet conscientious, bold yet rigorously technical, systematic yet sensitive to contingency and context. It will abundantly reward close study.

— Akeel Bilgrami, Sidney Morgenbesser Professor of Philosophy, Columbia University

Prashant Parikh has made fundamental contributions to the game-theoretic analysis of linguistic meaning. *Communication and content* summarizes and extends this important work, offering a truly novel approach to the strategic foundations of meaning. This approach finds a way out of the prison of methodological solipsism and opens up the study of linguistic meaning to scientific study.

— Robin Clark, Linguistics, University of Pennsylvania

Praise for Communication and content

A pioneering attempt to work out things like literal meaning, modulation, enrichment, implicature, etc. in mathematical detail within a game-theoretic framework.

— François Recanati, Chair, Philosophy of Language and Mind, Collège de France

Communication and content is the crowning achievement of a long line of research pioneered by Prashant Parikh. In this groundbreaking work Parikh introduces a fresh perspective on natural language pragmatics, by making a creative tie with game theory. Clearly written, *Communication and content* weaves together semantics, game theory, and situation theory to create a thought-provoking picture of natural language pragmatics. Every modern AI researcher interested in the foundations of natural language pragmatics owes it to him- or herself to become familiar with this picture.

— Yoav Shoham, Computer Science Department, Stanford University

Contents

Acknowledgments

This book owes more to John Perry, Dan Flickinger, and Ennio Stacchetti than I can possibly express.

In addition to his own invaluable comments, John organized an informal international seminar at Stanford University conducted over Skype whose members – Eros Corazza, Maria de Ponte, Kepa Korta, JJ Lang, Fenrong Liu, and especially Dan – provided extremely useful feedback. My grateful thanks to them all.

Ennio discussed and also contributed to the game-theoretic model I was developing over multiple conversations at New York University. I give him my everlasting thanks.

I would also like to thank Joel Sobel and Navin Kartik with whom I had an extended email interaction. I am grateful as well to the game theorists Ken Binmore and Vincent Crawford, both of whom provided useful comments.

Over the years, I have had interesting exchanges with and would like to thank Kent Bach, Anton Benz, Joan Bresnan, Nelson Correa, Michael Devitt, Fred Dretske, Dagfinn Føllesdal, Joe Halpern, James Hampton, Larry Horn, Gerhard Jäger, Gregory Murphy, Daniel Osherson, Liina Pylkkänen, Robert Stalnaker, Herb Terrace, Tom Wasow, Deirdre Wilson, Terry Winograd, and Ed Zalta.

My deepest debt is to my parents Jagdish and Shaila, my wife Avani, and my son Neal. Without Avani, this book would never have come to be.

Part I

Introduction

1 Why communication is central to meaning

There are many kinds of things in the world. There are individuals such as planets and stars, ants and humans, tables and chairs, and books and civilizations. Each of these individuals has an infinite number of properties such as having a size and shape or being more or less complex. And these individuals also stand in relations such as a planet circling a star, an ant being smaller than a human, a chair being beside a table, or a book belonging to modern times. Complexes of these, individuals having properties or standing in relations, form *infons*, and collections of infons form situations, parts of the world we occupy as agents and move about in. A situation may contain the earth circling around the sun, a person on earth sitting at a desk writing a book, and a scurrying ant. Such an inventory, consisting of individuals, properties, relations, infons, situations, and some other kinds of entities, makes up *reality*.

All agents, whether ants or humans, carve up reality and create an *informational space* or *ontology* to navigate the world. That is, while the entities in such a space are part of reality, they need to be individuated as distinct entities for them to become part of an agent's informational space. An ant does not individuate a mountain just as a person may not discriminate between two shades of blue.

In carving up reality, one of the key things agents do is discover or create connections or constraints between two or more items of information such as, say, smoke and fire or an utterance and its content. When such a connection occurs with a certain regularity, we call it a *meaning*. Indeed, we say that smoke means fire and, likewise, that an utterance of "There is a fire" in suitable circumstances means there is a fire in some situation. Reality is full of such constraints or meanings. Science, broadly construed, is just the attempt to comprehend them and, in this sense, it can be said to be *the search for meaning*.

1.1 Semantics

This vast domain of constraints in reality can be roughly divided into natural and artificial, that is, those that are independent of humans and those that depend on them. Smoke meaning fire is a natural constraint and an utterance of "There is a fire" meaning there is a fire in some situation is an artificial constraint.

Some artificial constraints like the one above involve *symbol systems* that enable a person to represent one object by another and convey this to another person. A traffic light tells one to stop or to go. Representation does not just exist in the world; it requires agency. It appears that a symbolic consciousness emerged a few hundred thousand years ago with the Neanderthals, or possibly much earlier with *Homo erectus*, although this has yet to be confirmed.

Semantics is or ought to be the study of symbol systems. In some traditions, this discipline is called semiotics or semiology. My use of "semantics" is considerably broader than simply the representing of utterance content as in formal semantics.[1] One of the most complex symbol systems is (verbal) language and so semantics has often come to be more narrowly identified with the study of linguistic meaning. A study of four traditions – Sanskrit, Greek, Hebrew, and Arabic – reveals that semantics emerged as an independent discipline roughly three thousand years ago from the exegesis of mostly religious texts and concerned the relationship between language, reality, and knowledge.[2] Its original context was *communicative*, initially between people and divine powers when the world itself was largely read as signs from above. Today, semantics belongs primarily to philosophy, linguistics, psychology, and artificial intelligence, and derivatively to other fields, each studying it from different points of view and with different ends. The reason for its wide scope is that symbolic meaning is central to life itself.

Arguably, its main problem is to understand how language acquires meaning. How is it that an utterance, whether spoken or written, carries one or more meanings from a speaker to an addressee? Because language has so many varied devices to convey meaning, this turns out to be a very large question made up of very many subquestions. To follow some parts of this problem, consider a simple puzzle concerning names.

1.2 A classic example

Consider the following sentence:

(1) Hesperus is Phosphorus. (ψ)

[1] I say more about this in Section 5.4.

[2] See van Bekkum et al. (1997: 286). See also Deutsch & Bontekoe (2007) for other traditions.

Many readers will recognize ψ as a variant of a sentence considered by Gottlob Frege (1892/1980), perhaps the founding figure of modern semantics. A first attempt at apprehending its content might be to say that a name refers directly to the object it names. For ψ this would imply that both the words HESPERUS and PHOSPHORUS refer to the planet Venus. Because IS can be taken to stand for equality, we can infer that ψ asserts the same thing as "Hesperus is Hesperus." But this is a problem because the latter utterance is a trivial identity whereas ψ involves an empirical discovery. This is Frege's puzzle of informative identities: how can we explain the new information conveyed by ψ?

Frege's solution was to posit an intermediate layer of meaning called the *sense* of each name and the object Venus they both stand for would be the *reference*. His idea was that every name expresses a sense which in some cases leads to a reference. Though the referents of the two names in ψ are identical, their senses are different, the first being the evening star and the second being the morning star, and it is these different senses that capture the cognitive significance of ψ unlike the trivial identity "Hesperus is Hesperus" where even the senses are the same. This solution runs into insuperable problems with more complex sentences although his two-tier account of meaning was on the right track. For now, I want to point out a pivotal thing he left implicit.

He did not sufficiently articulate the distinction between a sentence and an utterance. A sentence – like a word – is an abstract object whereas an utterance of a sentence is an action performed in some situation. Everyone knows the difference between the two but Frege mentioned it mainly when he discussed force.[3] But it is vital even for his theory of sense and reference because a sentence by itself does not have a content: it is only when it is *used* appropriately in an utterance situation that it acquires a content. It appears to us that ψ has a content only because we implicitly evaluate it as an utterance in some unspecified but familiar situation. But a little thought would reveal that in a different situation a speaker may be referring to the figure in Greek mythology who was a personification of the evening star and not to the evening star itself. In other words, HESPERUS is *ambiguous* and the addressee needs to know which meaning is intended. Differently put, Frege abstracted from its use and assumed the meaning of interest to him. A more complete theory would have to start by showing first how the utterance is disambiguated.

Frege and the logicist tradition he inaugurated correctly identified reference as a key property of language. In so doing, the ideal language philosophers succeeded in stating one of the two main subproblems of semantics: how does language relate to reality, how does it connect with the world? The other main

[3] By "force" Frege meant assertion or other such acts performed as part of uttering a sentence.

subproblem, bequeathed from ancient times, concerns the possible knowledge it brings about. But by sticking to sentences and ignoring utterances they left out the key means to their solution: the use of language – or *communication*. Just as Frege did, many who followed him chose to abstract from communication and focus on particular linguistic devices and their problems, as we have just seen with the puzzle of informative identities. They did this generally by employing truth as a way to get at meaning because Frege had pointed to a truth value as the reference of a sentence. It was the later Wittgenstein (1953/1968) followed by Austin (1961; 1975) and Grice (1989c) and other ordinary language philosophers who realized the importance of seeing language as a situated activity but did not quite succeed in unraveling this elusive concept.[4]

It would seem there are two ways to approach the connection between language and world. One is to break down language into various constructions such as noun and verb phrases, select some specific devices, and create a framework to see how they work via their truth conditions. Some adherents still stick to sentences as the bearers of truth values but many have shifted to utterances because the situated nature of language makes the former increasingly difficult as the field advances and more details come into view.

The other is to recognize that unless one first fathoms simple cases of communication in their full complexity across a number of devices, one cannot create a *general* framework that will be able to address the many problems they raise in a uniform way. And doing so allows us to largely avoid the tricky notion of truth because meaning would be derived from its use in communication.

The first approach is bottom-up and the second top-down. There is a grave danger that the first will lead to a proliferation of incompatible theories presupposing different views of meaning at a more foundational level. My view is that once communication and content are understood thoroughly with simple constructions in a general way, it will be much easier to tackle more complex constructions with an elaboration of the same tools so that the many subquestions of semantics all cohere into an integrated solution to its main problem.

[4]See the fascinating account by Sen (2003) of how the economist Piero Sraffa (and indirectly the political theorist Antonio Gramsci) may have influenced Wittgenstein in making his celebrated move from his early logicist ideas to his later use-based ideas. As Sen puts it, "Wittgenstein told a friend (Rush Rhees, another Cambridge philosopher) that the most important thing that Sraffa taught him was an "anthropological way" of seeing philosophical problems. In his insightful analysis of the influence of Sraffa and Freud, Brian McGuinness (1982: 36–39) discusses the impact on Wittgenstein of "the ethnological or anthropological way of looking at things that came to him from the economist Sraffa." While the *Tractatus* tries to see language in isolation from the social circumstances in which it is used, the *Philosophical Investigations* emphasizes the conventions and rules that give the utterances particular meaning. The connection of this perspective with what came to be known as 'ordinary language philosophy' is easy to see."

From a related but different vantage point, Dummett (1996: Chapter 3) offers the following insights.

> In fact, it is by grasping what would render [a sentence] true that we apprehend what it means. There can therefore be no illuminating account of the concept of truth which presupposes meaning as already given: we cannot be in the position of grasping meaning but as yet unaware of the condition for the truth of propositions. Truth and meaning can only be explained *together*, as part of a single theory. (page 15)
>
> ⋮
>
> Or, at least, they have to be explained together so long as Frege's insight continues to be respected, namely that the concept of truth plays a central role in the explanation of sense. On this Fregean view, the concept of truth occupies the mid-point on the line of connection between sense and use. On the one side, the truth-condition of the sentence determines the thought it expresses, in accordance with the theory developed by Frege and adapted by Davidson; on the other, it governs the use to be made of the sentence in converse with other speakers, in accordance with the principles left tacit by both of them. *That leaves open the possibility of describing the use directly, and regarding it as determining meaning, relegating the concept of truth to a minor, non-functional role.* [my italics] This was the course adopted by Wittgenstein in his later work. The concept of truth, no longer required to play a part in a theory explaining what it is for sentences to mean what they do, now really can be characterized on the assumption that their meanings are already given. (page 19)
>
> ⋮
>
> Rather than characterizing meaning in terms of truth-conditions, and then explaining how the use of a sentence depends upon its meaning as so characterized, this approach requires us to give a direct description of its use: this will then *constitute* its meaning.
>
> The disadvantage of this approach lies in its unsystematic nature. This, for Wittgenstein, was a merit: he stressed the diversity of linguistic acts and of the contributions made to sentences by words of different kinds. Systematization is not, however, motivated solely by a passion for order: like the axiomatic presentation of a mathematical theory, it serves to isolate initial assumptions. A description of the use of a particular expression or type of

> sentence is likely to presuppose an understanding of a considerable part of the rest of the language: only a systematic theory can reveal how far linguistic meaning can be explained without a prior supply of semantic notions. The ideal would be to explain without taking any such notions as given: for it would otherwise be hard to account for our coming by such notions, or to state in any non-circular manner what it is to possess them. It is unclear from Frege's work whether this ideal is attainable. The indefinability of truth does not, of itself, imply the inexplicability of truth, although Frege himself offered no satisfactory account of elucidations that fall short of being definitions. Perhaps the concept of truth can be adequately explained if a substantive analysis of the concepts of assertion and of judgement is feasible: but Frege leaves us in the dark about this. Wittgenstein, equally, leaves us in the dark about whether his programme can be executed: it is another disadvantage for the repudiation of system that it leaves us with no way of judging, in advance of the attainment of complete success, whether a strategy is likely to be successful. (pages 20–21)

Dummett provides a further reason to pursue meaning via communication: to sidestep the interdependence of truth and meaning. My approach squarely pursues the possibility of describing use directly and deriving meaning from use. However, contra Wittgenstein's "repudiation of system" I will attempt a complete and systematic account of communication and meaning.

There is a third advantage to reducing meaning to communication. It allows us to see meaning as a natural phenomenon as opposed to something otherworldly. With the truth-conditional strategy, this possibility remains murky at best. Grice (1989c) may have been the first person to outline a concrete program for such a reduction even though its details are different from my approach. I discuss his framework in Chapter 5.

Lastly, ever since the later Wittgenstein mentioned it, various writers have been trying to pin down the normativity of language and meaning. I will try to show that even this intangible property emerges more clearly once communication is grasped in a precise way.

1.3 A snapshot of semantics

Communication has at least six dimensions: its contextual or situated nature, its involving actions or utterances and interpretations, its encompassing epistemic, practical, and social interactions between speaker and addressee, and its being computationally tractable. For the moment, they can be understood intuitively.

Table 1.1 displays how some prominent approaches to communication fare against these dimensions.

Table 1.1: Summary of theories of communication

	Logicism	Wittgenstein	Austin	Grice	Lewis
context	marginal	implicit	implicit	implicit	implicit
action	partial	yes	yes	yes	yes
epistemic interaction	no	no	implicit	yes	yes
practical interaction	no	implicit	no	no	partial
social interaction	no	no	no	no	no
computable	no	no	no	partial	partial

The table identifies how different frameworks deal with the *how* of communication, not the what. It deals with its process rather than its content. Two distinct "yesses" in the same row do not imply that both accounts do equal justice to the dimension. As will become clear, my framework – called Equilibrium Semantics – scores a "yes" in all rows except the one pertaining to social interaction, where it scores a "partial" rating.

1.4 Equilibrium Semantics

Building on my books *The Use of Language* (2001) and *Language and Equilibrium* (2010), I present a new framework in this book that can serve as a core part of a science of communication usable primarily by philosophy, linguistics, and artificial intelligence, and also perhaps by cognitive and social science, especially economics, and related disciplines. I will try to show that the models I develop are not only philosophically sound and mathematically solid but also computationally tractable and empirically adequate, and I will connect them to the subjects above in multiple ways. The nature of communication is such that I will also have to draw upon insights from these diverse areas, especially philosophy, linguistics, psychology, and economics. This poses a challenge for both the writer and reader, as I have to strive to make all of it accessible and you may have to take an interest in occasionally unfamiliar ideas, concerns, and methods. But I believe the journey will be well worth the effort as I offer a substantial advance on the state of the art with results that are both more general and more precise.

In my view, the mathematics best suited for modeling communication comprises certain innovations in game theory and situation theory that I have described earlier and that I add to here.

Informal game-theoretic thinking is quite old, going back to classical times in multiple cultures. Possibly its modern roots can be traced to Machiavelli and Hobbes. Its first formal results were obtained by Zermelo (1913) in connection with the game of chess. As Myerson (1991: 1) says, *interactive decision theory* might have been a more descriptively accurate name for the theory as it allows one to consider practical and epistemic interactions between people in a computationally tractable way.

Lewis (1969) was the first to introduce game theory into semantics but he used rather simple games from Schelling (1960) and viewed them as dispensable scaffolding. At this early moment, game theory was not seen as offering a framework that could partly replace logic. The Austin-inspired idea of use conditions as an alternative to truth conditions was often entertained but such conditions were also seen as a part of logic.

As I have shown in my dissertation *Language and Strategic Inference* (1987b) and since, game theory allows one to discard such logically stated use conditions and offer precise mathematical derivations of content instead. This work, together with some aspects of Lewis's way of modeling games, has led to a small but growing field in recent times with many papers and also books such as Benz et al. (2006), Pietarinen (2007), Clark (2011), and Benz et al. (2011). Shoham & Leyton-Brown (2009: Chapter 8) describe some of my ideas in the context of distributed systems. The key difference between some of the game-theoretic linguistic approaches that have emerged and my framework is that the former are largely orthodox Gricean and the latter is not, as will become especially clear in this book. This has meant that the former have focused mainly on deriving Gricean implicatures whereas I have tried to derive literal meaning as well.

The theory of situations was originally developed by Barwise & Perry (1983) and Barwise (1989d) and drew upon ideas from Shannon & Weaver (1949) and Austin (1961/1979a). It is essentially a qualitative account of information and I have already described some of its key ideas informally, especially the idea of a situation which forms the context for an utterance. Later we will see how it also allows us to capture the content of an utterance in a very general way.

I will confine myself to language although the primary means by which we convey meanings[5] to one another include words, images, and gestures. Indeed, gestures preceded words and images in our evolution. My account can be extended straightforwardly to images and gestures as well as to other symbol systems and so will also be of interest to the corresponding fields.

One of my aims in this book is to show how little needs to be assumed to construct a science of communication, mainly facts about ontology, the partial rationality of agents, and a language and its grammar and phonetic system. I want to do this in an informally foundational way so I will introduce various concepts when they are required, in a more or less linear fashion. I will therefore start with information as the most basic "building block" instead of plunging directly into communication.

A science of communication rests on a science of information for two reasons. The first is that the world is information as such and we are embedded in it. Communication therefore occurs within an informational space and cannot be understood without it. The second is that communication involves conveying information in one form or another even when we ask questions, issue commands, or express feelings. It relays *content* and has an *effect* and both are informational, although the former affects us cognitively and the latter may involve the whole person. So I begin with a brief tour of this pervasive landscape.

[5] I will use "meaning" and "content" interchangeably in many contexts although I will also use "content" more widely to include the grammatical and phonetic structure of the sentence uttered. There will be times when I will explicitly distinguish conventional meanings from referential meanings, the latter being the same as semantic contents.

2 Information and agents

2.1 Information

I started the first chapter with a quick description of an informational space. I elaborate on it here.

2.1.1 Philosophical background

Reality is all the stuff that makes up the world, not only the tables and chairs in it but also a table's *being* white and a chair's *being* black. Perhaps Wittgenstein (1921/1961: 5) put this insight most grandly: "The world is the totality of facts, not of things." I include both facts and things as well as other entities.

Reality is sliced up differently by different living things, partly because they possess different kinds of bodies and partly because they have different concerns. That is, ants, dogs, and humans *individuate* and *discriminate* reality in different ways,[1] both at a basic categorial level that is constant across a species and at a more fine-grained level that varies with each agent. *Individuation* involves an analog-to-digital conversion, possibly along with an application of symbols to reality.[2] *Discrimination* involves just an analog representation without any digital conversion.

So ants may see one facet of reality, dogs another, and humans a third, each based on a distinct perceptual and cognitive apparatus. And two persons may further individuate their shared ontologies in contrasting ways, as happens when one sees the proverbial glass as half full and the other as half empty. Ontologies depend on reality and so cannot be arbitrary and must also be compatible with

[1] See Devlin (1991: Chapter 2). For some evidence of this, see Carruthers (2004).

[2] When some information is in analog form, there is always more specific information about the object also present. Digital information is the most specific information about an object in a representation. If someone says a cup has coffee in it, this information is in digital form; when a picture of the cup is shown instead, the same information is in analog form because it conveys many more specific facts, such as the extent to which the cup is filled, what kind of cup it is, and so on. See Dretske (1981: 177). The analog/digital distinction is related to but not the same as the continuous/discrete distinction. The former could perhaps be thought of as a generalization of the latter.

one another. That is, if a real glass is in fact half filled with water, one cannot without error see it as only a third full, but one can see it as either half full or half empty. An informational space can be viewed as a coordinate system an agent uses to orient itself in reality.

People find themselves in a "world" or *environment* from the start, as the phenomenological tradition[3] has emphasized. This is not the entire world or all of reality but a small part of it. This observation seems obvious but the analytic tradition of Frege and Russell missed it and missed its profound consequences. First, in all social matters and especially in the domain of communication, this environment or *context* or *background* is indispensable. Even today, a century later, its presence in the study of meaning is stubbornly resisted by many. Less obvious is the fact that our informational space, a subspace of our environment, results from society's interaction with reality partly via communication.[4] Third, this context of communication often has indeterminate boundaries and this results in many other uncertainties. Once individuated, each informational item is linked with others that form its context, although in formal domains such as mathematics and physics this context is more determinate. For example, a certain fact may be true of equilateral triangles and then this condition forms the context for the fact. In everyday life, the context is less determinate.[5]

There is a premature rush among analytically minded researchers to secure the respectability of science and it is not realized that their abstractions often miss the essence of the problem. A glaring example of this is the fundamental difference between a sentence and an utterance, something Wittgenstein (1953/1968), Austin (1961), and Strawson (1956; 1964) were the first to emphasize. When approaching the social sciences, one has to straddle both their abstract and concrete sides, and recognize that a complete science may not be possible because of the indeterminate and partly unstructured nature of context. And yet, as I will show, a remarkable degree of scientific success is possible if the task is approached in the right way. Surprisingly, philosophers especially, despite their desire to emulate the sciences, have often kept away from relatively new mathematical tools such as game theory and have continued to rely on commonsense insights. This is seldom the way of science. Earlier philosophers invented and used the then-most powerful methods of logic but most contemporary philosophers have largely avoided the more modern techniques of game theory.

[3]See Spiegelberg (1981). Also Husserl (1913/1967: 68–71) and Heidegger (1927/1962: Division One). Husserl's situatedness is more cognitive whereas Heidegger's encompasses the whole person.

[4]This interaction among persons, language, and world was described in what I called equilibrium metaphysics in Section 7.6 of *Language and Equilibrium*.

[5]Taylor (2013: 85–86) discusses this Heideggerian holism in the context of the evolution of modern epistemology.

As we move about, we encounter parts of this environment that we carve out as situations. From situations, we extract states of affairs or *infons*, informational complexes consisting of individuals having properties and standing in relations to other individuals. That is, situations are just more or less indeterminate collections of infons and there are often multiple situations within an agent's environment. There are further uniformities such as *types* and *parameters* that we abstract.[6] Perceptual information is transformed into cognitive information in our ongoing experience that keeps identifying and classifying new situations as we act in the world. Our sensory/perceptual awareness of situations tends to be analogical and our partial classification of them by infons digital.[7] In practice, we extract a fair bit of digital information from our surroundings but a great deal is omitted.[8] These entities – situations, infons, properties and relations, individuals, types, parameters, propositions, constraints – form our partly shared informational space.[9] Such information enters the science of communication in two ways, via the context of communication and via the content of communication.

[6]This book will not be dealing with such entities much, if at all, as I plan to use situation theory relatively informally here. They are described more fully in my previous book, *Language and Equilibrium*.

[7]Most classical Indian philosophies held that "perception must be of two kinds, each corresponding to a stage of its unfolding: at first a nonconceptual, nonlinguistic taking-in of whatever is presented to the senses, and then a conceptual, linguistic, predicative cognition in which the entities presented to the senses are knit together as qualifier and qualified." See Mohanty (2007: 31).

[8]During this conversion, there is always a *loss* of information because not all that is taken in is digitized, and language can handle only the digitized part. When we want to describe seemingly ineffable aspects of our thoughts – like the glimmer of an idea – we have to digitize some of our analog information. That is why *articulating* our ideas and experiences is seldom easy. Such a gap between word and world, possibly responsible for some of our sense of alienation, becomes most evident in the arts. Dance is relatively analog and literature is relatively digital, so when we try to capture our experience of dance in words, we sense correctly that we are losing a lot of information. That is why we say a picture is worth a thousand words.

[9]As Quine (1960; 1969a,b) has noted, the possible ontologies social groups can individuate are not unique. For example, Quine (1969b: 4–5) says: "English general and singular terms, identity, quantification, and the whole bag of ontological tricks may be correlated with elements of the native language in any of various mutually incompatible ways, each compatible with all possible linguistic data, and none preferable to another save as favored by a rationalization of the *native* language that is simple and natural to *us*." However, as I say in Parikh (2010: Section 7.6), it is not necessary to accept Quine's conclusions about *radical* indeterminacy because while many ontologies may be possible, only very few are *optimal*, as equilibrium processes eliminate all the suboptimal ones. The ontologies that remain are therefore likely to be identifiable and inter-translatable, and radical indeterminacy can be avoided. Similarly, the classical Buddhist doctrine of *dependent origination* maintains that reality is *conventional*, that is, socially constructed and nonunique, and becoming aware of this conventionality is the key to attaining wisdom. See Garfield (1995: 87–94) and also Mohanty (2000).

I generally use the word INFORMATION for the entire ontological space, including both true and false propositions and also entities that are neither true nor false. Dretske (1981; 2003; 2009) and Israel & Perry (1989) use it just for propositions that are true. Computer scientists, linguists, psychologists, and others use it for both true and false propositions. There is some warrant for both of these senses in ordinary usage, and my more inclusive meaning is a little nonstandard, but it is useful when accounting for titles of books such as *The Problems of Philosophy* or *Anna Karenina*, or unfinished utterances, or pictures of objects such as an apple or a ball, or similar uses of symbols that are "nonpropositional." My notion is coextensive with "content" but this word usually involves some representation whereas "information" need not.

2.1.2 Situation theory

The version of Barwise & Perry's (1983) situation theory I offer here is my own, its main innovation being partial infons. I use it relatively informally in this book so I give just an outline here.[10]

Consider an utterance of BILL SMITH RAN in some utterance situation u. To analyze it, we would need to represent the content of BILL SMITH, the content of RAN, the content of the whole sentence, and possibly other things. It is partial infons and full infons that enable us to do this.

Situation theory allows us to model partial information in a very fine-grained way. There are individuals a_i and n-ary relations R_i. Basic infons are $(n + 4)$-tuples $\langle\langle\, R;\ a_1;\ \ldots;\ a_n;\ l;\ t;\ 1 \,\rangle\rangle$ made up of individuals standing in relations at certain locations l and times t with the last item, the number 1, being its polarity, indicating the relation holds.[11] Partial infons such as $\langle\langle\, R;\ a_1;\ a_3 \,\rangle\rangle$ or even $\langle\langle\, a_1 \,\rangle\rangle$ are legitimate infons. Any arguments from the full infon $\langle\langle\, R;\ a_1;\ \ldots;\ a_n;\ l;\ t;\ 1 \,\rangle\rangle$ can be omitted. For example, the content *Bill Smith ran* can be expressed partially as $\langle\langle\, \mathit{ran};\ \mathit{Bill\ Smith} \,\rangle\rangle$ or more formally as $\langle\langle\, R^{\text{RAN}};\ b \,\rangle\rangle$ where R^{RAN} is a relation, b is an individual, and the location and time and polarity have been dropped. $\langle\langle\, \mathit{Bill\ Smith} \,\rangle\rangle$ or $\langle\langle\, b \,\rangle\rangle$ is also a partial rendition of this content and is just the

[10] The curious reader who wants to know more can refer to Parikh (2010). It is interesting to compare situation theory with semantic nets and studies by Sanskrit grammarians dating back to the first millennium BCE and up to the 18th century CE. As discussed by Briggs (1985), the latter two have a number of parallels. Situation theory is similar. For example, in an utterance of "John cooked the food and burned his mouth," Sanskrit schemata and semantic nets represent the missing information (that cooking involves heating food and heated food when eaten can burn a mouth) explicitly in the analysis of the utterance. Situation theory handles it in the same way.

[11] The polarity can also be 0 indicating the relation does not hold.

individual Bill Smith. This is the same as the identity between 7 and (7) in arithmetic. The angled brackets serve to gather the arguments.

Partial infons are not existentially quantified over the "missing" arguments because they are separate entities in their own right that can be merged with other appropriate partial infons in the right circumstances. In this sense, there are no missing arguments as such. More complex infons are formed from these basic infons via the operation of merging and they are all collected in the set $\mathcal{I}$. As I will not be much concerned with the internal structure of infons in this book, I will assume the operation of merging or unification parametrized by some situation s on $\mathcal{I}$ is given. This operation $\odot_s$, abbreviated to $\odot$, is neither associative nor commutative and has the identity **1** which stands for no information and zero **0** which stands for contradictory information. The product $\sigma \odot \tau$ is often written $\sigma\tau$.[12]

The relation between a situation s and an infon σ that holds in it is written $s \vDash \sigma$ or $\sigma \in s$, and is described by saying s supports σ or σ holds in s. The information expressed by $\vDash$ is special and $s \vDash \sigma$ is called an (Austinian) proposition. Only propositions can be true or false or indeterminate, the last in borderline cases involving vague terms as discussed in Chapter 11 or when the relevant infon is partial; infons by themselves do not admit of truth values. Utterances typically convey multiple propositions although these are usually multiple infons relative to a common described situation. I will show in Chapter 15 how propositions actually have a more general form because the infons in them may occur probabilistically.

Recall that a partial order over $\mathcal{I}$, the space of infons, is a binary relation $\Rightarrow_\ell$ over $\mathcal{I}$ which is reflexive, antisymmetric, and transitive; that is, for all σ, τ, and υ (i.e. the Greek letter upsilon) in $\mathcal{I}$, we have:

Reflexivity: $\sigma \Rightarrow_\ell \sigma$

Antisymmetry: If $\sigma \Rightarrow_\ell \tau$ and $\tau \Rightarrow_\ell \sigma$ then $\sigma = \tau$

Transitivity: If $\sigma \Rightarrow_\ell \tau$ and $\tau \Rightarrow_\ell \upsilon$ then $\sigma \Rightarrow_\ell \upsilon$

A partial order $\Rightarrow_\ell$ on $\mathcal{I}$ that captures the relation "is at least as informative as" or "is at least as strong as" is assumed. Certain infons are naturally more informative or stronger than others. For example, $\langle\langle\, P^{crimson};\ a \,\rangle\rangle \Rightarrow_\ell \langle\langle\, P^{red};\ a \,\rangle\rangle$ where a is some physical object because anything crimson is also always red. So the first infon is stronger than the second. Likewise, $\langle\langle\, P^{spinster};\ a \,\rangle\rangle \Rightarrow_\ell \langle\langle\, P^{female};\ a \,\rangle\rangle$

[12] The interested reader is referred to my previous book, Parikh (2010: Chapter 2), for more details about $\odot$.

where a now stands for a person. It is also true that $\langle\langle\, R;\ a;\ b\,\rangle\rangle \Rightarrow_\ell \langle\langle\, R;\ a\,\rangle\rangle$. If R is the relation of eating, then if a is eating b, a must be eating. Likewise, $\langle\langle\, R;\ a;\ 0\,\rangle\rangle \Rightarrow_\ell \langle\langle\, R;\ a;\ b;\ 0\,\rangle\rangle$ because if a is not eating, then a is not eating b. In each case, the infon on the left is more informative than the infon on the right.

Intuitively, it is clear that if we have two items of information, say, that a is red and b is blue, then it is possible to combine these states of affairs in two obvious ways, by conjoining or disjoining them. With this in mind, the partially ordered set $(\mathcal{I}, \Rightarrow_\ell)$ is further assumed to be a lattice, which is a partially ordered set in which every pair of elements has a unique supremum (called their join)[13] and a unique infimum (called their meet).[14] Let $\vee$ and $\wedge$ be the induced join and meet operations. If $\tau = \sup\{\sigma, \sigma'\}$, then $\tau = \sigma \vee \sigma'$, and if $\tau = \inf\{\sigma, \sigma'\}$, then $\tau = \sigma \wedge \sigma'$.

A lattice is complete if all of its subsets, finite or infinite, have both a join and a meet. There is no reason to restrict $\vee$ and $\wedge$ to finite subsets so we assume $(\mathcal{I}, \Rightarrow_\ell)$ is complete. The supremum of $(\mathcal{I}, \Rightarrow_\ell)$ is denoted by **1** and the infimum by **0**. Intuitively, **1** will hold in any situation because every situation supports "no information" vacuously, and **0** will not hold in any situation because no (coherent) situation can support contradictory information. **1** and **0** are just the identity and zero elements for $\odot$.

Since $\mathcal{I}$ now has the two binary operations $\vee$ and $\wedge$, we assume each distributes over the other. That is, it is assumed that $\sigma \wedge (\tau \vee \tau') = (\sigma \wedge \tau) \vee (\sigma \wedge \tau')$ and $\sigma \vee (\tau \wedge \tau') = (\sigma \vee \tau) \wedge (\sigma \vee \tau')$.

A valuation on $\mathcal{I}$ is a real-valued function $v : \mathcal{I} \longrightarrow \mathbb{R}$ such that $v(\sigma) + v(\tau) = v(\sigma \vee \tau) + v(\sigma \wedge \tau)$. A positive valuation is one where $\sigma \Rightarrow_\ell \tau$ implies $v(\sigma) < v(\tau)$. A metric lattice is a lattice with a positive valuation and the corresponding metric is given by:

$$\delta(\sigma, \tau) = v(\sigma \vee \tau) - v(\sigma \wedge \tau)$$

Valuations and therefore metrics always exist on distributive lattices and so one can define a metric δ on $\mathcal{I}$. Our interest will be in situated metrics that depend on some situation s. In other words, the valuation v the metric would correspond to would be a situated valuation. I will use this kind of metric in Chapter 21.

I will use the following concept in Chapter 19. A nonempty subset $\mathcal{F}$ of $\mathcal{I}$ is called a filter if

1. $\sigma, \tau \in \mathcal{F}$ implies $\sigma \wedge \tau \in \mathcal{F}$,
2. $\sigma \in \mathcal{F}$, $\tau \in \mathcal{I}$ and $\sigma \Rightarrow_\ell \tau$ imply $\tau \in \mathcal{F}$.

[13] The supremum or least upper bound of a pair of elements, if it exists, is the least element of $\mathcal{I}$ that is greater than or equal to each element of the pair.

[14] The infimum or greatest lower bound of a pair of elements, if it exists, is the greatest element of $\mathcal{I}$ that is less than or equal to each element of the pair.

For all situations s and all infons σ and τ, the following facts hold:

1. $s \not\models \mathbf{0}$ and $s \models \mathbf{1}$.
2. If $s \models \sigma$ and $\sigma \Rightarrow_{\ell} \tau$ then $s \models \tau$.
3. $s \models \sigma \wedge \tau$ if and only if $s \models \sigma$ and $s \models \tau$.
4. $s \models \sigma \vee \tau$ if and only if $s \models \sigma$ or $s \models \tau$.

Extensions are just the sets corresponding to properties or relations *relative* to some situation. Perhaps the most common type of extension is $\{x \mid s \models \langle\langle\, P;\ x \,\rangle\rangle\}$, which is the set of objects satisfying the property P in situation s. We can define $e(P, s)$ to be the individual a when the condition $s \models \langle\langle\, P;\ x \,\rangle\rangle$ yields one object a and the set $\{x \mid s \models \langle\langle\, P;\ x \,\rangle\rangle\}$ otherwise. This entity $e(P, s)$ occurs frequently in the study of noun phrases.[15] When *no* object has P in s, $e(P, s)$ will be the empty set. More formally:

(2) $$e(P, s) = \begin{cases} a & \text{if there is exactly one object } a \text{ having } P \text{ in } s \\ \{x \mid s \models \langle\langle\, P;\ x \,\rangle\rangle\} & \text{otherwise} \end{cases}$$

When one situation s (or situation type $\mathbf{s}$) *involves* another s' (or $\mathbf{s}'$), there is a constraint between them, written $s \Longrightarrow s'$ (or $\mathbf{s} \Longrightarrow \mathbf{s}'$). Constraints can be nomic, conventional, or of other types. They provide the mechanism through which agents perceive, infer, and act in the world and were introduced in Chapter 1 to account for meaning. Equilibrium Semantics can be compactly expressed as a system of constraints.

2.2 Agents

Our situated agents are finite in their capacities, have a range of concerns, and constantly face choices their environments make available. When concerns are articulated digitally as infons, they become *goals*. All agents have a complex and shifting hierarchy of concerns and goals, from survival at the top that is generally always present to very particular ones at the bottom such as a desire[16] for ice cream in some situation. An agent's goals can be equivalently expressed as preferences between situations, for example one in which he is eating ice cream and another in which he is not. It is convenient to use both goals and preferences

[15] See Chapter 26 and also Parikh (2010: Chapter 6).

[16] I use goal, desire, wish, purpose, and other synonyms interchangeably to relieve the tedium of repetition.

when discussing communication, remembering that we can translate between them.

When the choices afforded by a situation become articulate and explicit, they form, together with the agent's beliefs and preferences, a situated choice problem for the agent. When it involves just the agent by himself, it is a situated decision problem, and when it involves other agents, it is a situated game.

A situated game is a structure involving a set of situations along with a set of players, a set of interlinked choices of action for each player in alternative situations, and each player's preferences (or payoffs) for every combination of choices the players may select. That is, a situated game is a multi-person interactive situated decision problem. Being (partly) rational, agents try to do the best they can given how the other players may choose, and such a jointly optimal choice for each player is called an equilibrium.[17]

2.2.1 A simple example

Suppose two agents or players $\mathcal{A}$ and $\mathcal{B}$ are trying to meet in Manhattan without being able to communicate with each other first. Assume they could meet either at Grand Central Station which is closer to both of them or at Penn Station which is further away. If they go to different stations, neither benefits. Such a situation can be captured by the matrix in Figure 2.1.[18]

	Grand Central	Penn Station
Grand Central	(2,2)	(0,0)
Penn Station	(0,0)	(1,1)

Figure 2.1: A coordination game G in normal form

In the game G, $\mathcal{A}$ has the two choices indicated in the two rows, either to go to Grand Central or to Penn Station, and $\mathcal{B}$ has the same two choices indicated in the two columns. Their respective payoffs are mentioned in the four cells of the

[17] The words "best" and "optimal" are meant to be synonymous with "equilibrium." It is common to distinguish between an optimal choice and an equilibrium choice in game theory because an optimal choice, if evaluated in the absence of what other players may do, can diverge from the equilibrium. I just mean the best a player can do *given* the other players' (best) choices.

[18] Almost this very game was first considered by Schelling (1960). Throughout, I have used "he" for $\mathcal{A}$ and "she" for $\mathcal{B}$ as two-person game theory naturally lends itself to two pronouns, making it easier to differentiate between the agents.

matrix, the first number being $\mathcal{A}$'s payoff and the second being $\mathcal{B}$'s in each cell. If both agents go to Grand Central they get a payoff of 2 units, if they both go to Penn Station they get 1 unit, and if they end up at different stations they get 0. Both $\mathcal{A}$ and $\mathcal{B}$ are rational and prefer more payoff units to less but each of them can only select their own action even though the outcome depends on what they both do. As we will see later, this is exactly the situation with communication where the outcome depends on both the speaker and the addressee.

Thus, both players have to choose a course of action based on what the other player will choose. This sort of interactive choice structure represented as a payoff matrix is called a game in normal or strategic form. This particular game is also called a coordination game because both agents have compatible payoffs and there is no conflict, that is, they do not value the same outcome differently. By varying the payoffs it is possible to generate a range of games even in this simple two-player, two-choice setting. In general, there can be more than two players and more than two choices of action for each player. In some sense, G is one of the simplest nontrivial games where some interesting interactive phenomena occur.

Once a game like G is set up the next step is to see how rational agents wanting to choose the best action would act given that the other agents want to do the same. The resulting optimal strategies are called the solution to the game. Studying the solution process formally involves a number of definitions of terms like *strategy*, *equilibrium*, and the like as well as somewhat subtle analyses.

The key idea behind one prominent kind of solution is that optimal actions should be such that no agent will want to deviate from them unilaterally. If $\mathcal{A}$ were to choose Grand Central then it is optimal for $\mathcal{B}$ to choose the same and vice versa. If $\mathcal{A}$ were to choose Penn Station then it is optimal for $\mathcal{B}$ to choose the same and vice versa. In other words, both $(\text{Grand Central}, \text{Grand Central}) = (\text{GC}, \text{GC})$ and $(\text{Penn Station}, \text{Penn Station}) = (\text{PS}, \text{PS})$ are pairs of actions that neither agent will want to deviate from unilaterally. They possess a kind of stability. On the other hand, both (GC, PS) and (PS, GC) are, in this sense, precarious pairs of choices because both agents will want to shift their strategy. Each can do better by a unilateral change to a different action. For example, with (GC, PS) $\mathcal{A}$ would benefit by shifting to Penn Station because $\mathcal{A}$ would then receive 1 instead of 0 and, likewise, $\mathcal{B}$ would benefit by shifting to Grand Central because $\mathcal{B}$ would then receive 2 instead of 0, assuming the two agents do not both shift simultaneously. If any single agent can do better by a unilateral change to a different strategy then that pair would not be selected as optimal. A pair of strategies that is immune to any such unilateral deviation by any agent is called a Nash equilibrium.[19]

[19]See Nash (1951), Myerson (1991: Chapter 3), Watson (2002: Chapter 9) or Parikh (2010: 78).

(GC, GC) and (PS, PS) are Nash equilibria and such solutions always exist in a large class of games but they are often not unique. This requires ways to eliminate certain equilibria that are counterintuitive from a commonsense viewpoint. In G, both agents would be better off by selecting (GC, GC) rather than (PS, PS) because they would both receive higher payoffs, (2, 2) rather than (1, 1). Such Nash equilibria that make at least one player better off without making any other player worse off are called Pareto-Nash equilibria.[20]

I will use both these equilibria extensively in the games I construct to model communication. The optimal utterances and interpretations of speakers and addressees form Pareto-Nash equilibria although the details are more complex as the games required to understand communication are subtler than G.

A different way to express G is via the so-called extensive form where each agent's actions are laid out sequentially in an interactive decision tree as shown in Figure 2.2.

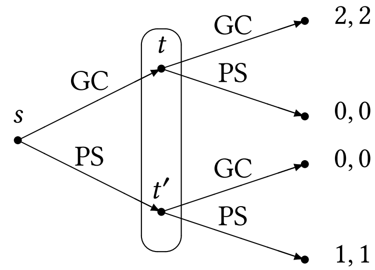


Figure 2.2: Extensive form for G

The nodes labeled s, t, and t' are situations and the idea is that an initial situation consisting of various infons results in some action by $\mathcal{A}$ and leads to a new situation which is followed by a further action by $\mathcal{B}$ leading to a third situation where payoffs are distributed to the two agents. So, for example, s may result in $\mathcal{A}$ going to Grand Central leading to a new situation t and then $\mathcal{B}$ may also go to Grand Central leading to a third unlabeled situation where both agents receive a payoff of 2. And similarly for the other three paths through the game tree. Figures 2.1 and 2.2 capture the same choices and payoffs.

The new element in Figure 2.2 is the oval enclosing the situations t and t'. This oval is called an information set and represents the epistemic state of $\mathcal{B}$ after

[20]See Parikh (1987b; 2001; 2010).

$\mathcal{A}$ has chosen his action, Grand Central or Penn Station. Because we assumed the two agents cannot communicate their choices, $\mathcal{B}$ does not know what $\mathcal{A}$ chose to do. She cannot distinguish between t and t'. This translates into two requirements: there must be the same choices available to $\mathcal{B}$ at both t and t' because otherwise she could tell the two situations apart, and she must make the same choice in t and t' because she is in the dark about where she is in the information set.[21] Every situation where a choice has to be made by either agent belongs to an information set. The oval represents $\mathcal{B}$'s information set. $\mathcal{A}$'s information set is trivial because he has just one choice situation s and it alone belongs to an information set which we do not bother to identify.

Consider Figure 2.3 where a slightly different game G' is shown with the same choices and payoffs as G but with different epistemic properties.

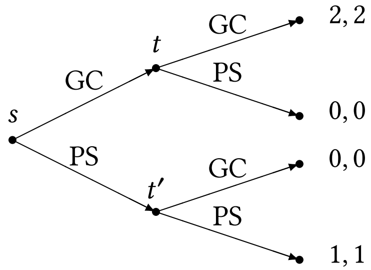


Figure 2.3: Extensive form for G'

In G' there is no oval which means $\mathcal{A}$'s choice of action is communicated to $\mathcal{B}$ and so she can distinguish between t and t' and make different choices in each of the two situations. Differently put, $\mathcal{B}$ now has two information sets, one containing t and the other containing t'. If $\mathcal{A}$ chooses Grand Central, $\mathcal{B}$ would find herself in t and knows this, and so can choose Grand Central, and if $\mathcal{A}$ chooses Penn Station, $\mathcal{B}$ would find herself in t' and again knows this, and so can choose Penn Station. It is an easier game for both agents to solve as they don't need to anticipate each other's choices. This is one of the ubiquitous reasons communication is so useful because society consists of a large number of games and communication makes them easier to solve. But communication itself involves

[21]There is a third requirement involving one situation in an information set not preceding another but I will ignore it as this possibility does not arise in the two-stage games we will be considering.

solving various games and so must be addressed first. G and G' are constructed from situations and so are situated games, structured sets of situations as I said above.[22]

2.2.2 Common knowledge

We have seen how the extensive form makes certain internal epistemic constraints on the agents explicit via information sets. Once knowledge and belief enter the scene, it becomes evident that both agents need to at least know the whole game G itself. In fact, they need to have *shared* knowledge of the choices and payoffs and of the interactive structure. The particular kind of shared knowledge required is called common knowledge.[23] It is interesting that this notion first arose outside of game theory in the context of communication itself, showing indirectly that there is some intimate link between games and communication.

There are two approaches to characterizing common knowledge. The orthodox idea, originating in Schiffer (1972) and standardly used in game theory following Aumann (1976), is that common knowledge of a fact σ is iterated knowledge of σ: $\mathcal{A}$ knows σ, $\mathcal{B}$ knows σ, $\mathcal{A}$ knows $\mathcal{B}$ knows σ, $\mathcal{B}$ knows $\mathcal{A}$ knows σ, $\mathcal{A}$ knows $\mathcal{B}$ knows $\mathcal{A}$ knows σ, and so on. This remains the mainstream approach. The shared situation approach, originating in Lewis (1969: Chapter II, Section 1) and made serviceable by Barwise (1988; 1989a), is as follows: $\mathcal{A}$ and $\mathcal{B}$ have common knowledge of σ just in case there is a situation s such that (a) $s \models \sigma$, (b) $s \models \mathcal{A}$ knows s, and (c) $s \models \mathcal{B}$ knows s. Keep in mind that "$\models$" can be read simply as *supports* or *contains* and "$\mathcal{A}$ knows s" as $\mathcal{A}$ *knows all the facts in* s. When a situation is relatively limited, it is quite common for an agent in that situation to know all the facts in it and thereby to know the situation itself.

Gintis (2000: 14) and others have argued against the possibility of common knowledge of the infinitely iterated kind for finite agents because it involves an infinite mental representation. I agree with them but when two ordinary human interlocutors are copresent, their copresence is surely common knowledge between them, and so this can only be explained by the shared situation approach, which involves only finite structures. If $\mathcal{A}$ knows s, then $\mathcal{A}$ knows each fact in s, and so $\mathcal{A}$ knows σ and $\mathcal{A}$ knows $\mathcal{B}$ knows s. This means therefore that $\mathcal{A}$ knows $\mathcal{B}$ knows σ. If this is continued indefinitely, it follows that the situational concept

[22]Watson (2002) is an introductory text on game theory. Also see the Appendix for more on situated games.

[23]The interested reader can see Parikh (2001: Section 5.3) where I discuss the motivation for common knowledge in detail.

implies the infinite chains of knowledge in the orthodox concept and so satisfies the intuitive requirement for common knowledge without its explicit mental representation. As Lewis (1969: 53) has said, "Note that this is a chain of implications, not of steps in anyone's actual reasoning. Therefore there is nothing improper about its infinite length." So common knowledge can be made plausible for finite agents.[24]

Common knowledge can be either conscious, nonconscious, or semiconscious. It typically arises from perceptual and other situations we share with others and then is maintained by our ongoing actions and communication. When $\mathcal{A}$ communicates something to $\mathcal{B}$ on the basis of some initial common knowledge, then the content communicated and related facts also become common knowledge and the process continues with an expanded base. Part of such growing common knowledge is retained and becomes their shared background information which is no longer directly perceptual. As members of society, agents can count on such shared information as part of their common knowledge.[25]

When common knowledge is used later, σ will be substituted by, say, g for the situated game under consideration. The agents will be said to have common knowledge of g as a situated game is a structured set of situations and situations are collections of infons which makes a situated game a structured collection of infons. This structure can itself be expressed via infons because infons involve entities standing in relations. Thus, a situated game is just a large collection of infons and so agents having common knowledge of situated games is the same as agents having common knowledge of all the infons comprising them.

[24]There is a mild cost to the shared situation approach. It requires the assumption of circular or non-well-founded situations where a situation may contain itself as a constituent. See Barwise & Etchemendy (1986). See also Fagin et al. (2003: Section 11.5) for a fixed-point approach to common knowledge. And see Clark & Marshall (1981) for the iterated approach.

[25]This has been a basic leitmotif of all my work. See, for example, Parikh (2001: Section 6.4). It has also been a fundamental idea in dynamic semantics as represented in, for example, Kamp (1981), Heim (1982), and Gronendijk & Stokhof (1991). However, as I pointed out in Parikh (2010: 5), "most of these developments remain squarely within the tradition of Montague-inspired formal semantics where the focus is on finding appropriate meaning representations rather than on *deriving* intended and optimal meanings through use. Discourse representation theory, file change semantics, and dynamic logic are concerned more with the *results* of the communicative process than with communication itself, with the *what* rather than with the *how*. They address what Austin (1961/1979b) called the perlocutionary act and effects of communication, not the locutionary and illocutionary acts and the securing of uptake and understanding. As such, they do not appear to question the *syntax-semantics-pragmatics* trichotomy and pipeline view of meaning bequeathed by Morris (1938) and Grice (1989c) despite their undoubted technical accomplishments." Also, the notion these accounts use is that of common ground which is a little different from common knowledge. See Stalnaker (2002).

I have defined common knowledge in terms of knowledge so a quick word about knowledge is in order. It is best explained as informationally caused belief, and belief as information carried in "completely" digitized form, as argued in Dretske (1981: 86, Chapters 7 and 8).[26]

2.2.3 Context

Our being situated agents implies that there is always a context when we communicate. A core aspect of this context, called an utterance situation, plays a profound role in communication, both in particular instances and in the large-scale evolution of language. It is seldom precisely specified or specifiable as the boundaries of situations are generally indeterminate. It contains all the ambient information that agents can draw upon in deciding what to say and inferring what has been communicated. There are other situations that are also part of the context such as a described situation, multiple resource situations, and a discourse situation made up of a sequence of utterance situations, but I will not need them much here. I will also often use "utterance situation" and "context" interchangeably. A central part of an utterance situation is a certain set of games that will be introduced in the next chapter.

I have given a brief description of our informational space and situation theory and, based on this, of agents and games and common knowledge and utterance situations. With these preliminaries, we are now ready to take a synoptic look at communication.

[26] In this sentence, the first use of "information" pertaining to knowledge stands for factual information and the second use of "information" pertaining to belief stands for propositional information that may be false.

3 A picture of communication

Communication in the small and in the large is a roughly *circular* process in a manner reminiscent of Dilthey's famous hermeneutic circle. In the small, I have shown in *Language and Equilibrium* how the interdependence of speaker and addressee as well as the mechanism of conveying and interpreting meaning make communication circular. In this book, I develop this theme further. In the large, this circularity is amplified because communication in society is ongoing and because such societal communication consists of smaller, interlocking conversations. That is, the conversations are linked with one another via the shared conventional meanings of words which emerge, in turn, from these conversations. The smaller model focuses on dialogues between two (or more) participants and describes what I call *micro-semantics*; the larger model focuses on several such interconnected micro-semantic models and describes what I call *macro-semantics*.

3.1 Micro-semantics

I start by showing a snapshot of the intricate process that occurs in micro-semantics in an utterance situation:

Utterance Situation

Setting Game

$\longrightarrow$ $\mathcal{A}$'s wish to elicit some response from $\mathcal{B}$

$\longrightarrow$ Content Selection Game

$\longrightarrow$ $\mathcal{A}$'s equilibrium content

$\longrightarrow$ Generation Game

$\longrightarrow$ $\mathcal{A}$'s equilibrium utterance

$\longrightarrow$ Interpretation Game

$\longrightarrow$ $\mathcal{B}$'s equilibrium content

$\longrightarrow$ Content Selection Game

→ $\mathcal{B}$'s equilibrium response

→ Back to the Setting Game

I now proceed to explain these terms in more detail.

3.1.1 The Setting Game

A widely prevalent picture of communication starts with something the speaker wishes to convey. This is only partly correct. A more accurate starting point is that the speaker and addressee are involved in some interaction and this utterance situation or *setting* induces the speaker to affect the addressee either by conveying information to her and triggering a corresponding belief or by eliciting information from her or, more generally, by getting her to act or react in some way. It may happen that the speaker does not care about the addressee's response but even here the conveying and reception of information is still desired.

This setting will generally be a game or decision problem of some kind. For example, $\mathcal{A}$ and $\mathcal{B}$ may be discussing where to eat and $\mathcal{A}$ may wish to influence $\mathcal{B}$'s choice. Such a situation, involving multiple options and involving possibly differing preferences for them, may take many forms, one of which may be the agonistically dubbed game *Battle of the Sexes.*[1] Its normal form typically looks as shown in Figure 3.1.

	Restaurant one	Restaurant two
Restaurant one	(3,2)	(0,0)
Restaurant two	(0,0)	(2,3)

Figure 3.1: A Battle of the Sexes game in normal form

The payoffs in such a Setting Game reflect the explicit or implicit goals and subgoals of the agents. Someone may have said something earlier like "Why don't we eat out tonight?" which makes their joint goal explicit. Or the goal may have evolved implicitly. These goals are similar to what Roberts (1996) calls "domain goals." They may be either cooperative or conflictual.

The Setting Game is crucial for a theory of communication because it contains information, especially about the agents' goals, some of it shared, that constrains

[1]Practically every textbook on game theory considers this game, which is almost as widely known as the Prisoner's Dilemma.

the future actions of both participants and helps in explaining them. There is often a hierarchy of goals in the utterance situation. For example, their joint goal may be to eat out but this leads to $\mathcal{A}$'s subgoal to influence $\mathcal{B}$'s choice of restaurant based on the payoffs in the Setting Game which reflect their preferences for different restaurants.

The Setting Game takes place in the utterance situation and results in $\mathcal{A}$'s wish to elicit some response from $\mathcal{B}$ as indicated in the snapshot above. This desire may lead $\mathcal{A}$ to select some content to convey to $\mathcal{B}$.

3.1.2 The Content Selection Game

Based on how $\mathcal{A}$ wishes to affect $\mathcal{B}$,[2] he considers the suitability of various *contents—not* utterances yet. As human beings are only partly rational, a speaker may not deliberate on all the relevant choices, only a few or possibly just one. $\mathcal{A}$'s pondering possible contents and the responses each would elicit from $\mathcal{B}$ results in a kind of game that seems superficially like what is called a *signaling* game.[3] But it will soon become clear why Content Selection Games are in fact very different from signaling games. For the moment, notice that in the latter the first player, the so-called sender, has to choose among a set of possible signals he can send. Signals are just utterances but in the Content Selection Game the sender has to decide among a set of *contents*, so while our game may *look* similar to a signaling game, it is different in substance.

[2] Other "external" factors such as the presence of eavesdroppers may also constrain $\mathcal{A}$'s choices.

[3] A signaling game is a particular kind of game among many different types of games studied in game theory. It starts with a random move by "Nature," that is, a kind of abstract device like the flip of a coin, which results in the first player, the Sender, acquiring some private information that the other player, the Receiver, does not know. The task of the Sender is to convey this private information to the Receiver by choosing one of a set of signals or utterances. The task of the Receiver is to take some action that affects both players based on the signal received. The first signaling games were invented by Lewis (1969). They were later studied by Spence (1973), Crawford & Sobel (1982), and Kreps (1986).

Some in the game-theoretic linguistics community combine features of signaling games with my former partial information games. For example, David Lewis was interested in the emergence of meaning and so he used initially meaningless utterances. Those writing since 2003, such as Robert van Rooij, Gerhard Jäger, and Anton Benz, were analyzing implicature and so could not assume initially meaningless utterances. So they adopted the idea from my games that the speaker's intentions constrain the range of sentences available, and related assumptions. I have compared signaling games and my games in several places and so I discuss such issues briefly in this book. The important point is not how different my games are from signaling games but whether the substantive structures employed are useful to compare directly or not. Besides, these theorists employ an orthodox Gricean framework, something I criticize in detail in Chapter 5. I hope readers from this community will find it useful to relate the substantive structures and ideas I offer to their frameworks.

I will call this structure the *Content Selection Game* because in it the primary task for the speaker is to select the best content from his choices and the primary task for the addressee is to select her best response from her choices. Concretely, if $\mathcal{A}$ wants to influence $\mathcal{B}$'s choice of a restaurant, he may have to choose between a content like *should we go to Bar Boulud?* or like *should we go to Bar Boulud? we haven't been there in a while*. The latter are *contents*, not sentences, and italics will be used to refer to particular contents throughout.[4] Once $\mathcal{B}$ learns what $\mathcal{A}$ wishes to convey, she would have a variety of choices, ranging from a simple *yes* or *no* to other possibilities such as *how about Bouley instead?* Just as the addressee cannot fully anticipate what the speaker might convey, the speaker cannot fully anticipate how the addressee might respond. So this game is typically one with unaware players without common knowledge of the game.[5] Despite this, I will assume common knowledge for such games as it is simpler to do so.

$\mathcal{A}$ then solves the Content Selection Game and identifies its equilibrium content. This search for his optimal action can be a tricky affair because this game can be conflictual. The interests of the two parties may not be perfectly aligned.

The way the addressee's choice influences the choice of content (and vice versa) has seldom been considered explicitly in semantics. Theories always start with a sentence the speaker has uttered as given. It is by attending to this game in the context of the whole communicative process that it becomes doubtful if the Gricean (1975) maxims work as they are meant to. A key reason is the presence of conflict; another is that in addition to conveying information, a speaker typically wishes to affect his relationship with the addressee in some way, either maintaining or altering it via the *effect* the utterance has on her.[6]

Relating the foregoing to the snapshot of communication above, the Content Selection Game arises from $\mathcal{A}$'s desire to evoke a response from $\mathcal{B}$ and results in the selection of an equilibrium content to convey to $\mathcal{B}$.

3.1.3 The Generation Game and the Interpretation Game

If telepathy were possible, $\mathcal{A}$ would directly transmit this choice to $\mathcal{B}$ and that would be the end of his effort. Unfortunately, he now has to find the right words to convey his chosen content. This involves considering different combinations of words based on their conventional meanings. For example, two options he may contemplate are SHOULD WE GO TO BAR BOULUD? and SHOULD WE GO TO BAR

[4] One can also use the infons of situation theory to represent such contents.

[5] See Parikh (2001: Sections 5.3 and 6.5) and, for a formal account, Halpern & Rego (2006; 2007).

[6] Asher & Lascarides (2013) address somewhat similar issues.

BOULUD? WE HAVEN'T BEEN IN A WHILE. Throughout, small capitals or quotation marks will be used for expressions.

Stepping back momentarily, $\mathcal{A}$ and $\mathcal{B}$ will generally have a language $\mathcal{L}$ they share. Sharing a language is a complex matter. For now, assume it is the linguistic community's nonconscious common knowledge of the conventional meanings of the words of the language as well as of its grammar and phonetic system. So $\mathcal{A}$ will look for a sentence or two in $\mathcal{L}$ that will express the content he has identified given the utterance situation he is in. If $\mathcal{L}$ is a very simple (artificially constructed) language that has neither ambiguous nor structured expressions, his task will be much simpler in one sense although an overly simple language may not be sufficiently rich for all that he wants to convey. When $\mathcal{L}$ is a full-blown natural language like English or Gujarati, things get a bit complicated because there are three dimensions at play simultaneously. One is the conventional meanings of the words, a second is the internal structure of the sentence, that is, their syntax, and the third is their phonetic properties. Assuming $\mathcal{A}$ is speaking and not writing, the content of a possible utterance will be determined by all three dimensions.

Since languages can be simple or complex, it should be evident that syntax and phonology are merely a wrinkle, a complicating factor – like friction in Newtonian mechanics – in the primary process of communication. The presence of syntax is highly interesting but contingent. It is the meaning (and effect) of utterances that are central, not their structure or phonetic patterns. Yet, linguists have focused primarily on these latter aspects and this is perhaps one reason why the science of communication has languished, including an account of how the *optimal* parse or set of words in an utterance is conveyed.

At a simplified level, $\mathcal{A}$ thinks about uttering a sentence such as SHOULD WE GO TO BAR BOULUD? WE HAVEN'T BEEN IN A WHILE by mentally trying out how $\mathcal{B}$ would interpret it if it were uttered. This interpretation process is part of his decision-making and is a game of partial information.[7] In other words, $\mathcal{A}$, so to speak, lines up alternative utterances, and then imagines the games of partial information each of them leads to, and, based on how $\mathcal{B}$ would interpret each utterance, he selects the utterance that will best convey the content he has identified in the Content Selection Game. This complex structure is called the *Generation Game* because in it $\mathcal{A}$ generates his best choice of utterance.

This is a simplified description of generation because I have made $\mathcal{A}$ consider whole sentences: I have conveniently omitted how the speaker assembles the

[7] These are games that were first developed in my doctoral dissertation *Language and Strategic Inference* (1987b) at Stanford University in the context of modeling communication and meaning. They have been elaborated a great deal since then.

parts of possible sentences to create whole utterances and also how the speaker decides how to pronounce the sounds corresponding to the words. I will say more about this later but it is a part of the model of communication that remains to be worked out fully.

We can now see fully why the Content Selection Game is not really a signaling game at all. Since telepathy is not possible, $\mathcal{A}$ cannot convey his chosen content to $\mathcal{B}$ without playing the Generation Game. So the Content Selection Game is *linked* to the Generation Game. And we will see below that the Generation Game is itself linked to another game because it isn't complete either. These linkages, together with the fact that there are no signals in it, make the Content Selection Game very different from a standard signaling game.

The games of partial information that are embedded in the Generation Game are called speaker games because they are the speaker's model of the linguistic interaction between $\mathcal{A}$ and $\mathcal{B}$. The speaker considers one set of speaker games for every utterance in the Generation Game but once he has selected his optimal utterance, he utters it, and this induces a corresponding set of partial information games on $\mathcal{B}$'s side, called addressee games. These are $\mathcal{B}$'s model of their linguistic interaction. $\mathcal{A}$ has multiple sets of speaker games to solve whereas $\mathcal{B}$ has just one set because $\mathcal{A}$ utters just the one best sentence from his choices. She has no idea about the other locutions $\mathcal{A}$ turns over in his mind. $\mathcal{B}$'s single set of addressee games is best referred to as the *Interpretation Game*. These games, too, have the three dimensions of meaning, structure, and sound.

As the Generation Game and Interpretation Game are $\mathcal{A}$'s and $\mathcal{B}$'s subjective models of their interaction, there is also a structure that is their *actual* objective interaction and this is also a game. In other words, once $\mathcal{A}$ utters a sentence publicly, the objective game comes into being along with $\mathcal{B}$'s model of it. For those who might find this third objective structure odd, think first of a single agent and their subjective representation of some part of the world in which they have to act. In this simpler scenario, there is the single agent's mental representation and an objective situation. An agent may look at the dark clouds in the sky and decide to carry an umbrella. When explaining their behavior and its outcome, a theorist needs access to both the actual state of the sky and the agent's mental representation of it. When communication is considered, there are two agents so there are two subjective representations along with the objective situation. In theorizing about communication, we need access to all three structures to understand the most general cases.

The Generation Game and Interpretation Game as well as the third objective game involve *sets* of games because there are games corresponding to each word, each phrase, and the whole sentence on both sides for each sentence considered.

So, in the example sentence SHOULD WE GO TO BAR BOULUD? there would be as many as five lexical games just for the semantic aspect of the interpretation. There would also be syntactic games and phonetic games. All of this will be dealt with in detail in Part III.

When the ideal conditions for communication are met, the speaker games in the Generation Game are identical to the addressee games in the Interpretation Game and both are also identical to the relevant objective games and the interlocutors have common knowledge of all these games. This will also become clearer in Part III. After $\mathcal{A}$ has completed his utterance and the Interpretation Game and objective game have emerged, it is $\mathcal{B}$'s turn to act. She solves the game to find the best interpretation of the utterance. If all goes well, she is now in possession of the content $\mathcal{A}$ wanted to communicate to her and she then has to choose her best response to this content in the Content Selection Game.

To sum up, once $\mathcal{A}$ has selected the optimal content he wants to convey to $\mathcal{B}$, he considers the Generation Game which enables him to convert this content into an equilibrium utterance for him to produce in the utterance situation. Reciprocally, once $\mathcal{A}$ utters his optimal sentence, $\mathcal{B}$ plays the Interpretation Game and attempts to infer the equilibrium content that $\mathcal{A}$ is conveying. Based on this inference, she now selects her best response in the Content Selection Game. And then both players resume the Setting Game.

3.1.4 The Communication Game

I once again display the process of communication schematically to help fix the sequence:

Utterance Situation

Setting Game

$\longrightarrow$ $\mathcal{A}$'s wish to elicit some response from $\mathcal{B}$

$\longrightarrow$ Content Selection Game

$\longrightarrow$ $\mathcal{A}$'s equilibrium content

$\longrightarrow$ Generation Game

$\longrightarrow$ $\mathcal{A}$'s equilibrium utterance

$\longrightarrow$ Interpretation Game

$\longrightarrow$ $\mathcal{B}$'s equilibrium content

$\longrightarrow$ Content Selection Game

→ $\mathcal{B}$'s equilibrium response

→ Back to the Setting Game

Everything starting and ending with the Setting Game occurs inside the utterance situation. It is unfortunate that there are so many different games and so much terminology, but the reader will immediately see that communication does in fact involve all of these steps and will also see the interdependence among these games. Very simply, the interlocutors are trying to decide where to eat and so $\mathcal{A}$ chooses to make a suggestion to $\mathcal{B}$ to which $\mathcal{B}$ chooses a response after which they return to their decision-making. This simple set of acts has the rather complex structure sketched above. In reality, perhaps the whole process happens in a more seamless way, but it is helpful to separate its parts analytically to understand the process better.

Further, even this quite complex picture is a simplification because the clean distinction between $\mathcal{A}$'s content and utterance selection embodied in the Content Selection Game and the Generation Game is often an idealization. A poet may care more about a word and its sound than its meaning and want the former to affect the addressee more than the latter and so may not first identify a meaning and then a word but in the reverse order or even together. Or they may struggle to articulate an inchoate thought and realize what their meaning is only through their words. So, in the most general picture, the model would merge some of these parts, especially the Content Selection Game and the Generation Game, possibly by considering ordered pairs of contents and sentences. We will keep them separate, however, for clarity and to simplify the model.

As we have seen, there are four large *interlocking* games:

1. Setting Game
2. Content Selection Game
3. Generation Game
4. Interpretation Game

I have built upon an innovation in my previous book involving separate but interdependent games. In that book, the linked games considered were just the games of partial information on the speaker's and addressee's sides. Here, the idea has been extended to the whole communicative process. I call the entire structure comprising these four games the *Communication Game* because it really is a single structure with interdependent parts. The Setting Game was described in some detail in *Language and Equilibrium* (2010: Chapter 3) so I will treat it summarily here. The other three games will be discussed in Parts III and IV.

Ever since the founding of modern semantics by Frege in the late nineteenth century, communication has been understood in a one-sided way, focused primarily on the perspective of the speaker. Arguably, the role of communication in semiosis was left implicit at best. It was the ordinary language philosophers such as the later Wittgenstein (1953/1968), Austin (1961; 1975), and Grice (1989c) who brought it to the fore but, though they grasped some of the complexities of the two-sided nature of the interaction, they appear to have missed the full structure articulated above. Their followers have seldom grappled with this ampler vision of communication and have confined themselves to elaborations and revisions of the original proposals.

In particular, all of Grice's ideas – communication as a form of rational activity, the conversational maxims, the analyses of speaker meaning and word meaning, and the distinction between a largely conventional semantics and an inferential pragmatics – fall short. Part of his goal was to naturalize "intentionality" or the aboutness of language by reducing linguistic meaning to speaker meaning and the latter to speaker intentions. Intentions can then presumably themselves be reduced to a digital notion of information as Dretske (1981) has suggested or to an agent's actions as Stalnaker (1984) favors. This led Grice to analyze speaker meaning but he missed other key concepts in the process of communication and the concomitant emergence of meaning. I look closely at some of the limitations of Gricean ideas in Chapter 5.

This completes my picture of micro-semantics. It is not difficult to extend it from a single utterance to discourse and dialogue but I will not pursue these natural extensions here.[8]

A key assumption of micro-semantics is that the conventional meanings of words[9] are fixed and given and are exploited by speakers and addressees in com-

[8]See Clark & Parikh (2007) for a preliminary account of discourse anaphora.

[9]In my framework, only words have conventional meanings. Longer expressions such as phrases and whole sentences do not, although conventional meanings can certainly be assigned to them if required for other reasons. The contents conveyed by utterances of longer expressions can be computed by using just the conventional meanings of words. Of course, idiomatic expressions (e.g. KICKED THE BUCKET) and such multi-word expressions that are not compositional would be treated differently but I do not consider such cases in this book. There are also sentences like CAN YOU PASS THE SALT? that involve conventional interpretations. I treat such situations as instances of modulation – see Chapter 17.

municating referential meanings, much as Fregean senses are used to convey references,[10] the former being generalized and refined versions of the latter.[11] This makes it possible to give a fairly complete account of communication and meaning in the small. Nothing else needs to be assumed besides the partial rationality of agents,[12] a language and its grammar, and an ontology. But this leaves unfinished the explanation of how semiosis gets off the ground, how language acquires meaning in the first place, the main problem of semantics. Conventional meaning itself must emerge through communication because there is nothing else, no other possible source. To explain this origination, it becomes necessary to shift from a single utterance to society-wide conversations, from micro-semantics to macro-semantics.

3.2 Macro-semantics

In macro-semantics, the assumption that conventional meaning is fixed and *externally* given is relaxed and other *cross*-conversational constraints are identified. In other words, it isn't possible to account for semiosis by looking at just one utterance. *Multiple* exchanges are required, either between the same two interlocutors over time or within a whole community. One can abstract from time and assume each member of a population is engaged in multiple conversations with other members synchronically. Each such utterance is described by a Communication Game. But now, instead of using a fixed set of conventional meanings for each word, all we ask is that *whatever* the conventional meanings may turn out to be, they are *consistent* across users. The same agent will not use a word with different conventional meanings every time they speak or interpret. In other words, the constraint of consistency entails that both parties to a conversation use the same conventional meanings not only in the exchange at hand but in *all* their communications. This severely limits the range of conventional meanings a word can acquire and makes it possible to derive not just referential contents via the Com-

[10] See Frege (1892/1980) and Section 1.2.

[11] While this will become fully clear in Part III, it might help readers who think of conventional meanings differently to see the following schema: word $\longrightarrow$ conventional meaning $\xrightarrow{u}$ referential meaning. Only the referential meanings of words enter into the content of the whole sentence uttered in the utterance situation u. Conventional meanings are a steppingstone to referential meanings and are more or less shared by the interlocutors.

[12] I discuss informally what I mean by "partial rationality" in Chapter 5. No clear definition of this notion has yet emerged and so I use it to indicate various sets of assumptions that fall short of full rationality. Other similar terms are "bounded rationality," "resource-limited rationality," and "behavioral rationality."

munication Game but also conventional meanings for each of the words. I will assume a toy language of one-word sentences as that sidesteps syntax, and assume there is no ambiguity, that is, there is one conventional meaning per word. Such a model is enough to demonstrate the basic idea behind macro-semantics. Working out the mathematical details of such a *meaningful* equilibrium for language presupposes a solution to the micro-semantic problem. An outline of such a model exists in Parikh (1987a) and I introduce a more complete version in Part V. The society-wide collection of interlocking Communication Games, linked via the Consistency Condition on conventional meanings, is called a *Language Game*, a highly suggestive but vague metaphorical idea of Wittgenstein's (1953/1968) that is given a precise definition here. There can be many different kinds of Language Games based on the kinds of networks that are possible in society. This suggests that Lewis's (1969) account of convention captures conditions that are sufficient but not necessary. That is, my network model broadens Lewis's more restrictive notion of convention as shown in Chapter 23.

Many generalizations of this picture are possible, not least introducing syntax and allowing ambiguity. These steps are difficult since ambiguity involves not just multiple conventional meanings per word but other complexities such as what is called modulation.[13] Another generalization is to note that people seldom share exactly the same conventional meanings so the notion of consistency has to be weakened. This also means the description of what it is to share a language in Section 3.1 has to be relaxed. Common knowledge of conventional meanings is too stringent. We can make the meaningful equilibrium dynamic by making utterances sequential rather than simultaneous to study meaning change. Another direction is to *derive* syntax and phonology from communication. These are all theoretical ideas and could be hard to work out, especially in computationally effective ways that enable empirically testable hypotheses to be developed.

The conventional meanings of a single word such as "dog" result from countless conversations and countless conversations are affected by these conventional meanings in a diachronic way. Indeed, it isn't clear how stable conventional meanings really are because it is through these very encounters that they also *change*, as becomes evident if one traces the word's evolution from the Old and Middle English "dogge" to its modern form and meanings. I address meaning change in Chapter 24.

Such macro-semantic models allow us to answer a question that is rarely posed in semantics. It is often said that language is a social institution but what this

[13]See Cohen (1985; 1986). An example of modulation is the use of a phrase like "the stone lion" where the content of "lion" has to be modulated to be made compatible with "stone." I treat this phenomenon in detail in Chapter 17.

consists in is seldom made explicit. One dimension of its being an institution is the conventional meanings of words as they are shared by the community and make communication possible. This is such a truism that it is seldom asked what purpose this conventionality serves.[14] Recanati (2004: Sections 9.6–9.7), based on his interpretation of Wittgenstein (1953/1968) and Austin (1961/1979c), entertains the possibility of eliminating conventional meanings altogether while still retaining an intermediate mechanism of abstraction from past uses of words to derive the referential meanings of present uses. One answer to this "Meaning Eliminativism" is that relatively stable, quasi-public conventional meanings facilitate communication by simplifying the choices agents have to make, not to mention the reduction in the burden on memory that results. Likewise, the purpose grammar serves is to reduce effort because there would be many more choices to consider otherwise. Syntax slots words in ways that automatically preclude many possibilities. These are two key elements that make language a social institution. Indeed, the standardization of the various devices themselves, whether words, images, or gestures, also eases communication immeasurably by reducing the possible alternatives. Whether the *existence* of such factors results from our rationality remains an open question but at least it can be so derived mathematically.

As I have said before, my inspiration for this network model of macro-semantics is a large analogy between the social institution of linguistic communication and a market economy, which is similarly ongoing and circular both with respect to the equation of supply and demand within a single market as well as within the wider interactions among different markets. Comparing the idea of meaningful equilibrium above with the idea of general economic equilibrium,[15] utterance situations are like markets, words are like commodities, conventional meanings are like the prices of commodities, and referential meanings are like the quantities of commodities bought or sold. Just as prices and quantities emerge from a general economic equilibrium so conventional meanings and referential meanings emerge from a meaningful equilibrium. Language change is roughly similar to the way in which the prices of goods in a market economy change.

[14] Direct reference theories deriving from Kripke (1980) err when they cut out conventional meanings, viewed as intermediate between words and referential meanings, for certain classes of words. A name N, for example, can be said to have a conventional meaning, the property of being named 'N', as Kneale (1962) and Burge (1973) have suggested, with the added proviso that this property together with a resource situation involving a causal chain of the kind Kripke has proposed determine the referent. My account retains an intermediate conventional meaning, thus avoiding direct reference, and also avoids Kripke's criticisms, including his noncircularity condition. See Kripke (1980: 20–21, 68–70) and also Section 26.2.

[15] See Arrow & Debreu (1954).

And just as a price system spares us the trouble of bargaining every time we buy or sell goods, so conventional meanings and grammar and standardization of the communicative devices we use ease our effort.[16]

A network model of this sort can be deployed at two levels. The first is a more abstract level where the empirical inputs are just assumed as given in order to show that conventional and referential meanings pop out of the model as expected. At this level, one does not attempt to derive the fact that the conventional meaning of DESK is, in fact, the property of being *something made for writing*, just that DESK acquires *some* conventional meaning. This is the level I will look at in Part V as it is the appropriate task for an initial theoretical study. But once such models have been developed in sufficient generality and detail, they can be examined empirically with historical linguists supplying the concrete initial inputs. This is exactly how the work in building models of general economic equilibrium has proceeded, with the early models establishing the theoretical existence of equilibrium prices and quantities and later models applying these early results to particular empirical economies. My purpose in this book is to start off the theoretical task with a relatively simple but revealing model. Later, it can be developed more fully and applied to specific languages. Such work will depend on what empirical data is available and how far back into the mists of time and into the origins of particular languages the data goes.

This completes my sketch of macro-semantics. I believe it goes significantly beyond the explanation envisaged by Grice (1968) and even Lewis (1969; 1972; 1975). Together, micro-semantics and macro-semantics make up my framework for the study of language that I call Equilibrium Semantics or Equilibrium Linguistics because equilibrium plays such a central role in it, not only as an idea and image but also as a practical computational tool. Equilibrium is a deep and widely used notion in the physical and social sciences, and is a part of the fabric of language and communication as well, not a *deus ex machina* imposed from the outside.

To summarize, communication occurs in the small and in the large. In the small, it is an interaction between two agents where an utterance situation, owing to the concerns and goals of the agents in it, induces one agent, the speaker, to elicit a response from another agent, the addressee, by conveying a content. Based on how the addressee might react, the speaker identifies the content and converts it into an utterance involving a sentence with a sound pattern, structure, and meaning. Upon receiving the utterance, the addressee has to interpret

[16]See Rubinstein (2000) for other interesting but quite different connections between economics and language.

it based on her model of the speaker, using the information the sentence provides as well as information from the utterance situation. Once she interprets it, and identifies the content, she can respond accordingly. In the large, there are multiple conversations in society that are linked to one another by an intricate structure. Both conventional and referential meanings arise through communication. As I have emphasized, communication in the small and in the large, what I call micro-semantics and macro-semantics, consists of *circular* processes at multiple levels, not in a "vicious" way, but in a way that makes modeling them quite difficult. It appears to require the special resources of game theory and is not amenable to the kinds of tools philosophers and linguists ordinarily use – logic and semi-technical English.

Those familiar with the literature in economics on communication and cheap talk, especially Crawford & Sobel (1982) and Farrell (1987; 1993), may now appreciate the key difference in my approach from theirs. First, both sets of authors consider just the micro-semantic level. Second, Crawford and Sobel assume there is no exogenously given meaning and Farrell assumes that literal meaning is entirely exogenous. These are two extremes and the truth lies somewhere in between. I discuss the former in Section 8.4 but suffice it to say here that for them meaning is entirely unconstrained and determined endogenously at the level of a single utterance. For Farrell, on the other hand, meaning is fully constrained because it is externally given. The latter begs the question because the problem is to show how meaning is determined. For me, at the level of micro-semantics, meaning is partly exogenous (i.e. conventional meanings are held fixed) and partly endogenous (i.e. literal and implied meanings are determined internally). At the level of macro-semantics, nothing is fixed, everything is endogenous, but an interplay among multiple utterance situations in society is required. Crawford and Sobel collapse these two levels. And Farrell just assumes the answer. There are other important differences too, mainly in the assumption of costless talk, which I discuss in Part III.

There are a number of moving parts in my framework. Different readers will undoubtedly find some of these more acceptable than others, given their own theoretical commitments and interests. I urge them to pick and choose, retaining what suits their tastes or reformulating some of the ideas in new ways. I have tried to identify and address a large swath of the problem of communication and meaning. Once the elements of the framework are grasped, it will be easy to tinker with it, replacing or improving one or other part, or even adapting a substantive idea to a different framework. I invite the reader to be a co-creator in just this way.

From the foregoing account, it should be evident that communication is a complex and many-splendored thing. Viewed broadly, it is what enables symbols and language to emerge and, through them, human civilization itself. And yet, as late as the twenty-first century, we do not have a firm grasp of the science of communication.[17] Some have even felt such a theory of *performance* is an impossibility. Perhaps this is because understanding communication turns out to require a number of novel ideas.

[17] Shannon & Weaver's (1949) celebrated theory of communication is really a theory of information *transmission*. Transmitting information, say from one hard disk to another through a channel, is very different from communication, although the former occurs within the latter. Incidentally, transmitting information and conveying information are one and the same thing.

4 Language and structure

In Chapter 2, I described informational spaces as well as agents and their interactions. Among other things, I introduced the infon space $(\mathcal{I}, \odot)$ which will supply the semantic contents of utterances in Equilibrium Semantics. Assessing utterances requires certain operations on words and parse trees analogous to the operation $\odot$ on infons and so I now define two systems similar to $(\mathcal{I}, \odot)$. One serves as the language to be analyzed and the other provides the syntactic contents.

4.1 Language

Assume a finite set of words $\mathcal{W}$. For example, $\mathcal{W}$ might include $\{$BILL, RAN$\}$. A string on $\mathcal{W}$ of length n is a function from $\{0, 1, 2, \ldots, n\}$ to $\mathcal{W}$. We usually write just a list of words such as BILL, BILL RAN, or RAN BILL. The unique string of length 0, the empty string, is denoted by e. The unique zero string 0 can be thought of as an illegitimate string. Both e and 0 are always members of $\mathcal{W}$. The concatenation $\cdot$ of two nonzero strings is defined in the usual way[1] and $w \cdot w'$ is written ww'. For example, BILL $\cdot$ BILL = BILL BILL and RAN $\cdot$ BILL = RAN BILL. The concatenation of any string with 0 is 0. $\mathcal{W}^*$ is the set of all strings on $\mathcal{W}$ and is called the free monoid on $\mathcal{W}$. It is an infinite set and has an identity e and a zero 0.

Assume G is a context-free grammar (CFG).[2] Here is a simple example:

$$\text{S} \rightarrow \text{NP VP}; \quad \text{NP} \rightarrow \text{N}; \quad \text{VP} \rightarrow \text{V}; \quad \text{N} \rightarrow \text{Bill}; \quad \text{V} \rightarrow \text{ran}$$

Here, "S" stands for sentence, "NP" for noun phrase, "VP" for verb phrase, "N" for noun, and "V" for verb. The only sentence generated by this CFG is BILL RAN. I will use "parse" and "parse tree" to refer also to the subtrees of phrases with the root node being any relevant nonterminal symbol. A tree such as $[_{\text{NP}}[_{\text{N}}\,\text{Bill}]]$ would count as a parse or parse tree of the noun phrase BILL.

[1] See Wall (1972: 164–166) for example.

[2] A context-free grammar is a grammar whose rules all have the form A → w. That is, there is no (linguistic) context surrounding A and w so that the rule can be freely applied. See Wall (1972: Chapter 9).

I now use concatenation to define an operation that yields exactly the subsentential expressions and sentences of the language $\mathcal{L}$ by forming an appropriate proper subset of $\mathcal{W}^*$. Define the following modification of concatenation: $w \circ_G w' = w \cdot w' = ww'$ when ww' is a substring of some string in $\mathcal{W}^*$ that has at least one parse by G and $w \circ_G w' = 0$ otherwise. The set formed by freely generating all strings from $\mathcal{W}$ by this special concatenation operation is the language $\mathcal{L}$. In our little example, $\mathcal{L} = \{e, 0, \text{BILL}, \text{RAN}, \text{BILL RAN}\}$. For instance, the string RAN BILL has been dropped from $\mathcal{W}^*$ because it is not a substring of any string parseable by G.[3] This operation is called *grammatical concatenation* and is abbreviated to $\circ$. It is associative but not commutative and has e as its identity and the zero element 0. In other words, $(\mathcal{L}, \circ)$ is a monoid with a zero. A sentence φ of $\mathcal{L}$ is made up of individual words $\varphi_1 \circ \varphi_2 \circ \ldots \circ \varphi_n = \varphi_1\varphi_2\ldots\varphi_n$ for some natural number n. For the sentence $\varphi = \text{BILL RAN}$, $\varphi_1 = \text{BILL}$, $\varphi_2 = \text{RAN}$, and $\varphi = \varphi_1 \circ \varphi_2 = \varphi_1\varphi_2$.

4.2 Algebraic system of trees

So far, I have defined two algebraic systems $(\mathcal{I}, \odot)$ and $(\mathcal{L}, \circ)$ that capture the structure of infons and linguistic expressions. One describes the world and the other language. The third system involves a new way to describe the grammar G as a system of trees with a product operation.

Each of the five rules of the CFG above can be re-described as a tree. For example, the tree corresponding to the first rule is $[_{\text{S}}[_{\text{NP}}\,][_{\text{VP}}\,]]$ and the tree corresponding to the fourth rule is $[_{\text{N}}\,\text{Bill}]$. Thus, G can be expressed either as a set of rules or as a set of trees. However, we cannot define the desired operation on these trees directly and a little work is required to get them into the appropriate form.

The product operation is defined in two stages. First, an intuitive substitution or merging operation on parse trees is specified as follows. A tree such as $t' = [_{\text{X}}\ldots]$ can be substituted into $t = [_{\text{Z}}[_{\text{X}}\,]\ldots]$ to form $t'' = [_{\text{Z}}[_{\text{X}}\ldots]\ldots]$ where the … from t' have been entered into t because the outer category X of t' matched an inner category X of t. If there is more than one X in t that matches the X in t', then the leftmost one is substituted into. This operation is denoted $\triangleleft$ and we write $t \triangleleft t' = t''$. It is neither associative nor commutative and is identical to the substitution operation defined in Joshi (1985) and Joshi & Schabes (1997).

[3] Consider the sentence I HANDED YOU THE SALT. Then the string YOU THE is a member of the corresponding $\mathcal{L}$ even though it is not a legitimate phrase because it is a substring of the whole sentence which would be parseable by the corresponding G. I owe this example to Tom Wasow.

Consider now the following merging of the second and fourth tree above: $[_{\text{NP}}[_{\text{N}}\ ]] \lhd [_{\text{N}}\,\text{Bill}] = [_{\text{NP}}[_{\text{N}}\,\text{Bill}]]$. This can be informally described by saying that we have merged the simple tree $[_{\text{N}}\,\text{Bill}]$ as far as it could go up the parse tree of the whole sentence without encountering another branch. It cannot go any further because the tree encountered is $[_{\text{S}}[_{\text{NP}}\ ][_{\text{VP}}\ ]]$ which has another branch involving the verb phrase. I call this procedure *chaining*, so we say that simple lexical trees are chained as far up as possible. It is possible to chain the third and fifth trees in the same way to get $[_{\text{VP}}[_{\text{V}}\,\text{ran}]]$. Thus, we are left with just two maximally chained trees from the original five trees and we also have the first tree – $[_{\text{S}}[_{\text{NP}}\ ][_{\text{VP}}\ ]]$. These are:

$$[_{\text{S}}[_{\text{NP}}\ ][_{\text{VP}}\ ]]; \quad [_{\text{NP}}[_{\text{N}}\,\text{Bill}]; \quad [_{\text{VP}}[_{\text{V}}\,\text{ran}]]$$

The chained trees, also called elementary trees, are collected and given names as follows:

$$t_1 = [_{\text{NP}}[_{\text{N}}\,\text{Bill}]]$$

$$t_2 = [_{\text{VP}}[_{\text{V}}\,\text{ran}]]$$

The subscripts of t_1 and t_2 are determined by the sentence being considered, namely, BILL RAN, so that the tree involving the first word BILL is indexed by 1 and the tree involving the second word RAN is indexed by 2. In more complex sentences, there will be more than nine elementary trees and then the indexes will be written (10), (11), (12), and so on. Keeping track of the indexes in this way makes it easier to describe the operation below. The unchained trees – just $[_{\text{S}}[_{\text{NP}}\ ][_{\text{VP}}\ ]]$ in our case – are left in G.

Now, I describe a more complex product operation $\star_{G,u}$ on these trees. It is parametrized by G as $\circ_G$ was and by u as $\odot_u$ was. The utterance situation is needed because the sentence being parsed enters via u and it is on the basis of the sentence that the trees can be properly indexed as explained above.

Because the chosen CFG is so simple, there is just one nontrivial product:

$$\begin{aligned} t_1 \star_{G,u} t_2 &= [_{\text{NP}}[_{\text{N}}\,\text{Bill}]] \star_{G,u} [_{\text{VP}}[_{\text{V}}\,\text{ran}]] \\ &= ([_{\text{S}}[_{\text{NP}}\ ][_{\text{VP}}\ ]] \lhd [_{\text{NP}}[_{\text{N}}\,\text{Bill}]]) \lhd [_{\text{VP}}[_{\text{V}}\,\text{ran}]] \\ &= [_{\text{S}}[_{\text{NP}}\,\text{Bill}][_{\text{VP}}\ ]] \lhd [_{\text{VP}}[_{\text{V}}\,\text{ran}]] \\ &= [_{\text{S}}[_{\text{NP}}\,\text{Bill}][_{\text{VP}}\,\text{ran}]] \\ &= t_{12} \end{aligned}$$

In other words, the tree product draws upon relevant trees in G such as $[_{\text{S}}[_{\text{NP}}\ ]$ $[_{\text{VP}}\ ]]$ to enable them to be merged or substituted. In this product, only one such tree in G was introduced but in general there could be more, both to the left of t_1 or t_2. This is why the operation is parametrized by G.

We now add two special trees. The first is the empty tree, t_e, and the second is a tree t_0. The former serves as the identity of the set of trees obtained via this product and the latter as the zero of the set. That is, $t \star_{G,u} t_e = t_e \star_{G,u} t = t$ for all t and $t \star_{G,u} t_0 = t_0 \star_{G,u} t = t_0$ for all t, the latter being true by definition. When two incompatible trees are multiplied, for example $t_1 \star_{G,u} t_{12}$, the result is stipulated to be t_0.

Note the special vector index "12" of the product t_{12}. First, this index should not be confused with an elementary tree index such as (12) which would be expressed as $t_{(12)}$. In the case of t_{12}, the vector index has two components; in the case of $t_{(12)}$ the vector index has just one component. Second, it may seem at first sight that we could have obtained the product $t_2 \star_{G,u} t_1 = t_{12}$ in the same way, by first substituting the verb and then the noun. But instead the product $t_2 \star_{G,u} t_1 = t_0$ by stipulation. The basic rule is that the subscript of the first multiplicand must be strictly lower than the subscript of the second multiplicand to potentially yield a nonzero tree. When higher-level trees are multiplied in a more complex setting, the rule is that the first component of the first vector index must be strictly less than the first component of the second vector index. For example, a tree labeled t_{34} would have a lower vector index than one labeled t_5 because $3 < 5$. So $t_{34} \star_{G,u} t_5$ could potentially be nonzero. In the reverse order, the product is always t_0.

There are thus five trees in the operation table for this sentence with respect to this CFG, the three above and t_e and t_0. All combinations of these five trees yield one of these five trees. This gives us closure for the operation. The multiplication table for $\star_{G,u}$ is shown in Figure 4.1. Observe that $t_1 \star_{G,u} t_2 = t_{12} \neq t_2 \star_{G,u} t_1$.

$\star_{G,u}$	t_e	t_1	t_2	t_{12}	t_0
t_e	t_e	t_1	t_2	t_{12}	t_0
t_1	t_1	t_0	t_{12}	t_0	t_0
t_2	t_2	t_0	t_0	t_0	t_0
t_{12}	t_{12}	t_0	t_0	t_0	t_0
t_0	t_0	t_0	t_0	t_0	t_0

Figure 4.1: The operation table for $\star_{G,u}$

The basic rules for forming the product of trees are as follows:

1. Merge or substitute via $\triangleleft$ if possible.
2. Otherwise, successively introduce trees from G to the left of either or both multiplicands until one or more substitutions via $\triangleleft$ are possible. For example, if two trees t' and t'' have to be introduced in that order to the left of t_i in a product $t_i \star_{G,u} t_j$, it would be as follows: $(t'' \triangleleft (t' \triangleleft t_i)) \triangleleft t_j$.
3. If the above fails, the result is t_0.

This procedure always gives a unique result as only trees such as $\mathrm{Z} \to \mathrm{X}\ \mathrm{Y}$ from G can left multiply a product like $t_i \star_{G,u} t_j$ where $t_i = [_{\mathrm{X}} \ldots]$ and $t_j = [_{\mathrm{Y}} \ldots]$ and we stipulate that there cannot be another rule $\mathrm{Z}' \to \mathrm{X}\ \mathrm{Y}$ with $\mathrm{Z}' \neq \mathrm{Z}$ in the CFG. Like $\triangleleft$, the operation $\star_{G,u}$ is neither associative nor commutative.

Full parsing may be done in either of two ways: by successive application of compatible rules in the CFG to yield S or by successive application of the $\star_{G,u}$ operation to yield a tree like t_{12}. All other combinations will ultimately result in t_0. In more complex CFGs, there will be multiple trees like t_{12} that may be the end result of combining subtrees, which corresponds to multiple parses for the sentence. The word order of the sentence is automatically taken into account by the product operation owing to the indexing procedure so the yield of successive applications of $\star_{G,u}$ is guaranteed to match it.

Consider the algebraic system $(\mathcal{T}, \star_{G,u})$ where $\mathcal{T}$ is the set of five trees. This captures the relevant subset of the CFG for this sentence and so is an equivalent way to express a context-free grammar sentence by sentence. Each sentence corresponds to a separate algebraic system, all of which can in principle be combined into a larger system but this is unnecessary in practice. If we start with the elementary trees t_e, t_1, t_2, and t_0, we can freely generate the whole set $\mathcal{T}$.

I will assume the grammar G for $\mathcal{L}$ can be rewritten as a system of trees $(\mathcal{T}, \star_{G,u})$ for each sentence in the manner described above. For convenience, I will drop the parameters from the notation and write just $\star$ henceforth and also frequently write $t_i t_j$ instead of $t_i \star t_j$.

4.3 Summary of assumptions

Three algebraic systems have been constructed: $(\mathcal{I}, \odot_u)$, $(\mathcal{L}, \circ_G)$, $(\mathcal{T}, \star_{G,u})$ or, more simply, $(\mathcal{I}, \odot)$, $(\mathcal{L}, \circ)$, $(\mathcal{T}, \star)$. The second system is a monoid with a zero but the first and third have just an identity and a zero. As the parameter u is fixed at the outset, each sentence in it is identifiable, and so the corresponding subsystem $(\mathcal{T}, \star)$ is also identifiable.

In Chapter 3, I sketched what communication looks like in the small and in the large and briefly mentioned the games of partial information that arise. The interaction between speaker and addressee can be partly described by these partial information games and they lead to two more monoidal systems $(\mathcal{G}, \otimes)$ and $(\mathcal{G}', \otimes')$, the first involving semantic games and the second involving syntactic games. These are introduced in the context of the sentence φ = BILL RAN that we will consider in detail in Chapter 7.

A peculiarity of all these systems is that every element of each system that is not an identity element is what is called a zero divisor, that is, a nonzero element for which another nonzero element exists such that their product is zero. Also, there is just an operation of multiplication and the zeros are stipulated rather than arising as identities of a second operation of addition as happens in rings.

Part II

Foundational perspectives

5 Grice and conversation

While the later Wittgenstein was the first to bring communication into semantics by tying meaning to use and while Austin produced some profound insights into meaning and use, it was Grice[1] who pioneered a systematic approach to deriving meaning *from* use. Unsurprisingly, therefore, his work has been enormously influential not only among those who make communication the basis of meaning but even among more traditional researchers still attached to logicism for whom communication remains something of an afterthought.[2] Indeed, he managed to straddle the two traditions of ideal language and ordinary language philosophy by dividing the subject into two parts, one largely conventional and consonant with logic and the other largely inferential and based on rationality, even though communication ultimately underlay both for him.

There are four elements in Grice's (1989c) synthesis of communication and meaning: the idea that rational agency applies to communication as it does to other actions, a theory of conversation based on this idea, a reduction of various semantic notions such as speaker meaning and word meaning that are involved in communication to the intentions of speakers, and a division of the study of meaning into a largely conventional semantic component and a largely inferential pragmatic component, the former preceding the latter in interpretation with both preceded by syntax in turn.

What was once an extremely fresh set of insights has now been shown to have serious defects, many of which have been pointed out by Grice's followers who may be divided broadly into orthodox Griceans, neo-Griceans, and post-Griceans. In my dissertation, *Language and Strategic Inference* (1987b), and in my first book, *The Use of Language* (2001), I myself sought to set some aspects of Grice's thought on firmer ground by using game theory and situation theory. But the inner dynamics of this process undermined these very ideas and an entirely new framework has emerged, leading to quite different construals of the four elements listed above. Not least among these is a mathematical rendering of the underlying phenomena.

[1] An overview of Grice's (1989c) writings can be found in Neale (1992).

[2] I discuss the shortcomings of logicism in some detail in *Language and Equilibrium* (2010: Chapter 1). Key figures include Frege, Russell, and, more recently, Montague. See the references in Chapter 1 of my book.

5.1 Communication as rational activity

Grice's (1989b) explanations were couched in a philosophical psychology that had two shortcomings.

First, his adherence to a belief-desire model based on rationality was at such a high level of generality that it appears that he only accounted for the optimality of *single* actions – essentially against a background of *no* action rather than within the full-fledged context of *choice* (i.e. a whole set of explicitly identified alternative actions) as studied in game theory. In a nutshell, he did not translate desires into preferences over situations and, indirectly, over alternative actions. In addition, this fledgling belief-desire model was one-sided rather than two-sided and did not account for both speaker and addressee.

> This is that for any particular object X and for any feature F and for any activity or type of behavior A, if the creature C believes that the object X both has the feature F and is nearby, or within reach, and that things of type F are suitable for activity A, then the creature wants to A with respect to the object X. In other words, the law harnesses the object to the type of activity. (Grice 1989b: 285)

In this passage, only one behavior A is identified. The problem here and elsewhere in Grice's writings is that he never *explicitly* discusses a choice between two or more actions, say A and B, whether these actions are utterances or interpretations. If he had done so, it would have compelled him to consider a more choice-theoretic perspective. He goes on to say:

> The laws I have mentioned are vulgar laws. The kind of theory in which I think of them as appearing would not be a specialist or formalized psychological theory, if indeed there are such things; I am perhaps not very comfortable with the word "theory" being applied to it. It would be the rough kind of system with which we all work, and the laws in it are to be thought of as corrigible, modifiable, and *ceteris paribus* in character. (Grice 1989b: 285)

This quote indicates, first, that he was probably not aware of decision and game theory in 1976 and 1980 when this late paper of his first appeared and so could not envisage the possibility that there might be a formalized theory for "the rough kind of system with which we all work." The second is his use of *ceteris paribus* in describing such laws which shows that he did have an *implicit* notion of choice that he never fully realized explicitly. Later in the same essay, he even uses the word "optimal" for the first time in his writings but the discussion veers off into other territory and does not address the choice of utterance or interpretation and how such choices might be optimal. Since this paper represents his mature thinking on the topic, we must take it as final. It seems that he was very much on the threshold of choice-theoretic thinking but as choice theory itself burgeoned only in the 1970s, the two passed each other like strangers in the night.[3]

Second, Grice introduced the all-important idea of *interaction* between speaker and addressee in a slightly more structured way than Wittgenstein (1953/1968) or Austin (1961) by seeing it through the lens of strategic reasoning. But this interaction and reasoning remained *epistemic* – he knows that she knows that he knows and so on – rather than *practical* – given what he knows, if he does this, she could do that, and so he should do something else instead. Here is a typical statement of Grice's that purports to account for implicature:

> He said that P; he could not have done this unless he thought that Q; he knows (and knows that I know that he knows) that I will realise that it is necessary to suppose that Q; he has done nothing to stop me thinking that Q; so he intends me to think, or is at least willing for me to think, that Q. (Grice 1989a: 30–31)

All his statements are similarly phrased where the reference to *alternative* actions and their desirability – what the speaker could have said but did not – remains implicit: "he has done nothing to stop me thinking that Q" where the main conclusion Q has already been arrived at *without* any explicit consideration of choice – "he could not have done this unless he thought that Q."

Thus, the two dimensions of rationality Grice enlisted – a one-sided belief-desire model of action and two-sided epistemic strategic reasoning – were not connected to each other and failed to explicitly compare the two-sided desirability of alternative actions. Moreover, he applied this limited sense of rationality only to the derivation of implicature and left it open how literal meaning was to be determined given its many sources of ambiguity: lexical and structural ambiguity, reference resolution and saturation, modulation, free enrichment, and other such phenomena that he mentioned only in passing if at all.

[3]I had started work on my dissertation at Stanford University in 1985 and had tried to contact him at U.C. Berkeley where he was based, but he was already unwell by then and was racing to complete his volume so he demurred. Sadly, I lost the opportunity to meet him and also to tell him about the possibility of using game theory to study language and meaning.

Although he never used the term, perhaps his account was *abductive*, that is, it involved an inference to the best explanation, but choice is present even in abduction, and in all the examples he considered, a *choice* among alternative explanations based on the agent's preferences never arose.[4] In other words, despite superficial similarities, his framework was never intended to be *choice-theoretic* and his sense of strategic rationality was always epistemic rather than based on a two-sided belief-desire model as in game theory.

Indeed, perhaps because it was logicians who took up the study of meaning and not social scientists, the almost exclusive emphasis in the entire field has been on the epistemic dimension of agency. Even when the ordinary language philosophers introduced a way of seeing utterances as actions, it was this epistemic level that persisted. Knowledge and belief were simply brought into practical reasoning but without all the accompaniments of practical action such as choices and preferences and utilities *even* in the domain of pragmatics. This is possibly why game theory has been such a latecomer to the field.[5]

As I stressed in my dissertation,[6] ambiguity applies not just to the lexical and structural aspects of a sentence but much more broadly to the presence of alternative interpretive possibilities as such, and then a framework for choosing among

[4]See C. S. Peirce (1867–1913/1955: 151) for the first use of the term "abduction." It is possible William Whewell (1840) originated the idea. Here is a quote from Hobbs (2004: 727–728): "In deduction, from P and $P \Rightarrow Q$, we conclude Q. In induction, from P and Q, or more likely a number of instances of P and Q together with other considerations, we conclude $P \Rightarrow Q$. Abduction is the third possibility. From an observable Q and a general principle $P \Rightarrow Q$, we conclude that P must be the underlying reason that Q is true. We assume P because it explains Q.

Of course, there may be many such possible P's, some contradictory with others, and therefore any method of abduction must include a method for evaluating and choosing among alternatives." Indeed, as I will show in Section 12.2, the choice-based approach of Equilibrium Semantics is just a species of abduction in which the choice structure is far more sophisticated and detailed than that provided by Hobbs.

[5]Kant's *Critique of Pure Reason* and *Critique of Practical Reason* concern these two dimensions of agency. The trouble arises when one tries to use the epistemic notion to do work that can only be done by the practical notion. The first is generally regarded as falling under metaphysics and epistemology and the second as falling under ethics and political philosophy. Philosophy of language in modern times was pursued primarily by logicians and so has long been practiced as if it were more like the first domain than the second. But it is arguable that even metaphysics and epistemology, let alone philosophy of language, have an inescapable practical dimension. See Parikh (2010: Section 7.6) and Section 2.1 of this book where I argue that Quine's (1960) view of radical indeterminacy fails to take the *optimal actions* of individuating agents into account. To its credit, choice theory – including both decision theory and game theory – unifies these two aspects of human agency. See Myerson (1991).

[6]See *Language and Strategic Inference* (1987b).

such alternative interpretations arising from various sources becomes indispensable. It is only through certain innovations in game theory that we are able to show how the meaning of an utterance may be computed from first principles assuming only the partial rationality of agents. To the best of my knowledge, no other contemporary approach, whether based on Grice or not, comes even close to doing this, and this is because they all lack the constraints on practical reasoning that the situated game-theoretic framework provides. If one wishes to understand linguistic communication and meaning, the appropriate philosophical psychology to adopt is the variant of game theory I call Equilibrium Semantics.

Going further, if one reflects on how human beings act, it becomes clear that in addition to the epistemic and practical levels of agency, there is a third level that might be called *social agency*. While people are, of course, separate entities enclosed in their bodies, there are ways of acting that may involve a direct connection between the minds and bodies of two or more agents through perception rather than inference. Many neuroscientists believe so-called *mirror neurons*[7] may play an important role in imitation, empathy, and understanding others. This close connection operates in conjunction with the more individualistic modes of acting involved in practical agency and practical reasoning. Partly, as Hegel argued, individual persons are *molded* through social interactions in ways we do not as yet fully understand, and partly they operate as separate individuals. Game theory grew out of single-person decision theory and so largely shares its methodological individualism making it difficult to study social agency. Likewise, cognitive science is also largely individualistic in its approach. But there are more holistic traditions in the social sciences, especially in sociology and anthropology, and some of their insights could be incorporated into the more finely articulated and more rigorous forms of choice theory. To a certain degree, I argued for something like this when I said that the preferences of agents in games of partial information are *endogenous* and shared, unlike those in traditional game theory where they are exogenously given and individualistic.[8] Further, this endogenous development is in response to the speaker's possible intentions. This implies that the individual preferences of two or more agents are shaped by common social factors as indicated by social agency.

Notions like *shared situation* are holistic in ways that partly elude us. As we saw in Section 2.2, common knowledge involves introducing non-well-founded situations where a situation may contain itself as a constituent. It is exactly this sort of change that allows us to bring less individualistic social action into com-

[7]See Rizzolatti & Sinigaglia (2008).

[8]See *The Use of Language* (2001: 28).

munication. Endogenous shared preferences can also be developed along similar non-well-founded lines.

There are thus three dimensions of agency: epistemic and practical – which are individualistic – and social, and this last level has barely been mentioned in studies of communication, although it could play a central role. I reproduce Table 1.1 from Section 1.3 as Table 5.1 as it is worth contemplating it in light of this discussion.

Table 5.1: Summary of Theories of Communication

	Logicism	Wittgenstein	Austin	Grice	Lewis
context	marginal	implicit	implicit	implicit	implicit
action	partial	yes	yes	yes	yes
epistemic interaction	no	no	implicit	yes	yes
practical interaction	no	implicit	no	no	partial
social interaction	no	no	no	no	no
computable	no	no	no	partial	partial

The key missing idea in all of this literature is that a rational decision-maker should make decisions that maximize his expected utility,[9] a view that goes back to Bernoulli (1738/1954) and to the early days of probability theory, though its modern justification is due to the classic work of von Neumann & Morgenstern (1944/1947). In the context of game theory, which is two-sided, this modus operandi becomes substantially more complex to describe and requires ideas of equilibrium as we have seen.

Such a procedure can be used prescriptively or descriptively. The seminal work of Allais (1953), Simon (1955; 1956), and Kahneman & Tversky (1979); Kahneman et al. (1982) has demonstrated the descriptive inadequacy of rational decision-making (understood as the maximization of expected utility) in many situations

[9]For an introduction to utility theory, see Myerson (1991: Chapter 1). The utility scale is like the temperature scale: the numbers are not unique (think of Fahrenheit and Celsius) and can be scaled by a positive number k and translated by any number k', that is, $v' = kv + k'$ where $k > 0$, while still representing the *same* underlying preferences. If action a is preferred to a' we write $a \succeq a'$. If v is the corresponding utility function then, given the assumptions of rationality, $v(a) \geq v(a')$ is equivalent to $a \succeq a'$. Making choices involves actions that are more preferred and this is equivalent to choosing actions with higher utilities. This allows us to deal with numerical utility functions instead of qualitative preferences, a great convenience that allows expected values of utilities to be computed. Incidentally, the payoffs I have shown in earlier chapters are utilities.

by establishing that people are just partially rational. The consequences of this behavioral revolution for theories of single-person and interactive decision theory have yet to crystallize[10] despite much experimental research (e.g. Camerer 2003; Vlaev et al. 2011) and many popular books (e.g. Ariely 2008) on the subject.

Such behavioral considerations will be a leitmotif throughout. The Interpretation Game has the kind of structure that does not lead people to deviate from utility theory. The only place where partial rationality may enter is the Content Selection Game. And this is where alternative theories of choice may play a role. There is also a subtler way in which I have already built partial rationality into the structure of communication. By separating it into different and independent levels – the four components of the Communication Game described in Section 3.1 – rather than trying to build a single large encompassing game, I have implicitly assumed that people are finite agents and cannot process everything at once. At the level of the brain, however, it is likely that there is one seamless parallel processing structure.

5.2 The theory of conversation

As is well-known, Grice's framework consists of a general principle called the Cooperative Principle and four lower-level sets of conversational maxims that enable addressees to infer meanings beyond the literal content of an utterance. These are as follows (Grice 1989a: 26–27):

The Cooperative Principle: Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

- The maxim of Quality:
  1. *Ceteris paribus*, do not say what you believe to be false.
  2. *Ceteris paribus*, do not say that for which you lack adequate evidence.
- The maxim of Quantity:
  1. *Ceteris paribus*, make your contribution as informative as required for the current purposes of the exchange.
  2. *Ceteris paribus*, do not make your contribution more informative than is required.

[10] Rubinstein (1998) is one attempt to formalize bounded rationality.

- The maxim of Relation:
  - *Ceteris paribus*, make your contributions relevant.
- The maxim of Manner:
  - *Ceteris paribus*, be perspicuous.
    a. *Ceteris paribus*, avoid obscurity.
    b. *Ceteris paribus*, avoid ambiguity.
    c. *Ceteris paribus*, be brief.
    d. *Ceteris paribus*, be orderly.

Grice's basic idea was that the speaker generally observes the Cooperative Principle and the four maxims when they are communicating. The addressee knows this and so, when she detects an infringement of one of the maxims, she supplies a further meaning, an implicature, to preserve the assumption that the higher-level Cooperative Principle has not been violated. This is an extremely powerful mechanism to explain how implied meanings can be derived rationally. When a speaker does not follow the maxims, he may mislead, opt out of cooperating, be faced with a clash of two or more maxims, or *flout* or blatantly fail to fulfill a maxim. Grice then divides implicatures into three groups: where no maxim is violated, where there is a clash of maxims and so a maxim is not obeyed, and where a maxim is flouted.

For our discussion, it is convenient to consider just situations where a maxim is violated, either literally or by what might be meant. Otherwise, presumably, the maxims play no role in the calculation and it isn't clear how the framework provides a way to obtain such implicatures. The actual example Grice gives to illustrate the case where no maxim is violated is where a stranded motorist beside an immobilized car is told by a passerby that there is a garage around the corner. Grice's own gloss here is that the speaker *would* be infringing the maxim of relevance unless he thinks the garage might be open. This potential violation of the maxim of relevance is not effected by the literal content but by what might be meant by the utterance. This potential violation serves to rule out the possible meaning that the garage may be closed and enables the addressee to opt for the related implicature that the garage may be open. Perhaps one can say that some maxim is violated in all implicatures, either literally or by what might be meant, that is, either actually or potentially, and in certain cases what is meant includes the literal meaning of the utterance and in other cases what is meant excludes the literal meaning.

Consider one of Grice's examples of the flouting of the maxim of Quality:

> *Irony*: X, with whom A has been on close terms until now, has betrayed a secret of A's to a business rival. A and his audience both know this. A says *X is a fine friend.* (Gloss: It is perfectly obvious to A and his audience that what A has said or has made as if to say is something he does not believe, and the audience knows that A knows that this is obvious to the audience. So, unless A's utterance is entirely pointless, A must be trying to get across some other proposition than the one he purports to be putting forward.) (Grice 1989a: 34)

This example makes clear that certain implicatures arise not because people are generally truthful as required by the maxim of Quality but rather because certain falsehoods are *obvious* to the audience. It is the *manifest* nature of the falsehood that generates the implicature, not its mere presence. That is, its falseness is common knowledge. Why should an overt lie matter in this way? Not because people obey the maxim of Quality which enjoins sticking to the truth but because when they knowingly utter a falsehood they generally try to conceal this fact. In the example, one would expect some stress on the word "fine" or an accompanying gesture or facial expression to make the falsehood more obvious. This additional information reveals to the audience that the falsehood is meant to be detected and is therefore not to be taken literally. A and his audience seem to be engaged in a cooperative conversation. If, instead, their own exchange were adversarial and the rest of the setting were identical (and there was no stress or gesture), the audience might merely think A was trying to mislead or had forgiven X; in any case, there would be no implicature and no irony.

Going back to the immobilized car, the lack of relevance of a possibly closed garage is common knowledge between the speaker and addressee and that is why the implicature follows. The speaker does not have to do anything special like stressing a word as in the case of irony above; the implicature flows from background knowledge every adult member of contemporary society has. But it is again common knowledge of the lack of relevance of what might be meant – and not lack of relevance per se – that leads to an implicature.

The same analysis applies to the other two maxims, the maxim of Quantity and the maxim of Manner. A person may not convey all that he needs to if the speaker and addressee are negotiating or may convey too much if he is trying to be helpful. In either case, he would violate the maxim of Quantity. It is only when such a violation is common knowledge that an implicature may be generated. As mentioned in Section 3.1, if $\mathcal{A}$ wants to influence $\mathcal{B}$'s choice of a restaurant, he may have to choose between conveying a content like *should we go to Bar Boulud?* or *should we go to Bar Boulud? we haven't been there in a while.* $\mathcal{A}$ may pick the latter as it would give a reason to $\mathcal{B}$ to accept his suggestion but it may also be unduly prolix and not the least-cost utterance and so violate the maxim of Manner. In such cases, too, it is only when it becomes common knowledge that an utterance is not perspicuous that it triggers an implicature.

Thus, the Gricean maxims will not do. People simply do not speak the way the maxims dictate because they lie or withhold information or do somewhat irrelevant things or fail to be perspicuous more often than not, all without setting off implicatures. A key reason the maxims don't work is the presence of conflict.[11] In addition, utterances affect addressees in positive or negative ways and the desire to produce such effects constrains speakers in ways that lead to routine violations of the maxims that also do not result in implicatures. These observations become especially clear when we consider the Content Selection Game, which is generally partly cooperative and partly conflictual.

Once one admits the influencing of relationships in order to maintain or alter them as part of the story of communication and, equally importantly, also includes the basics of human psychology and motivation, many large-scale phenomena like indirect speech, politeness, and just plain but ubiquitous *framing* – something politicians and used car salesmen know a great deal about – become possible to explain via the Communication Game. Indeed, *all* communication involves framing. $\mathcal{A}$ may choose to give $\mathcal{B}$ a reason why they should go to Bar Boulud because he knows enough human psychology to know that only some content of that kind would be effective.[12] And so he might choose a costlier utterance than the cheapest possibility.

The Content Selection Game and the Generation Game are precisely the sites where a speaker chooses the best way to frame his utterance given the situation he is in. This explains a curious paradox. As a rule, we do *not* choose the cheapest utterances overall, as I also observed in my first book.[13] But once a *way* to realize

[11] As discussed in Section 2.2, all that is required for an interaction to be conflictual is that the interlocutors value the same situations differently. In game-theoretic terms, their payoffs would be misaligned. $\mathcal{A}$ and $\mathcal{B}$ may both want to eat out but may have somewhat dissimilar preferences about restaurants as shown in Figure 3.1. This is such a common occurrence that conflict in this sense is more or less ubiquitous. On the other hand, pure conflict is rare and some cooperation is also typically present because the payoffs are seldom completely misaligned. This is why most games are *mixed-motive* games rather than *coordination* games involving pure cooperation or *zero-sum* games involving pure conflict, the two extremes of a continuum.

[12] See Cialdini (2006), for example.

[13] See *The Use of Language* (2001: Section 8.3).

the speaker's goal has been identified, that is, once a *framing* has been identified, then the speaker *does* try to choose the cheapest utterance that corresponds to the framing. That is, the contents we select are not usually the cheapest but once they have been identified, we do normally choose the cheapest way to convey them. Of course, cost includes not just the effort of producing a sentence but also includes the kinds of things Pinker et al. (2008) mention (e.g. the avoidance of awkwardness in a relationship) based on the Politeness Theory of Brown & Levinson (1987).

If one followed Grice literally, it would seem that speakers always choose the cheapest sentence (unless they want to implicate something via the maxim of Manner). But the foregoing shows why we often do not utter the cheapest sentences as such; we do so only relative to the contents we have chosen to convey. *This* is why language in ordinary human interactions and in literature is far more complex and colorful than we would expect based on straightforward cost-minimizing of the kind Grice envisaged. It is the Communication Game and its multi-level structure that can give a clear account of such large-scale phenomena, as we will see. It makes possible a kind of *thick description*[14] of communication.

In fact, the Cooperative Principle itself appears ambiguous. One way to interpret it is to say that people generally cooperate when they communicate. But this would render it completely implausible owing to the pervasive existence of conflict. In clarifying it, Grice (1989a: 29) says the following:

> The participants have some common immediate aim, like getting a car mended; their ultimate aims may, of course, be independent and even in conflict – each may want to get the car mended in order to drive off, leaving the other stranded. In characteristic talk exchanges, there is a common aim even if, as in an over-the-wall chat, it is a second-order one, namely, that each party should, for the time being, identify himself with the transitory conversational interests of the other.

There are as many as *three* levels at which the Cooperative Principle may be interpreted. As Grice makes clear above, there are so-called *ultimate* aims that could be cooperative or conflictual. This level is *not* germane for the Cooperative Principle. But there are also *immediate* aims like getting a car mended. As Grice lacked the model of a Communication Game with its distinction between various games, he could not say that the immediate aims or "conversational interests" are the goals the agents have in the Setting Game.

[14]See Gilbert Ryle's 1971 university lectures and especially Clifford Geertz (1973).

As mentioned in Section 3.1, in the interaction between $\mathcal{A}$ and $\mathcal{B}$ involving the choice of a restaurant, there is a shared goal of eating out together in the Setting Game. This is the same as the conversational goal. Its subgoals, whether to eat at this or that restaurant, may not be shared. Both the goal and its subgoals are reflected in the payoffs of the Setting Game (e.g. possibly a *Battle of the Sexes*). If the Setting Game induces a Content Selection Game, then the latter's payoffs reflect the former's payoffs. It is the Content Selection Game through which cooperation and conflict primarily enter the process of communication, as shown later.

Some conflict also occurs in the choice of utterance the speaker faces in the Generation Game, especially in what to make explicit and what to leave implicit. Such conflict can occur because of the requirements of politeness or appropriateness or to avoid error. The speaker may prefer to be more direct or less explicit whereas the addressee may prefer greater deference or greater explicitness to prevent error. But such conflict is related more to the maxims just discussed and less to the Cooperative Principle. I discuss this in Section 8.4.

But once a speaker utters something and the addressee tries to interpret it, the Interpretation Game involved is usually one of cooperation because the interlocutors identify with their mutual interests in communicating even if one is insulting the other. This identification is partly what makes their preferences *endogenous* reactions to the speaker's possible intentions, as I said earlier. But even here, as will be seen in Sections 8.1 and 12.2, conflict can occur because certain interpretations may be more or less preferred by the addressee. This is the *third* level at which the Cooperative Principle could be interpreted.

The correct level at which the Cooperative Principle applies is the intermediate level of "conversational interests" or "immediate aims." It does not apply at the level of ultimate aims and it does not apply at the level of the Generation and Interpretation Games. At this intermediate level of the Setting and Content Selection Games, however, there can be cooperation or conflict or both. So the Cooperative Principle does not apply universally to all conversations. However, Grice was right in assuming that it nevertheless plays a crucial role in determining illocutionary meanings such as implicature. This is a subtle point that has been missed by many and so I repeat it: even when there is no common immediate aim or shared conversational goal, the Cooperative Principle still influences implicature. Grice's error lay, as I will show in Chapter 16, in identifying *precisely* what implicatures are generated when the Cooperative Principle does not hold.[15]

[15] I believe my analysis is different from but related to the proposal by Asher & Lascarides (2003), who distinguish Strong Cooperativity (underlying goals coordinated) from Rhetorical Cooperativity (conversational goals coordinated).

Thus, the Gricean maxims might be rephrased into a single maxim of Communication as follows:

The maxim of Communication: If you wish to imply something beyond or in place of what you are literally conveying, ensure that it is common knowledge between you and the addressee that you are either actually or potentially conveying a falsehood or not providing the appropriate amount of information or not being relevant or not being perspicuous and ensure that it is common knowledge that you are not opting out of cooperating at a conversational level.

Implicatures that themselves do not need to become common knowledge can be secured by weaker forms of shared knowledge than common knowledge, as in cases of hinting or suggesting. If a speaker acts in accordance with such a maxim or a weaker variant, then the desired implicature will be calculable from this maxim together with the Cooperative Principle: the addressee will notice that the speaker has uttered a falsehood or is not providing the expected amount of information or is being irrelevant or is not being perspicuous, either actually or potentially, and that this is common knowledge (or something weaker) between them and, given the Cooperative Principle, will be able to infer the implicature.

The requirement of common knowledge is different from Grice's notion of flouting. Flouting is just one way in which common knowledge is realized but common knowledge can also occur in other ways, as indicated by the garage example discussed above. The key thing that makes implicature possible is common knowledge.

The maxim of Communication is still open to the charge often expressed with regard to the Gricean maxims – that they are approximate and vague and may also be vulnerable to other counterexamples.[16] Thomason (1990: footnote 32) says, in fact:

> I have the impression that Grice was tentative in the William James Lectures about his account of conversational implicatures, and may not have been entirely satisfied with its ability to generate detailed explanations of a wide range of cases. On the whole, I believe that linguists and computer scientists have taken the details of Grice's theory more seriously than they perhaps should have. It is important to remember that Grice's William James Lectures were never prepared for publication.

[16] See Horn (1972; 1984) and Levinson (1983; 2000) for example.

This is why the informal expression of maxims, whether Grice's or mine, is itself an inadequate approach to the problem. In Part IV, I will show how all such informal principles may be replaced by formal notions and how these formal notions play a role not only in the derivation of implicatures but also in computing other aspects of what I call illocutionary meaning, such as modulation, free enrichment, and figures of speech, which belong to the literal meaning of an utterance. In other words, the Gricean maxims are completely superfluous.[17] As we will see, however, the Cooperative Principle is required.

Grice's theory of conversation also assumed the pipeline theory of meaning where the syntax of the sentence uttered is determined first, then its semantics (i.e. its literal meaning), and finally its pragmatics (i.e. its implicatures). The implicatures were derived from the literal meaning and other facts of the utterance. As I have shown earlier,[18] literal meaning and implicature can in fact be interdependent as all meanings tend to be. This two-way influence will be discussed in detail in Part IV. I will show, in fact, that not only are the various meanings of an utterance interdependent but so are these meanings and the optimal parse of the utterance. In other words, the implicature can depend in part on the optimal parse and vice versa.

To conclude this section, I point out that the Gricean theory of conversation is really a theory of interpretation and not communication. While it is two-sided, it has little to say about the speaker's calculations. It also offers nothing about how literal meaning is to be derived and one is forced to surmise that Grice probably meant this to be addressed by convention. While conventional meaning does play an important role in the determination of literal meaning, the process is ineluctably contextual as I have shown before and will show in even greater detail in this book.

5.3 Speaker meaning and word meaning

As I said toward the end of Section 3.1, Grice wanted to reduce the intentionality of language, its aboutness or acquisition of meaning, to the intentions and beliefs of speakers, with the idea that these mental states could be further reduced to physical facts. This would explain how meaning could be part of the natural order. In his view, the key notion to effect such a reduction was that of speaker meaning.[19] If that notion could be defined in terms of the speaker's intentions,

[17] This was also argued in *The Use of Language* (2001: Chapter 7).

[18] See *Language and Equilibrium* (2010: Section 4.6).

[19] He used the term "utterer's meaning" as he had in mind the wider scope of what he called nonnatural meaning of which linguistic meaning was one important part.

then word meaning[20] (and, presumably, literal meaning and implicature) could be reduced to what speakers mean.

Grice (1957; 1969; 1989c) gave many definitions of speaker meaning, each successively designed to meet various counterexamples and counter-definitions proposed mainly by Strawson and Schiffer. Consider the following restatement of an early version by Strawson (1964: 446):

> *S* meant something by (or in) uttering *x* iff *S* uttered *x* intending
>
> 1. that this utterance of *x* produce a certain response *r* in a certain audience *A*;
> 2. that *A* recognize *S*'s intention (1);
> 3. that *A*'s recognition of *S*'s intention (1) shall function as at least part of *A*'s reason for *A*'s response *r*.

To summarize the complex intention involved in meaning, the term "M-intention" was introduced. Schiffer (1972: 63) argued against both the sufficiency and necessity of these conditions and gave his own more adequate definition:

> *S* meant that *p* by uttering *x* iff *S* uttered *x* intending thereby to realize a certain state of affairs *E* which is such that *E*'s obtaining is sufficient for *S* and a certain audience *A* mutually knowing that *E* obtains and that *E* is conclusive evidence that *S* uttered *x* with the primary intention
>
> 1. that there be some ρ such that *S*'s utterance of *x* causes in *A* the activated belief that $p/\rho(t)$;[21]
>
>   and intending
> 2. satisfaction of (1) to be achieved, at least in part, by virtue of *A*'s belief that *x* is related in a certain way *R* to the belief that *p*;
> 3. to realize *E*.

While Schiffer's definition may be sufficient, it is not necessary for three reasons. First, his kind of M-intention is too complex for people to entertain in ordinary speech. In fact, even Grice's M-intention is too complex. An argument

[20] In Equilibrium Semantics, there is no need for the notion of sentence meaning so I omit it from the discussion here but Grice did want to reduce both word meaning and sentence meaning to speaker meaning.

[21] The symbol "$p/\rho(t)$" is to be read as the belief *p* for which *S* intends *A* to have the truth-supporting reasons $\rho(t)$.

against both is the inability to form such intricate intentions when we rattle off multiple sentences in quick succession. And yet speakers certainly mean things so such M-intentions cannot be required. Second, there are many utterances such as "It is raining" or "Smith weighs 150 pounds" involving free enrichments of the content where a location is meant but not always explicitly intended so again the conditions are not necessary. By "explicitly" I just mean there is no corresponding mental representation. This objection can be fixed by stipulating that the speaker *implicitly* intends contents that he would assent to if asked. Third, when a language exists with conventionally meaningful words there is no need for M-intentions. Both Grice and Schiffer miss the *circularity* of language where communication explains the emergence of conventional meanings and conventional meanings (partly) explain communication.

In Parikh (2000; 2001; 2010), I give game-theoretic definitions of speaker meaning and other related concepts, especially communication and information flow. The main idea is that the kind of state of affairs E that Schiffer envisaged is replaced by an appropriate game and this game is shown to occur naturally so it does not need to be intended. Moreover, because utterances are ambiguous in many ways as I have remarked above, these ambiguities need to be resolved in some way in order for a speaker to mean the right thing. As a result, the key to speaker meaning is the solution of a certain game.

Definition 5.1. $\mathcal{A}$ means p by uttering φ in u if and only if there is an addressee $\mathcal{B}$ such that $\mathcal{A}$ intends to convey p to $\mathcal{B}$ in u (a proper part of which intention may be implicit) and the solution to the game $G_u^{\mathcal{A}}(\varphi)$ between $\mathcal{A}$ and $\mathcal{B}$ that $\mathcal{A}$ considers in uttering φ in u and (nonconsciously) believes to be common knowledge between them is p.[22]

$G_u^{\mathcal{A}}(\varphi)$ is $\mathcal{A}$'s mental model of part of the interaction with $\mathcal{B}$ and belongs to the Generation Game. It was introduced in Section 3.1 and will be described in Part III so the definition's evaluation has to be deferred. There is only a simple intention to convey p, part of which may be implicit. No complex intentions

[22]The word "convey" in the definition merely signifies a transfer of information as from one computer disk to another. We could replace "convey" with "bring about a transfer of."

One approach to the definition would include other attitudes to p than the having of information by $\mathcal{B}$. For example, $\mathcal{A}$ might intend to cause an activated belief that p in $\mathcal{B}$. Or he might intend to get $\mathcal{B}$ to do something. If we adopt this approach, we can think of this definition as focusing on the case of pure information transfer. Other possibilities are then easy to define.

Alternatively, transferring information is a prerequisite for having other attitudes to p, or doing something, so that one can circumvent having to specify the other possibilities by simply using "convey." This seems like a better approach because the response is left open.

are required. It should also be noted that φ can be a linguistic or nonlinguistic action.[23]

Notice that an *infinite* amount of "work" is being done in both Schiffer's and my definition above. This infinity cannot be avoided. Early attempts by Grice and Strawson led to an unending series of explicit intentions. Schiffer introduced iterated common knowledge which substituted an unending epistemic chain for the unending chain of intentions. Both may be sufficient but are unnecessary as I have argued above. Definition 5.1 renders this explicit infinity harmless by requiring a single simple intention and a shared situation orientation to common knowledge. The explicit infinity becomes implicit – mere entailments rather than mental representations – because of the circular nature of this account of common knowledge.

Just as there are two ways of approaching the iterated type of definition – either via infinitely iterated intentions or iterated common knowledge – so there are also two ways of approaching the circular type of definition, not just via common knowledge but also via circular intentions. Indeed, Grice's (1957) very first attempt was precisely along these lines: "A meant$_{nn}$ something by x" is (roughly) equivalent to "A intended the utterance of x to produce some effect in an audience by means of the recognition of this intention." Here, the speaker's intention contains a reference to itself and so is circular.

Bach & Harnish (1979) and Bach (1987a; 2012a,b) but also Searle (1969) and Harman (1974; 1986) have championed such circular intentions, filling out Grice's somewhat incomplete stab at the definition. In my view, none of their formulations satisfy entirely because they lack a clear language to express the circularity. Barwise (1989c: 194–195) points out that such circular intentions involve circular (or non-well-founded) situations but does not spell out a definition. My account is as follows.

Definition 5.2. $\mathcal{A}$ means p by uttering φ in u if and only if there is an addressee $\mathcal{B}$ and a situation $s \subset u$ such that (a) $s \models \mathcal{A}$ utters φ and (b) $s \models \mathcal{A}$ intends to convey p to $\mathcal{B}$ by $\mathcal{B}$'s recognition of s.

Definition 5.2 shows plainly that the relevant intention is circular and entails an infinite chain of intentions just as happened with common knowledge in Section 2.2. "Recognition of s" is just recognition of all the facts in s. One of these is precisely the fact in item (b) in the definition and so we get $\mathcal{A}$ intends to convey p to $\mathcal{B}$ by $\mathcal{B}$'s recognition that "$\mathcal{A}$ intends to convey p to $\mathcal{B}$ by $\mathcal{B}$'s recognition of s"

[23] As I will not be dealing with nonlinguistic cases here, see Chapter 6 of *The Use of Language* for further discussion.

and so on ad infinitum. This kind of clear formulation is better because it identifies precisely what is involved in adopting a self-referential intention, something that remains obscure in Grice's original formulation and in subsequent ones by Bach and others as well. Just as with the situational account of common knowledge, such an infinite chain of entailments is harmless because it does not involve an agent actually representing these nested intentions mentally. So while Recanati (1986) was right to say that an infinite number of sub-intentions are implied by a reflexive intention, he was wrong to say this poses a problem.

Despite the infinity of entailed intentions being innocuous, such circular intentions are also too complex, certainly compared with the simple intention in Definition 5.1. This is especially true when successive rapid utterances are considered. So while the conditions in Definition 5.2 may be sufficient, they are also not necessary. However, they are simpler than Strawson's and Schiffer's definitions. In other words, both types of iterated approach are worse than the circular intention approach and the latter is worse than the situated common knowledge approach. Indeed, my claim is that Definition 5.1 also gets things right.

Some authors (e.g. Devitt 2013a) argue that such definitions, presumably including Grice's and Schiffer's, confuse the causes of a speaker's meaning *p* with what *constitutes* a speaker's meaning *p*. For Devitt, speaker meaning is constituted by the speaker's "thought content" *p*. In light of the above, this thought content is just a necessary condition that is not sufficient. After all, a speaker will have many thought contents and only those satisfying either Definition 5.1 or some such definition will count as what the speaker meant.

A related issue in Devitt (2013a) is the use of the term "metaphysics of meaning" to describe speaker meaning and distinguish it from the "epistemology of meaning" to describe the addressee's (inferred) meaning. Speaker meaning and addressee meaning are certainly kept distinct in my framework. While Contextualists such as Recanati (2004) confuse them by discussing the addressee's interpretation in place of the speaker's production, the terms themselves are misleading and best avoided because the speaker has to take account of how the addressee will interpret his utterance and so the metaphysics and epistemology are intertwined and not separable. This interpenetration of the two will become completely obvious in Part III.

Speaker meaning is supposed to help with three things: defining word meaning, literal meaning, and implicature. But, in my view, it is too one-sided a notion as it is $\mathcal{A}$'s mental representation of part of the interaction with $\mathcal{B}$. This makes it possible for $\mathcal{A}$ to mean things even when they are not communicated to $\mathcal{B}$. The right notion to use, as suggested by Section 3.2 on macro-semantics, is communication which is genuinely two-sided as the interaction with the addressee is not just a representation in $\mathcal{A}$'s mind. The cross-conversational constraint of consistency on conventional (i.e. word) meanings introduced there applies to *all* users, whether speakers or addressees. That is, conventional meanings arise because both speakers and addressees participate in creating them, not just speakers as Grice supposed. Speakers exploit conventional meanings in speaking and addressees exploit them in interpreting this speech.

Just as there is $\mathcal{A}$'s subjective game $G_u^{\mathcal{A}}(\varphi)$, it was mentioned in Section 3.1 that there is $\mathcal{B}$'s subjective game $G_u^{\mathcal{B}}(\varphi)$ and there is also the actual objective game $G_u(\varphi)$, the latter two of which emerge after the public utterance of φ in u. Incidentally, it is these distinct games that keep speaker meaning distinct from addressee meaning. Since all these games will be described later, the reader should just think of them for the moment as the agents' subjective representations of their interaction and their actual interaction.

Definition 5.3. $\mathcal{A}$ communicates p to $\mathcal{B}$ by uttering φ in u if and only if $\mathcal{A}$ intends (possibly partly implicitly) to convey p to $\mathcal{B}$ in u, $\mathcal{B}$ intends (possibly partly implicitly) to interpret $\mathcal{A}$'s utterance of φ in u, and the games $G_u^{\mathcal{A}}(\varphi)$, $G_u^{\mathcal{B}}(\varphi)$, $G_u(\varphi)$ induced thereby are equal and common knowledge and their solution is p.

Communication involves an actual transfer of p from $\mathcal{A}$ to $\mathcal{B}$ rather than just an intended transfer by $\mathcal{A}$ as with speaker meaning. It is such actual transfers that result in conventional and referential meanings emerging from the interlocking Communication Games in society that make up a Language Game. So speaker meaning is not the right concept to be chasing after despite all the ink that has been spilled on it; it is communication that is foundational. Grice's (1968) attempts at defining word meaning from speaker meaning were in any case extremely informal and vague. He also failed to realize that the two notions – conventional and referential meaning (i.e. literal meaning and implicature) – have to be defined together, as we will see in Part V.

Despite the fact that communication underlay meaning for Grice, he never made explicit the connection between his theory of conversation and speaker meaning as indicated in Definition 5.3 and as explained in Part III. As a result, the speaker and addressee in his account remain solipsistic, locked within their mental states, and their interaction never gets addressed. It is because of this lack of integration of the two sides of communication that orthodox Griceans focus on speaker meaning and revisionist Griceans focus on addressee interpretation and the heart of the problem – communication – eludes the grasp of both.

5.4 Semantics and pragmatics

When Morris (1938) introduced his trichotomy of syntax, semantics, and pragmatics within semiotics, the general science of signs, content, the subject matter of semantics, was thought of as predominantly conventional. For example, Austin (1961/1979c) assumed that demonstrative *conventions* correlated statements with historic situations. Because content and convention were believed to coincide, it is arguable that the use of "semantics" to describe this aboutness of utterances was ambiguous between the two. Later, when Grice (1975) realized they diverged, he chose to identify semantics with the mostly conventional aspect of content, with what is said, rather than with content per se. This latter possibility dropped out of view.

This also meant that implicature, the part of content left over after what is said is assigned to semantics, was relegated to pragmatics. This created a parallel ambiguity in "pragmatics" as applying to communication (as Morris had originally envisaged) and as applying to implicature, an ambiguity Devitt (2013b: Section 4) mentions, although without tracing it back to Morris and Grice. There were thus two shifts: semantics was narrowed from all content to literal meaning and pragmatics was altered from communication to implicature.

Although Grice was aware that literal meaning required the disambiguation of lexical and structural ambiguities and the fixing of pronoun references, he seemed to ignore the uncomfortable fact that the context would have to play a role in the determination of literal meaning and not just implicature. As a result, because Grice held literal meaning to be primarily conventional, there was an implicit third shift: semantics was tied to what comes from the linguistic representation alone and pragmatics was tied to what comes from the utterance situation.

There were thus two distinct scopes for Gricean semantics: one was literal meaning, the other was conventional meaning, and the two were thought to coincide – except for the somewhat unusual category of conventional implicature. Just as Morris had implicitly identified the full content with convention so Grice implicitly identified literal meaning with convention. And just as the presence of implicature meant that the full content could not be entirely conventional, so Searle (1978), Cohen (1985; 1986), and others made it equally clear that the presence of contextual aspects of literal meaning, such as saturation, modulation, and free enrichment, meant that literal meaning could not be entirely conventional. This was in addition to the already known contextual aspects of literal meaning – lexical and structural disambiguation and fixing the references of pronouns – that Grice had chosen to disregard.

This divergence between literal and conventional meaning has led to deep schisms in the understanding of semantics and pragmatics. Contemporary writers such as Bach (1987b; 2001), Levinson (2000), Recanati (2004), Bianchi (2004), Cappelen & Lepore (2005a), Szabó (2005), Carston (2008), Devitt (2013b), and others have tried to redraw the line between semantics and pragmatics in a variety of ways based on different motivations to include either no context in semantics (e.g. Carston – linguistically encoded or conventional meaning, and possibly Levinson – generalized versus particularized conversational implicatures) or restricted bits of context (e.g. Bach – "narrow" and "wide" context, Cappelen and Lepore – shared minimal semantic content, Devitt – a narrow notion of what is said) or all of context (e.g. Recanati – truth-conditional content or a wide notion of what is said or full literal meaning). However, the Gricean distinction between a largely conventional semantics and a contextual/inferential pragmatics surprisingly continues to underpin all these efforts.

Carston (2004; 2008) does an able job of showing the inadequacy of several competing views (Levinson, Bach, Cappelen and Lepore) so I will not discuss their details here. However, her own context-independent linguistically encoded meanings of sentences expressed as "incomplete, gappy meaning structures (semantic schemata or templates, propositional radicals or matrices)" have their own problems. First, it isn't clear that sentences have conventional meanings because only words do. At least, such sentential conventional meanings are not required for generating and interpreting utterances, the real problem of communication, though it is possible that the kinds of gappy structures she posits might be definable with some difficulty. In any case, Carston and the Relevance Theorists lack a precise theory of communication as I will show in Chapter 13, so such posits remain otiose. This is a problem for all the writers mentioned above because none of them have a worked-out theory of communication that enables one to derive the meanings of utterances from first principles so the vigorous redrawing of lines between semantics and pragmatics becomes a little scholastic and reminiscent of pre-Copernican astronomy.

Second, these lexical conventional meanings are *intermediate* steps toward referential meanings, the former being generalizations of Fregean senses and the latter of Fregean references, as I said in Section 3.1. So conventional meanings are not contents that utterances are *about*; they are rather exploited by speakers to convey referential meanings. That is why conventional meanings belong to word *types*; it is only when words are uttered in some situation that they acquire referential meanings based on their conventional meanings.

Third, as I argued in Section 3.2, the context-independence of conventional meaning, that is, of linguistically encoded meaning, is just a convenient assumption in what I call micro-semantics; the real theater of conversation is macro-semantic and, in this society-wide setting, conventional meanings themselves keep changing and getting reset. While speakers and addressees do generally have access to some initial lexical conventional meanings during communication, the final conventional meaning at the end of the process either remains what it was or gets altered, as will be shown in Part V.

On account of the above, especially the first and second points, it is far from clear what purpose is served by calling this complex but relatively small area of lexical conventional meaning within the much larger field of content and communication *semantics*. Carston (2008: Section 4) herself hints at this by saying that semantics could be omitted entirely, leaving just syntax and pragmatics.

Communication involves fundamental indeterminacies of many different kinds.[24] To take just one pertinent example, the basic distinction between literal meaning and implicature is itself indeterminate. In the situation of the stranded motorist referred to earlier where the speaker tells the addressee that there is a garage around the corner, there is no simple way to decide whether there is an implicature that the speaker thinks the garage might be open or there is just a modulated literal meaning that there is a possibly open garage around the corner. Some speakers and addressees might see it one way and others might see it differently. There may not even be a fact of the matter because the speaker may not have had an explicit intention regarding it. Such indeterminacies are pervasive so using, say, the distinction between literal meaning and implicature as the basis for the distinction between semantics and pragmatics seems unwarranted.

The important issue is that this additional information, whether obtained via implicature or via modulation, involves what I call *illocutionary* computations. However, implicature will turn out to be purely illocutionary whereas modulation will have both locutionary and illocutionary dimensions. In general, there are meanings obtained via lexical and structural disambiguation as well as by the fixing of pronoun reference or saturation that involve exclusively locutionary techniques. On the other side, there are meanings such as free enrichment and implicature and direct and indirect illocutionary force that involve exclusively illocutionary techniques. Modulations, the most complex of all, straddle the two. The distinction between locutionary and illocutionary meaning describes *how* the meanings are derived. It crosscuts the other classes of content such as literal meaning which includes everything except for implicature and force. That is, the latter classes are based on the nature of the content, on the *what*, not on how it is computed. These two crosscutting classifications of content will become very

[24]See *Language and Equilibrium* (2010: Chapter 5).

clear in Chapter 20 by which point I will have shown how all types of content are obtained by the speaker and addressee.

Interestingly, even though it is useful to distinguish between locutionary and illocutionary meaning derivations, these processes are more similar than different. Both involve partial information games and there is an underlying uniformity in communication that Equilibrium Semantics brings out clearly. This common feature is that *all* content computation requires context. Thus, contrary to Grice's distinction between a largely conventional semantics and an inferential pragmatics that undergirds all the more recent approaches, it is better to identify semantics with the problem of inferring the entire content, regardless of what contributes to this content, the linguistic representation or the context. As one reading of Morris's trichotomy suggests, the concern of semantics should be with content per se, a view that has disappeared from sight.

There are three basic reasons why such an identification of semantics with deriving content is desirable: it allows a uniform view of all representations whether they are linguistic or visual or gestural or mental; it makes the flow of information rather than representation primary; and, as Austin (1961/1979b; 1975) argued, truth conditions parallel felicity conditions, which makes a construal of the semantics-pragmatics distinction that connects the former with semantics and the latter with pragmatics untenable.

The reason why it is more appropriate to call everything semantics rather than pragmatics as Carston (2008: Section 4) suggests is that semantics refers to the *content* itself and pragmatics is ambiguous between communicative processes and content. There is, therefore, just phonetics, syntax, and semantics and all three mutually determine optimal phonetic, syntactic, and semantic values in communication in a *circular* way, as I will show. Such an expanded semantics that swallows all of pragmatics, if articulated mathematically as Equilibrium Semantics does, can then be seen as a generalized and contextual model-theoretic account of meaning within some wider notion of a *situated* logic. It also reinforces the broad notion of semantics as the discipline founded by Frege, Russell, Wittgenstein, and Austin that I started with in Section 1.1.

This completes my discussion of Grice and why I believe it is time that his synthesis of communication and meaning is superseded by a fully mathematical framework such as Equilibrium Linguistics that offers a much more comprehensive and correct account of the subject. The full proof of this can only be realized through the rest of the book.

6 Incorporating elements of the Romantic tradition

In my previous book, I tried to combine the strengths of two twentieth-century analytic traditions, so-called *ideal language philosophy* and *ordinary language philosophy*. I identified reference as the key notion contributed by the former and use as the key notion contributed by the latter. To these two notions, I added the new ideas of indeterminacy and equilibrium. Thus, Equilibrium Semantics rests on the four fundamental ideas of reference, use, indeterminacy, and equilibrium. Both the analytic traditions are rooted in the rationalist and empiricist doctrines of Descartes, Hobbes, Locke, and others issuing from the seventeenth-century revolution in science and later Enlightenment ideas. They leave out ideas developed in the Romantic reaction to these doctrines embodied by Herder, Humboldt, Hegel, and especially Heidegger. The two sets of ideas, one identified largely with Anglo-American philosophy and the other largely with Continental thought, although these geographic markers no longer hold, have tended to be opposed to each other and consequently neglected by each other. One reason why analytic and science-based approaches have generally ignored the latter is that their expression is often mired in obscure (and sometimes obscurantist) language and another is that their articulation is sometimes genuinely difficult. A third important factor is that most such accounts tend to be nonnaturalistic and hence opposed to science.

Nevertheless, there are some central insights in this Romantic tradition that are worth assimilating. One way to attempt this is by expanding certain aspects of the notion of use that ordinary language philosophy developed in a partial way. My own notion of use was partly based on game theory which, though far ampler in its understanding of action and choice than ordinary language philosophy and much contemporary semantics, is still somewhat narrowly circumscribed. This broadening is obviously a large undertaking and to accommodate it within a single chapter of this book, I restrict myself to the papers of Taylor (1985/1999), who is unusually clear in his formulations of these Romantic elements.

In *Language and Equilibrium* (2010: Section 1.4.2), I identified the following interconnected aspects of use:

- Belief, desire, intention, and agency
- Sentence and utterance
- The situatedness of language
- The efficiency of language
- Ambiguity
- Communication and information flow

Of these six, only the first one, the concept of action, needs some direct alteration; the rest change as a result. I divide my comments on Taylor's papers into two parts.

6.1 Human agency

Compressing greatly, it is possible to pinpoint the following key aspects of action Taylor identifies:

1. Human agents are self-interpreting animals and their feelings in particular (e.g. pride, shame, sense of worth, love, etc.) essentially involve interpretations or articulations in language that may be more or less adequate to their objects. Further, these articulations partly constitute their objects. That is, these formulations may simultaneously be right or wrong and also partly constitutive. So they are representations but not of extra-linguistic and independent objects. (Taylor 1977/1999a; 1981/1999)

2. Some actions involve evaluations (e.g. emotional, moral, political, aesthetic, etc.) that cannot be reduced to a calculus of preferences as required by decision and game theory. (Taylor 1977/1999b)

3. Action is not to be explained in terms of the mental (i.e. beliefs, desires, intentions) but the mental is to be explained in terms of action. Understanding does not precede action but is achieved by performing the action and this understanding also transforms the action. (Taylor 1983/1999)

4. Collective action is not reducible to individual actions. (Taylor 1983/1999)

6.1.1 Persons as self-interpreting animals

Taylor contrasts two interpretations of persons, one the broadly Enlightenment view and the other the broadly Romantic view, both updated to our current times. Both start with the ordinary concept of persons as beings with certain capacities: "a person is an agent who has a sense of self, of his/her own life, who can evaluate it, and make choices about it." (Taylor 1981/1999: 103)

The first view treats all agents as entities that act on the basis of beliefs and desires. There is no essential difference between animals and complex machines. The ends of agents are taken as unproblematic. What is unusual about persons, therefore, is their ability to conceive and make more complex choices: their *strategic* power. The capacities listed above are seen in the context of this ability to plan. The power to represent with clarity plays a crucial role in executing this strategic power. Computation is the key and the difference between persons and other animals and machines is the relative complexity of the calculations involved.

This is the dominant view in current scientific writing about people. However, all modern semantics, including Grice, operates with a much thinner version of even this partial picture. It is only game theory that allows us to address the full range of this strategic power underlying communication and language.

The second view conceives persons as beings for whom things matter in certain special ways, as subjects of *significance*. Things matter for all agents but persons have qualitatively different concerns that are *sui generis*. In other words, the peculiar ends of persons come into focus. Thus, the difference between humans and other animals is not just their greater strategic abilities but also their unusual goals. This means consciousness consists not just in the power to represent but also in the power to constitute our concerns.

What makes human ends qualitatively different is that they involve a sensitivity to certain kinds of situated *standards*. To understand what is meant by "standard," first consider a situation where two persons are preparing food and one says to the other, "The oven is hot." Here, the statement is made relative to some situated standard involving temperature and heat issuing from a situated goal, say that of baking pizza.[1]

[1] I discussed this example from Travis (1996) in Parikh (2010: Section 5.6): the key idea is that the standard or goal relative to which the statement is made arises from the utterance situation and is located in the described situation c of the (Austinian) proposition $c \models \sigma$ conveyed by the utterance. More broadly, Austin's (1961/1979b) felicity conditions are likewise standards that belong to the described situation part of the proposition conveyed by an utterance. Sadly, mainstream semantics has largely ignored Austin and so has largely ignored the facts about standards as dealt with by Travis and in my previous book. See Barwise (1989b) for a discussion of Austinian propositions.

But standards involving heat and temperature are immediately physical and other animals and machines could also be sensitive to such standards (and similar ones like experiencing pain or fear). The kind of standard Taylor (1977/1999a: 54) has in mind is *subject-referring*. For example, the emotion of shame involves (failing with respect to) standards which are experience-dependent properties related to our desires and to our emotional life and in particular to a sense of dignity, of worth, of how we are seen by others, all things connected with the life of a subject of experience. Such standards are part of the situation in which shame is experienced and are bound up with a whole gamut of other emotions, goals, and experiences. So shame and its accompanying standards refer us back to the subject or agent and so are subject-referring.

Further, such situated standards are partly constituted by accompanying articulations or interpretations and so, consequently, are the corresponding emotions. There cannot be an experience of shame without some awareness of certain distinctions because shame is felt with respect to falling short of standards like worth and this is related to levels of high and low. These distinctions or articulations, while partly constitutive of shame, are not arbitrary: they can be right or wrong in delineating the feeling. As I said above, they can be more or less adequate to their objects which means that such feelings are, to a certain degree, independent of their articulations even though they are partly constituted by them.

Being a subject of significance is then being an agent that experiences subject-referring standards and emotions (such as shame, pride, remorse, moral goodness, admiration, contempt, aspiration, dignity, worth, love, certain joys and anxieties, and countless others) that necessarily involve ongoing articulations or interpretations. And being a subject of significance is what makes us self-interpreting animals. We are partly constituted by such self-interpretations. This self-interpreting nature is the distinctive mark of our humanity, not our strategic capabilities, which we share to a greater or lesser degree with other animals and machines.

For these self-interpretations to partly constitute our emotional life and, therefore, partly constitute us, a symbolic medium is required, that is, language in the widest sense, which includes words, images, and gestures among other things. In this wide sense, language articulates feelings and makes them clearer and more defined and hence transforms them. So language partly constitutes us and, as a result, the pre-symbolic hominins must have been much impoverished in their mental lives compared with us.

This second view, in highly condensed form, captures Heidegger's concept of a person according to Taylor. Taylor (and Heidegger) claim that persons and, in particular, their subject-referring emotions that are articulated in and partly constituted by language are irreducible, that is, they cannot be naturalized or reduced to physical things. Further, it is this nonnaturalizability of persons that makes them essentially distinct from other animals and machines.

This is a delicate argument and we can leave this question of nonnaturalizability open, although I believe persons are not sharply different from other animals and machines and are naturalizable. One has to concede, however, that any reduction of the interconnected web of subject-referring emotions and goals and their partly constitutive articulations would have to be quite complex and seems beyond our reach today.[2] Analytic philosophy has generally not faced up to the difficulties raised by such Romantic intuitions about the nature of persons. It continues to focus on strategic man (and, in the case of mainstream semantics, on an extremely thin version of strategic man) and has, relatively speaking, ignored these subtler dimensions of symbolic man.

If the argument is right about the partial constitution of subject-referring emotions by language, then the insight of real import for our discussion is that the linguistic formulations of feelings are representations but not of wholly independent objects. It has been a basic assumption of referential semantics since Frege (in fact, even earlier – Bolzano, for example) that the referents of words are wholly independent of them. This is true of objects like tables and chairs but not of subject-referring emotions. In other words, depending on what they are about, utterances can partly constitute the propositions they convey. Despite this partial constitution, propositions remain abstract entities; that is, they are not linguistic.

But this realization does not affect the fundamental fact that language is nevertheless representational, as Taylor (1981/1999: 101) himself concedes. That is, the key concept of reference still applies to all words, as I have argued in Parikh (2010). In some cases, the infons represented will be independent of the uttered words, and in others, they will be partly constituted by them. This is a relatively small adjustment in the framework though it should be acknowledged that it alters the basic nature of a large number of referents.

6.1.2 Irreducible evaluations of choices

Taylor (1977/1999b) argues that some choices involve *strong* evaluations (e.g. emotional, moral, political, aesthetic, etc.) that cannot be reduced to a calculus of

[2]Think, for example, of *Othello* and of the evil Iago's sharp awareness of the moral and emotional standards that prevail in his community that he exploits so adroitly.

preferences as required by game theory. This is because the evaluations concern considerations of worth that require contrastive articulations like good/bad, high/low and so entail potential conflicts among different self-interpretations. For example, someone may forgo cake because it would adversely affect their cholesterol. This is an instance of weak evaluation and involves utilitarian reasoning. But they might also forgo it because they do not wish to be the kind of person who responds abjectly to bodily appetites: this implies contrasting self-interpretations such as preserving one's dignity versus succumbing to degradation. In other words, desires do not only count in virtue of the positive consummations that result but also in virtue of the kind of life and kind of person considered worthwhile. Taylor maintains that such strong evaluations cannot be reduced to a calculus of preferences and to what he calls instrumental action.

In my view, there is no problem with accepting that, contrary to Taylor, strong evaluations can be so reduced because the preference relations of game theory abstract from the motivations that underlie them, whether they happen to be strong or weak. The contents of the two kinds of evaluation mentioned in forgoing a cake are suppressed in registering that *forgo cake* ≻ *eat cake* because preference is a relation among *situations* or *actions*. At this abstract level, the distinction between the two kinds of evaluation – avoiding cholesterol or avoiding degradation – can be ignored. And these preferences over situations or actions are all that is needed for the *calculation* of the best course of action. Grice (1989b), for example, does not draw a distinction between these two kinds of evaluation but he does use the words "value" and "optimal" and, here, I agree with Grice. It is often not realized by critics of choice theory how thin its demands are with respect to the assumptions it needs to make to draw very powerful conclusions.

But this reduction to preferences and calculation does not thereby imply a reduction to *instrumental* action if the latter means nonmoral and non-self-interpreting action. This is because the choice-theoretic description does not fully exhaust our understanding of action, it only provides a means to predict and *partially* explain it. If we also want to understand its underlying motivations, then it will be necessary to return to the contents that we abstracted from in arriving at their preferences. This is true even if we wish to know whether it was cholesterol or fat that drove them to avoid cake, not just higher-level motivations.

All kinds of evaluation occur in communication, especially in the Content Selection Game described in Section 3.1 and later. The game-theoretic framework abstracts from the content of the reasons why $\mathcal{A}$ wishes to communicate something to $\mathcal{B}$ and focuses on explaining why certain choices are made in light of certain abstracted preferences. In carrying this out, it becomes necessary to explain informally how these preferences follow from the relevant motivations but this informal step does not raise any special difficulties. If required, it can be formalized through the notion of higher-order preferences though this would be unduly pedantic.

6.1.3 The expressive dimension of action

Supported by Hegel's philosophy of mind and action, Taylor (1983/1999) contrasts two views of action, the causal view and the qualitative view.[3] The first view, rooted in the rationalist/empiricist paradigm, takes the performance of an action to be an event (typically involving some physical movement) caused by an agent's beliefs, desires, and intentions that are further reducible to physical things. This can be recognized as the informal underpinning of the notion of action in game theory as well.

The second view, originating especially in Vico but also in Kant and Hegel, takes the performance of an action to be a primitive thing inseparably connected with some direction or purpose that is either semiconscious or fully conscious and that requires articulation to be better understood. This understanding via articulation is an achievement and it transforms the action itself just as it did subject-referring emotions.

This coming to fuller understanding through articulation in a symbolic medium is the expressive dimension of action as developed in Romanticism which, according to Taylor, is missing from the first view as the latter takes actions as fully determined by their causes. Understanding is itself an activity that does not take beliefs, desires, and intentions as immediately and transparently given but as objects that have to be clarified and partly constituted in their expression. For example, our desires that issued in some prior action may crystallize as we reflect on them. And becoming clear about an action in this way changes the action itself.

Thus, action is not to be explained in terms of the mental as the causal view does, rather the mental is to be explained in terms of action as the qualitative view recommends. The mental is not a primitive given but something to be achieved via the primitive capacity for action and its subsequent clarification. Further, because utterances are themselves actions, this reversal applies to utterances as well.

There appear to be two irreconcilable views of action here. In one, actions are caused by beliefs, desires, and intentions, and, if we come to appreciate these

[3]In the paper, Taylor's goal is to explicate Hegel's philosophy in light of the qualitative view, not to infer the latter from the former.

causes better upon reflection, the action itself is not transformed, only our understanding of it. In the other, these causes are at best partial, and they are, as it were, completed when we express and clarify them, and that changes the action itself. For example, someone who voted a certain way may have done so on the basis of inchoate intuitions. The first view would say the action remains unchanged even if the voter subsequently thinks about it and clarifies their initially unformed intuitions. The second view would say the action is transformed by this clearer insight.

It seems the first view is right because in everyday life and also in the law we hold people responsible for their actions based on their initial beliefs, desires, and intentions, however inarticulate, and not on their later cogitations. In the case of subject-referring emotions, it is possible for the emotions themselves to change from their rudimentary forms to more fully developed forms, but not in the case of actions. This is because emotions are ongoing and evolving, actions are not. As Omar Khayyam's memorable verse says, "The Moving Finger writes; and, having writ, Moves on: nor all thy Piety nor Wit Shall lure it back to cancel half a Line, Nor all thy Tears wash out a Word of it."

Besides, Taylor overstates his case in contrasting what he takes to be mutually exclusive views of action. Sometimes (e.g. when we turn on the lights in a room), we understand our beliefs, desires, and intentions perfectly well prior to our actions and no further clarification is required. At other times, an agent such as a voter may not understand their action until they mull over it. And there is a whole continuum of cases in between. For the entire spectrum, the action itself is what it is independent of its symbolic expression.

So the reversal mentioned above, explaining the mental in terms of the action, is to be understood in a qualified way. It is certainly true that we sometimes come to more fully understand what we did upon further thought. But this understanding is *not* involved in causing the action in any way. The action has its own determinants, however embryonic.

It needs to be noted that the beliefs, desires, and intentions assumed in explaining action are seldom fully developed. They are likely to be partial and partially understood in a variety of ways.[4] This affects game theory as much as it does any other causal theory. But the effect is quite limited because a partial intention must nevertheless be sufficiently formed to bring about an action or utterance.

[4]Incidentally, for Hegel complete understanding is possible whereas for Heidegger, who shares the qualitative view of action, understanding is always partial. An implicit part always persists. This latter view of partial understanding is similar to the partiality of situations described in Section 2.1. They can seldom be exhaustively described; there is often something left over.

6.1.4 Collective action

According to Taylor (1983/1999), it follows from the qualitative view that collective action cannot be reduced to individual actions because action is primary and the mental is secondary. This part of his argument is not clear because if there are irreducible collective actions, there also has to be irreducible collective understanding that partly constitutes them. But unless one is willing to accept collective minds, this is not possible.

Collective actions are ubiquitous: for example, newspapers daily report the collective actions of various composite associations. Indeed, communication itself can be seen as a joint action of speaker and addressee like two persons together pushing a cart uphill. And language in the large, the subject of macro-semantics, might also be viewed as being a fully public object potentially irreducible to individual actions.

I think it is true that communication and language involve joint or collective action, something that is seldom acknowledged by mainstream semantics as it is rather remote from taking models of actions and agents as seriously as game theory requires one to do. But there is no reason to think that such action cannot be reduced to individual actions. The model of communication and language developed in this book shows concretely how such a reduction might be carried out in these special cases.

For other types of collective action, all that is required is a suitable generalization of Definition 5.3.[5] This involves shared situations and common knowledge and their non-well-foundedness, which are less individualistic ways of conceiving holistic facts than a directly individualistic account but, in the end, it nevertheless remains a reduction of such facts to the actions and physical states of individuals.

6.2 Language

A good bit of the groundwork has already been laid by the discussion of human agency. We have already encountered some key roles of language adumbrated by Herder, Humboldt, and Heidegger. Part of Taylor's (1978/1999; 1980/1999) goal is to trace the genesis of two contrasting modern approaches to meaning. I will set aside this interesting historical aspect. I also pay less attention to elements where both perspectives are more or less in agreement (e.g. the importance of *use* as a key idea for any theory of meaning).

[5] See Parikh (2001: 58–59).

For Taylor, language has two dimensions, the representational and the expressive, the latter being more fundamental. Expression is realized through the activity of speech, through *use*, so communication and the speech community are central to language. Language shapes the community as much as the community shapes language.

Expression has four aspects: the articulation of inchoate and dimly grasped contents; the making public among interlocutors of such contents; the partial constitution of subject-referring emotions and concerns; and the partial constitution of relationships among interlocutors and among members of the wider community.

The first of these – making something implicit explicit – is in essence similar to what happens with emotions and actions but this process of giving shape covers all of language. There are times when speakers know exactly what they want to convey and just have to convert their mental contents into words; at other times, they may struggle and only partially succeed in expressing their thoughts. This wider scope of what happens in speech is a fact of life and so the simple separation of Content Selection Games from Generation Games must be seen as an idealization, as I said in Section 3.1. The full picture must involve a range of cases allowing contents and words to be identified in any order and even together. For now, we can stick with this idealization as it is useful to first clarify how an easier instance works.

Making public is just the securing of common knowledge between speaker and addressee of the content conveyed. Here, Taylor has something irreducibly holistic in mind and implicitly criticizes the Gricean tradition's iterated approach to common knowledge. As shown in Sections 2.2 and 5.1, the shared situation approach to common knowledge does have a holistic side but it is fully reducible. Parikh (2001: Section 6.4) and Parikh (2010: 225) discuss the publicity of communication and its content in some detail and it is also mentioned in passing in Section 2.2.

The partial constitution of subject-referring emotions and concerns has already been addressed. The key here is that language remains representational but the entities represented are not always entirely language-independent. Some contents expressed will be partly formed by the words uttered by the speaker.

The partial constitution of relationships (e.g. equality between the interlocutors or in a wider political sense) has two senses, one immediate as happens via a Communication Game, and another mediate as happens in society via a Language Game. The first was discussed in Section 5.2 under *framing* and will be addressed further through an example in Section 18.2. The second is discussed in Part V.

With regard to both emotions and social relationships, special mention must be made of the significance of vagueness in language, a new point that is missing from the Romantic tradition as well. It is because almost all words in a natural language are vague to some degree that a language can constitute both human emotions and concerns and large-scale relationships in society like political equality. If our words were perfectly precise, it might be impossible to have the kinds of conversations we do in fact have that allow language to play a constitutive role. Vagueness is essential to language because without it we could not be fully human. This shows how remote such a conception of language is from that of ideal language philosophy. I will say a great deal more about vagueness in Chapter 11.

Taylor (1980/1999: 273–292) argues successfully, I believe, against truth-conditional theories of meaning. A key point seems to be that once an expressive dimension in its four aspects is admitted, understanding must *precede* truth conditions, contrary to truth-conditional meaning theories where understanding consists in grasping truth conditions. A further difficulty with these theories is that they posit a neutral observer who has to apply the theory in order to understand an utterance.[6] I repeat the middle paragraph from Dummett (1996) quoted in Section 3.2:

> Or, at least, they have to be explained together so long as Frege's insight continues to be respected, namely that the concept of truth plays a central role in the explanation of sense. On this Fregean view, the concept of truth occupies the mid-point on the line of connection between sense and use. On the one side, the truth-condition of the sentence determines the thought it expresses, in accordance with the theory developed by Frege and adapted by Davidson; on the other, it governs the use to be made of the sentence in converse with other speakers, in accordance with the principles left tacit by both of them. *That leaves open the possibility of describing the use directly, and regarding it as determining meaning, relegating the concept of truth to a minor, non-functional role.* [my italics] This was the course adopted by

[6] A third point has to do with the primacy of literal meaning in these theories. Here, I believe Taylor is mistaken because all that analytic theories require is that there be some primary meaning from which secondary, indirect meanings such as implicature can be derived. This primary meaning is what is usually called literal meaning. If what Taylor calls "invocative" meanings are primary in other cultures – though I very much doubt this can ever be the case because the demands of everyday life undergird everything including special invocative occasions – then they have to make possible the derivation of other secondary meanings (e.g. literal meanings); otherwise, the very things that language makes possible will be rendered impossible. But doing justice to this part of Taylor's argument and rebutting it would take us too far afield.

> Wittgenstein in his later work. The concept of truth, no longer required to play a part in a theory explaining what it is for sentences to mean what they do, now really can be characterized on the assumption that their meanings are already given. (page 19)

As I wrote in Section 3.2, Equilibrium Semantics squarely pursues the possibility of describing use *directly* and deriving meaning from use. So my approach agrees with Taylor (and Dummett) that use is central and comes first. Moreover, as I have said in Parikh (2010: 208), there is no question of a neutral observer: the ethnographer or theorist may know only part of the Communication Game.

I hope I have made a case for taking the insights of the Romantic tradition seriously. They call for expanding the notion of use that prevails in ordinary language philosophy and incorporating certain elements into Equilibrium Semantics. The adjustment is slight in material terms but the arguments sensitize us to the expressive dimension of language that has been missed by the analytic tradition. It does not conflict with the representational dimension and the four ideas that underlie Equilibrium Semantics – reference, use, indeterminacy, and equilibrium – continue with straightforward alterations. Indeterminacy, especially as it occurs via vagueness, becomes central rather than something to be shunned. A key lack in Romantic theories, including those of Heidegger, is that they remain inadequately theorized from a scientific standpoint. Part of this has to do with their accepting context fully but, as I said in Section 2.1, this does not prevent us from going quite far down the scientific road contrary to what many Continental theorists believe. They have also typically adopted an anti-scientific stance and this explains why such attempts have not been forthcoming. For our purposes here, the two traditions have been amalgamated as far as that is possible.

In closing this chapter, I remind the reader of what I set out to do in Section 1.4 and reiterate that it is desirable and possible to develop a framework that not only brings together arguments in philosophy, linguistics, and artificial intelligence, but also incorporates a few relevant elements of psycholinguistics, sociolinguistics, and historical linguistics in order to build a *unified* science of language and meaning based on communication. Secondly, it has become evident since the beginning of the last century that language and meaning occupy a central place in the human sciences, and to realize this essential role it is necessary to build a solid theory of communication as their foundation.

Part III

Communication Games

7 Defining Communication Games

I start by illustrating the four parts of a Communication Game with a very simple example.

7.1 The Setting Game

Consider the following situation or setting u. Bill Smith, a politician, and Bill Jones, a runner, are acquaintances of $\mathcal{A}$ and $\mathcal{B}$, who are discussing a local election. This is their implicit, more or less shared goal that both agents are able to extract from the exchange. In such an open-ended conversation, there is a very loosely specified Setting Game, a dialogue with alternating discourses by $\mathcal{A}$ and $\mathcal{B}$.

7.2 The Content Selection Game

The implicit goal and the Setting Game induce $\mathcal{A}$ to simply continue the conversation without affecting $\mathcal{B}$ in any very specific way or eliciting any very particular response from her. The only reaction he wishes to draw is that she tacitly accept the information he conveys and continue talking. Since there is nothing conflictual here, let us say that she considers just one response a, which is to accept the information he conveys and respond to it with her own contribution. More complicated scenarios are not uncommon. $\mathcal{B}$ may reject what $\mathcal{A}$ says and respond accordingly. But my goal is first to explain the basics of the model.

In such a loosely constrained dialogue game, all that is required is that $\mathcal{A}$ and $\mathcal{B}$ stick broadly to the topic though even this is not really true as topics in such free-floating conversations are likely to change unpredictably. Given this happy situation, $\mathcal{A}$ may want to convey to $\mathcal{B}$ that *Bill Smith ran in the election.* Call this content σ. He may not bother to think about alternative contents, especially as σ would be a perfectly fine thing to convey in u. This single pair of actions, conveying σ and a, make the Content Selection Game shown in Figure 7.1 trivial.

The game has a situation s at the start that is represented by a large black node. This situation is a part of u and contains $\mathcal{A}$'s intention to evoke the response a by conveying σ. Sometimes the whole content conveyed is not *explicitly* intended

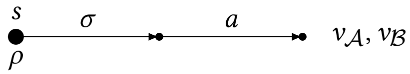


Figure 7.1: Content Selection Game for conversation about local politics

as when someone says "My weight is 150 lbs." and conveys that their weight is 150 lbs. *on earth*. The weight being relative to earth may not have crossed the speaker's mind even in the relevant Content Selection Game let alone in the Generation Game. It is simplest to stipulate that the intention is situated or implicit in such cases. It is important to be aware of such intentions because contrary to what Grice thought and what most of his followers think, all that is conveyed is *not* always explicitly intended, as was pointed out in Section 5.3.

The other label for the initial node is ρ and this is just the (prior) probability of that node. It is the probability of $\mathcal{A}$'s intention to convey σ and evoke a and, derivatively, the probability of s. Given its nature, it is generally a shared subjective probability. Since there is just one initial situation s, $\rho = 1$, but when there are more initial situations, as happens in more complex settings, there will be multiple prior probabilities between 0 and 1 that sum to 1.

Next, there is an arrow issuing from s that represents $\mathcal{A}$'s single choice of action, conveying σ. This leads to an unnamed intermediate situation from which issues another arrow representing $\mathcal{B}$'s single choice of response, the action a. Finally, in the unnamed terminal situation, $\mathcal{A}$ receives the payoff $v_{\mathcal{A}}$ and $\mathcal{B}$ receives the payoff $v_{\mathcal{B}}$. These payoffs are algebraically represented numbers standing for *utilities*, which are themselves representations of the agents' preferences over the outcomes that are delivered through the terminal situation. Preferences and payoffs are determined in a variety of ways and are based partly on the benefits received, the costs incurred, and possibly other factors such as the other agent's payoffs. The benefits and costs are themselves fixed in many ways, some exogenous depending on the payoffs in the Setting Game and some endogenous depending on the other agent's intention.

As indicated above, it is possible to accommodate $\mathcal{B}$'s rejecting rather than accepting the information $\mathcal{A}$ conveys. Rejection may be called a'. Now there would be two arrows issuing from the intermediate node of Figure 7.1, one labeled a and the other a', and there would be corresponding payoffs $v'_{\mathcal{A}}$, $v'_{\mathcal{B}}$ awaiting the players in the new terminal situation. This suggests two possible scenarios. One is that $v_{\mathcal{A}} = v'_{\mathcal{A}}$, which makes $\mathcal{A}$ indifferent to what $\mathcal{B}$ chooses. Another is that, based on what $v_{\mathcal{B}}$ and $v'_{\mathcal{B}}$ are, $\mathcal{A}$ might contemplate some alternative content σ' if he feels that σ may not achieve his desired result based on $v_{\mathcal{A}}$ and

$v'_{\mathcal{A}}$. This leads to a new initial situation s' where $\mathcal{A}$'s intention is to convey σ'. This makes the overall tree structure of the game more complex. I will consider such elaborations later. I note in passing that it is the psychology and sociology of the interaction that result in such complexities in the Content Selection Game and, correspondingly, in different choices of content and utterance and, therefore, in altered communication.

The entire game has been described as a succession of situations, first s, then the intermediate situation, and then the final situation. This idea can be developed formally and I have done so in my two earlier books. It enables games to be represented as structured sets of situations and thereby in terms of our informational universe as I observed in Section 2.2. This reduction suggests a new way to understand games that I call *situated* game theory and allows us to *prove* that such interactive choice situations are *games*. In other words, the construction provides an argument for using games and demonstrates the fit between the model and the facts assuming more basic and self-evident starting points like situations and possibilities for action.

Recall from Section 2.2 that the agents need to share information about this and other games in order to be able to play them fully. At this stage, however, only $\mathcal{A}$ is aware of this game as it is all in his mind and nothing has been communicated yet. He considers the game, identifies his optimal action, and then converts this content into a sentence via the Generation Game that I have yet to describe. It is only after he utters the corresponding sentence and only after $\mathcal{B}$ interprets the utterance by playing the Interpretation Game that she becomes aware of the Content Selection Game. Thus, there is a time lag between $\mathcal{A}$'s having information about the Content Selection Game and $\mathcal{B}$'s having information about it, a factor that matters for the psycholinguistics of communication. Also, $\mathcal{A}$ may ponder alternatives that $\mathcal{B}$ may never imagine – and vice versa – so that their shared information is generally partial. Despite these important qualifications, I will assume the Content Selection Game is common knowledge between $\mathcal{A}$ and $\mathcal{B}$, as also mentioned in Section 3.1. This common knowledge is part of the definition of the game when it is handled formally. In other words, it is not just the theorist who thinks about the game but also the agents, though in our situation the latter do it nonconsciously or semiconsciously.

$\mathcal{A}$ now has to identify the optimal content for him to convey. As we saw in Section 2.2, this is called *solving* the game. The key idea is that each agent has to take account of what the other agent might do because it is their joint actions that affect their payoffs. In other words, when rational decision-makers interact, their individual decision problems have to be solved together, like a system of equations. In addition to making explicit the details of their interaction in a clear model, this circular process where each agent puts himself or herself in the other agent's shoes and tries to guess what the other agent will do and then chooses his or her best action lies at the heart of game theory.

In our very simple game, the agents do not have any *real* choices; they just convey the one content σ and respond with the one action a. So, the solution is trivial and $\mathcal{A}$ perforce picks σ as his equilibrium action. (The action is *conveying* σ, not σ, but we can slur over this nicety.) More generally, the agents would have to take account of the payoffs and prior probabilities and compute solutions by evaluating the expected utility of different *strategies*, as explained in Section 2.2.

As I said in Section 3.1, the Content Selection Game is very different from a standard signaling game because it does not contain any physical action for $\mathcal{A}$ to perform and because of its linkages with the other parts of the whole Communication Game. All that it can do is enable $\mathcal{A}$ to identify what to convey to $\mathcal{B}$ in order to elicit action a from her. This game induces a corresponding Generation Game as a way to achieve this goal. The two games are interdependent because the former relies on the latter for its realization and the latter depends on the former for its existence.

7.3 The Generation Game

Assuming $\mathcal{A}$ wants to convey σ, he searches for sentences that will do the job in u. He may find BILL SMITH RAN IN THE ELECTION by a kind of direct conversion of σ but he may also be cleverer and wonder about BILL RAN. The latter involves less physical effort, though it may require a little more mental effort because he has to check if $\mathcal{B}$ will be able to fill in the rest by using the context u. There are also intermediate possibilities such as BILL RAN IN THE ELECTION.

In practice, he has to construct a sentence from individual words that get identified from bits of partial content. Psycholinguistically speaking, the communication process does not happen with the whole content and the whole utterance being identified in that order, but happens with partial contents and partial utterances and partial interpretations in an ongoing sequential and partly circular process on both sides. This is why the sentences we utter are seldom optimally constructed – their optimization is dynamic and sequential and occurs in chunks. Even when we write, we frequently have to revise our sentences to make them smoother and more graceful. Constructing good sentences is an art and most of us do not ever manage to solve this complex optimization problem fully. This is a consequence of our limited resources and finite rationality, though subjective factors of taste and style also play a role. However imperfect the final result, considerations of efficiency, social norms, and aesthetics guide us in our choices, whether spoken or written.

I'm not sure as yet how the details of this process – the conversion of bits of content to parts of sentences – might be modeled, so I will assume the whole content σ is identified first through the Content Selection Game and candidate sentences are magically available to $\mathcal{A}$ to evaluate in the Generation Game. I will consider BILL RAN first – call it φ. In order to evaluate φ, $\mathcal{A}$ has to imagine uttering it and examine how $\mathcal{B}$ might interpret it given $\mathcal{A}$'s information about u. $\mathcal{B}$'s hypothetical interpretation involves four constraints:

- Phonetic Constraint
- Syntactic Constraint
- Semantic Constraint
- Flow Constraint

All of these operate simultaneously and in a "circular" way. A key fact of natural language is pervasive ambiguity in all aspects of content – phonetics, syntax, and semantics. The basic idea is that, owing to this ambiguity, the Phonetic Constraint generates all the possible words associated with the speech wave corresponding to φ in u, the Syntactic Constraint generates all the possible parse trees of φ in u, and the Semantic Constraint generates all the possible meanings associated with φ in u. The Flow Constraint then disambiguates all these ambiguities simultaneously and identifies the optimal or intended content, that is, the words uttered, their parse, and their meaning. In practice, there are shortcuts which considerably simplify this process.

I will abstract from the first constraint and assume the actual words in the speech wave are immediately available to $\mathcal{B}$. In principle, an algebraic system corresponding to the phonetic properties of utterances can be constructed just as we built up $(\mathcal{I}, \odot)$ and $(\mathcal{T}, \star)$ for semantics and syntax. Showing how semantics and syntax work together to determine content is enough to suggest that a phonetic system can also be added to the mix to deliver its part of the content of an utterance because the principles used in this process can be seen to be quite general. The common idea is to generate all the possibilities and then disambiguate them in an interdependent way. It does not matter what the particular possibilities are, whether words associated with the speech wave from the Phonetic Constraint, parses from the Syntactic Constraint, or referential meanings from the Semantic Constraint. It does not even matter how they are generated, which will naturally be different for each of the three Constraints. That is why they are all *contents* of the utterance. Indeed, the method can be extended even to different but related problems such as detecting spam or sentiments in a message, as will be shown in Section 12.2. All of these are *classification* problems, sorting inputs (i.e. utterances) into separate bins, each set of bins being a different kind of content. Identifying an utterance as spam or its sentiment as positive is just classifying it in terms of more abstract contents. And the interdependence among the three types of contents, phonetic, syntactic, and semantic, also occurs in a uniform way, just through conditioning variables in conditional probabilities, as will be seen presently.

For example, consider an utterance situation with a potentially ambiguous speech wave whose words could be either I WANT TO GET UP LATE or I WANT TO GET A PLATE. The Phonetic Constraint would be tasked with generating these two possibilities. Once they are available, they can be added to the Flow Constraint via phonetic games of partial information that interact with similar syntactic and semantic games of partial information in the manner described below. The Flow Constraint disambiguates phonetic, syntactic, and semantic contents in exactly the same way. Once the interdependence between syntactic and semantic games is understood, it will be more or less obvious how phonetic games can be added to the system.

The traditional way to interpret an utterance is to first obtain its words, then its parse, then its literal meaning, and finally its implied meaning. This pipeline view can be abbreviated as phonetics → syntax → semantics → pragmatics. In Equilibrium Linguistics, semantics and pragmatics are unified into just one process called semantics, and phonetics, syntax, and semantics codetermine each other. I describe the framework below for syntax and semantics. I first consider just the locutionary part of semantics which corresponds to that part of the meaning that derives directly from the words in an utterance. Illocutionary meaning – modulation, free enrichment, implicature, and force – is addressed in Part IV. All that is assumed by Equilibrium Linguistics in addition to the (partial) rationality of agents is an ontology, a language, and a grammar for the language, expressed via the three systems $(\mathcal{I}, \odot)$, $(\mathcal{L}, \circ)$, and $(\mathcal{T}, \star)$.

We are now ready to consider how $\mathcal{A}$ imagines $\mathcal{B}$ interpreting φ in u in light of his desire to convey σ in u. Recall from Section 7.1 that in u there are two Bills, Bill Smith the politician and Bill Jones the runner, either of whom could be the potential referents of φ_1 = BILL in u since it is ambiguous. It is quite possible that $\mathcal{A}$ and $\mathcal{B}$ know yet other Bills, so there may be further possibilities but I will ignore these. Likewise, φ_2 = RAN in u is also ambiguous since it could mean *stood for election* or *competed in a race*,[1] two quite different kinds of running. In fact, my dictionary gives over twenty different meanings for the verb RUN of which the ones corresponding to an election are *be a candidate for, stand for, be a contender for* and the ones corresponding to a race are *compete, take part, participate*. The word RUN can also be a noun and has over ten different dictionary meanings. At the start, it cannot be known that the word is being used as a verb so all meanings would have to be entertained and the inappropriate ones eliminated via the Flow Constraint. In practice, the full range of theoretical possibilities is not included because agents seldom know all the different meanings a dictionary (i.e. the language) can offer. However, at a nonconscious level, it appears that several different meanings do get activated even though we are almost never aware of this.[2] I will consider just the two meanings related to elections and races. Such facts about semantic ambiguity are captured formally by the Semantic Constraint. Only words have conventional meanings so this is all that is mentioned here. In a similar way, as we saw in Section 4.2, $\varphi_1 = \text{BILL}$ has just one parse $t_1 = [_{\text{NP}}[_{\text{N}}\,\text{Bill}]]$, $\varphi_2 = \text{RAN}$ also has just one parse $t_2 = [_{\text{VP}}[_{\text{V}}\,\text{ran}]]$, and so does $\varphi = \text{BILL RAN}$ whose parse is $t_{12} = [_{\text{S}}[_{\text{NP}}\,\text{Bill}][_{\text{VP}}\,\text{ran}]]$. These syntactic facts are captured formally by the Syntactic Constraint.

7.3.1 The Syntactic Constraint

We have already done the work of parsing the words and the sentence in Section 4.2 so all that is required here is to write down the results.

$$\begin{aligned}\varphi_1 = \text{BILL} &\longrightarrow [_{\text{NP}}[_{\text{N}}\,\text{Bill}]] = t_1\\ \varphi_2 = \text{RAN} &\longrightarrow [_{\text{VP}}[_{\text{V}}\,\text{ran}]] = t_2\\ \varphi = \text{BILL RAN} &\longrightarrow [_{\text{S}}[_{\text{NP}}\,\text{Bill}][_{\text{VP}}\,\text{ran}]] = t_{12}\end{aligned}$$

The Syntactic Constraint uses the elementary trees in $(\mathcal{T}, \star)$ to derive the possible parses of each word, phrase, and the whole sentence. In this example, there is no syntactic ambiguity so there is just one syntactic content per word and for the whole sentence. I look at situations with syntactic ambiguity in Chapter 9.

7.3.2 The Semantic Constraint

The Semantic Constraint consists of two subconstraints, the Conventional Constraint and the Referential Constraint. The first maps words into their conven-

[1] Throughout, I will mean a certain kind of sporting event when I use "race."

[2] See, for example, Onifer & Swinney (1981) and Kawamoto (1993). I will be saying more about this later.

tional meanings and the second maps each of these conventional meanings into their potential referents relative to u.

(3) BILL:

- Referential Use: $\varphi_1 \longrightarrow P^{\varphi_1} \xrightarrow{u} \sigma_1$
- Referential Use: $\varphi_1 \longrightarrow P^{\varphi_1} \xrightarrow{u} \sigma_1'$

As mentioned in Section 2.1, properties or relations corresponding to words are written formally as R^{ω} where ω is the relevant word. Since $P^{\varphi_1} = P^{\text{BILL}}$ is a property, I have used P instead of R. It is the conventional meaning of the word BILL and, as I said in footnote 14 in Section 3.2, it is the property *named 'Bill'*. This part of the Semantic Constraint is the Conventional Constraint because it provides the map from a word to its conventional meanings.

The Referential Constraint then maps each conventional meaning into potential referents. The situation u gives rise to two distinct resource situations r_u and r_u', the first containing a causal chain originating from Bill Smith and the second from Bill Jones. There are only these two possibilities given the Setting Game.

This mapping is a bit complicated and I discuss it in Section 26.2 in the context of Frege's puzzle. For now, to keep things simple, just note that these two Bills are obtained via $e(P^{\varphi_1}, r_u) = \textit{Bill Smith}$ and $e(P^{\varphi_1}, r_u') = \textit{Bill Jones}$ based on the definition of $e(P, s)$ on page 19 in Section 2.1. The first possibility is $\sigma_1 = b = \textit{Bill Smith}$ as mentioned on page 16 and the second possibility is $\sigma_1' = b' = \textit{Bill Jones}$. I will assume the two individuals $\sigma_1 = b$ and $\sigma_1' = b'$ are given to us directly and return to how names work in more detail in the very last chapter where I discuss the HESPERUS IS PHOSPHORUS example from Section 1.2.

The symbol u on top of the second arrow above is required to constrain the possibilities by the information in u. It is because of u that BILL gets mapped into Bill Smith or Bill Jones and not any of the other countless Bills that doubtless abound. On the other hand, the first arrow in each case does *not* have a u on top of it because words acquire conventional meanings through communication as discussed in Section 3.2 and then, until a change in meaning occurs, they remain attached to the words independent of context.

Both uses of BILL are of the same kind and are called referential. There are other ways of using names, most commonly attributively but also generically and predicatively, just as with descriptions.[3]

[3] See Parikh (2010: Chapter 6) for more details. For example, the name JANUARY in an appropriate use of the sentence JANUARY HAS 31 DAYS may be used generically.

(4) RAN:

- Predicative Use: $\varphi_2 \longrightarrow P_1^{\varphi_2} \xrightarrow{u} \langle\langle P_1^{\varphi_2} \rangle\rangle = \sigma_2$
- Predicative Use: $\varphi_2 \longrightarrow P_2^{\varphi_2} \xrightarrow{u} \langle\langle P_2^{\varphi_2} \rangle\rangle = \sigma_2'$

These predicative uses of RAN involve two distinct conventional meanings as noted above and each is mapped into itself by the Referential Constraint. Specifically, $\langle\langle P_1^{\varphi_2} \rangle\rangle = \sigma_2$ is the meaning *stood for election* and $\langle\langle P_2^{\varphi_2} \rangle\rangle = \sigma_2'$ is the meaning *competed in a race*. Despite the Referential Map's being an identity map, we write $P_1^{\varphi_2} \xrightarrow{u} \langle\langle P_1^{\varphi_2} \rangle\rangle$ instead of writing $P_1^{\varphi_2} \xrightarrow{u} P_1^{\varphi_2}$ as the angled brackets serve to emphasize that a potential referent of RAN is an infon, which in this case is just the property $P_1^{\varphi_2}$ written as $\langle\langle P_1^{\varphi_2} \rangle\rangle$. I also use the term "referent" much more broadly: in Equilibrium Semantics, every word more or less *refers* to its content via the Conventional and Referential Maps and so do speakers.

Notice that the word RAN actually contains two bits of information: the possible referents of the verb TO RUN and the time of running, some time prior to the time of utterance. In other words, strictly speaking, RAN should generate the two possibilities σ_2 and σ_2' as well as temporal infons σ_3 and σ_3' (or continuous ranges of temporal infons). Indeed, if it is known to the interlocutors when the election under discussion took place, then this time can be more precisely fixed than simply being some interval in the past. I will ignore these complexities, both how a single word might generate two distinct sets of possibilities and how temporal possibilities might be narrowed by taking contextual information into account. They do not pose insuperable obstacles but do complicate the framework and my purpose is to delineate just the basic ideas. For that reason, I will drop σ_3 and σ_3' and work just with the infons identified above.[4]

Thus, the Semantic Constraint, based on its two component constraints, provides the *possible* referents of each word in the utterance. Larger units, such as phrases and the whole sentence, do not have conventional meanings. Their possible referents are derived by multiplication of the partial infons σ_1, σ_1' and σ_2, σ_2' via the unification operation $\odot$ as will become clear shortly.

Equilibrium Linguistics treats all values, whether syntactic, semantic, or phonetic, as *contents* and treats them uniformly. The outcome of the three constraints is to produce the possible corresponding contents for each relevant part of the utterance.

[4]The details involve conditioned infons dealt with in Parikh (2010). Essentially, σ_3 and σ_3' condition σ_2 and σ_2' respectively.

7.3.3 The Flow Constraint

It is helpful to keep the coordination game in Figure 2.1 in mind for this section.

The Flow Constraint *disambiguates* among the possible contents thrown up by the Phonetic, Syntactic, and Semantic Constraints. There are four lexical games of which two are syntactic and two are semantic, corresponding to the two words in φ. (If we had included the Phonetic Constraint, there would also have been phonetic games except that, in this case, it is the lexical items that have to be determined from the speech wave.) It is convenient to start with the semantic games as these are more likely to be familiar. Recall that in the Generation Game $\mathcal{A}$ is trying to imagine how $\mathcal{B}$ might interpret an utterance of φ in u. Figure 7.2 shows the first semantic game of partial information he constructs in order to do this.

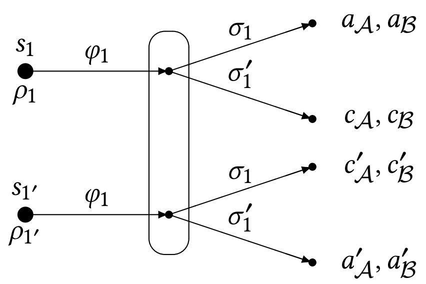


Figure 7.2: Semantic lexical game g_1

This game should not be confused with the Content Selection Game in Figure 7.1, although there are superficial similarities because both structures involve communication. In the latter game, $\mathcal{A}$'s choice concerns the content σ and $\mathcal{B}$'s choice concerns her response a. In the present game, based on the Content Selection Game, $\mathcal{A}$'s choice involves a hypothetical utterance of φ_1, the first word of the whole utterance φ, and $\mathcal{B}$'s move involves her possible interpretations of this word uttered in u. In terms of complexity, g_1 is slightly more complex than the game in Figure 7.1 as there are now two initial situations s_1 and $s_{1'}$ instead of just one. (Note that the prime is on the 1 and not on the s. The reason for this will become clear presently.) Each initial situation is part of the utterance situation and includes an appropriate intention to convey some content. Recall from Section 7.2 that this intention may be implicit.

In s_1 in Figure 7.2, the speaker is assumed to be conveying σ_1, *Bill Smith*, although because his action of uttering φ_1 = BILL is semantically ambiguous between σ_1 and σ_1', the addressee could potentially interpret the word as referring to σ_1', *Bill Jones*, as well. This choice is represented by two branches labeled σ_1 and σ_1' emanating from the upper intermediate situation inside the elongated oval. That is, if $\mathcal{A}$ utters φ_1 in s_1, the resulting situation, the upper intermediate node inside the oval, offers $\mathcal{B}$ two choices of interpretation σ_1 and σ_1'. In the terminal situations that result from each choice, the payoffs for the two agents are $a_{\mathcal{A}}$ and $a_{\mathcal{B}}$ for making the right choice σ_1 as this is what $\mathcal{A}$ is conveying in s_1, and they are $c_{\mathcal{A}}$ and $c_{\mathcal{B}}$ for making the wrong choice σ_1' as this is not what $\mathcal{A}$ is conveying in s_1. The a payoffs are therefore greater than the corresponding c payoffs as they represent the correct interpretation by $\mathcal{B}$.

In $s_{1'}$ in the lower half of g_1, a similar choice presents itself to the addressee but this time the payoffs are reversed, because in this situation $\mathcal{A}$ would be conveying the other possible meaning σ_1'. Again, the primed a payoffs are greater than the corresponding primed c payoffs.

As we saw in Section 2.2, the oval enclosing the two intermediate nodes is an *information set.* It represents $\mathcal{B}$'s inability to distinguish between them as she has seen the same utterance φ_1 in both cases and cannot know if the originating situation was s_1 or $s_{1'}$ (since what $\mathcal{A}$ is conveying is not antecedently available to $\mathcal{B}$). So she is constrained to make the same choice in both situations: either she must choose σ_1 in both or σ_1' in both. If the oval were not there, she could make different choices in the top half and the bottom half of the game tree. This constrains the possible solutions of the game. Depending on the other features present, such games are called games of *imperfect*, *incomplete*, or *partial* information. In this case, this is a game of partial information for reasons given below.

The symbol ρ_1 represents the prior conditional probability that $\mathcal{A}$ is conveying σ_1 *given* that he is conveying the other meanings and parse trees of the rest of the sentence in the utterance situation u.[5] In other words, $\rho_1 = P(\sigma_1 \mid x_2, y_1, y_2; u)$ where the x_2 is a variable ranging across the possible corresponding meanings σ_2, σ_2', and the y_1, y_2 are variables standing for the possible corresponding parse

[5]In retrospect, this idea could be seen as a generalization of the concept of selection restrictions, which are the semantic restrictions a word imposes on the environment in which it occurs. See Chomsky (1965). For example, a verb like "eat" generally requires that its subject refers to an animate entity and its object to something concrete. A violation of the selectional restrictions of a word results in an anomaly: in "the mountain eats sincerity" both restrictions are violated, rendering the sentence anomalous. Such restrictions are often not absolute as there are uses of "eat" such as "the meeting ate up a lot of time," an example I owe to Tom Wasow. The question whether selectional restrictions should be treated in syntax or semantics, or even outside grammar, as a matter of knowledge of the world, has been a point of debate. My idea for the prior probabilities was developed independently of this notion and it occurs entirely outside grammar.

trees t_1, t_2, respectively.[6] I have omitted writing other conditioning variables involving phrasal and sentential meanings and parse trees such as $x_{12} = x_1 \odot x_2$ and $y_{12} = y_1 \star y_2$ as their influence occurs through the corresponding component lexical items anyway. The situation u is *not* a random variable but a parameter that is held fixed. This notation is a slight abbreviation for the actual probability which is a conditional probability of a *speaker's conveying a content* rather than just the conditional probability of a content per se. This is a very important difference because a speaker could be likely conveying a rare content and it is the former likelihood that matters. Note that ρ_1 is actually a *function* of the conditioning variables and parameter u and not a constant.

Likewise, the other prior probability $\rho_{1'} = P(\sigma_1' \mid x_2, y_1, y_2; u)$ and we must have $\rho_1 + \rho_{1'} = 1$ as there are only two possible meanings the speaker could be conveying. Since $\rho_{1'}$ is also a function of the conditioning variables and parameter u, there are different *versions* of essentially the same game g_1 when x_2, y_1, or y_2 take on different values. These versions differ only in the numerical values of the prior conditioned probabilities; in all other respects, they are identical. Thus, for example, when $x_2 = \sigma_2$ the values of ρ_1 and $\rho_{1'}$ may be 5/6 and 1/6 and when $x_2 = \sigma_2'$ the values of ρ_1 and $\rho_{1'}$ may be 1/4 and 3/4. g_1 has just two versions. One reason why it is a game of partial information and not a game of incomplete information (or a signaling game, one kind of game of incomplete information) is that the prior probabilities are not fixed in advance.

If I had worked out how the Phonetic Constraint generates its possible contents, that is, the possible words in the speech wave, then these possibilities would just have been added to the conditioning variables just as the syntactic possibilities are. That is all the interdependence among phonetic, syntactic, and semantic contents consists in. So, in principle, it is easy to add phonetic contents into the mix once the details of the Phonetic Constraint are spelled out. I have done this for both the Syntactic and Semantic Constraints so augmenting the system with phonetic contents is a natural extension. There is a slight complication because in semantic and syntactic games one is interpreting words via meanings and parses and in phonetic games one is interpreting a speech wave via words, but this does not pose any insuperable technical problems.

Returning to Figure 7.2, $\mathcal{A}$ knows what his intention is and so he knows the real situation is s_1 and not $s_{1'}$. But because $\mathcal{B}$ does not know $\mathcal{A}$'s actual intention in advance – as represented by the information set – $\mathcal{A}$ realizes that $\mathcal{B}$

[6] There is no syntactic ambiguity in this example and so y_1 just stands for t_1 and y_2 for t_2. If there had been other alternative parses, for example, some tree t_1' as a possible parse of φ_1, then y_1 would have ranged over both t_1 and t_1' just as x_2 ranges over σ_2 and σ_2'. Likewise with y_2. I will consider such an example in Chapter 9.

might think the real situation was $s_{1'}$. So $\mathcal{A}$ has to consider this counterfactual situation $s_{1'}$ as well. He has to portray the interaction as $\mathcal{B}$ would see it and, therefore, as both should see it. The prior probabilities indicate $\mathcal{A}$'s view of what chances $\mathcal{B}$ would assign to the two situations. But these chances depend on various conditioning variables which gives rise to different versions of the game. Which version to play is a further decision each agent has to make. That is, the speaker has to choose not just what to utter but also the probability distribution $(\rho_1, \rho_{1'})$ by specifying the values of the conditioning variables – in this case, just whether $x_2 = \sigma_2$ or $x_2 = \sigma_2'$ since the variables y_i can take on just one value each. Likewise, the addressee also has to choose not just an interpretation but also the probability distribution. Expressed concretely, both agents have to choose between the pair $\rho_1 = P(\sigma_1 \,|\, \sigma_2, t_1, t_2; u)$, $\rho_{1'} = P(\sigma_1' \,|\, \sigma_2, t_1, t_2; u)$ and the pair $\rho_1 = P(\sigma_1 \,|\, \sigma_2', t_1, t_2; u)$, $\rho_{1'} = P(\sigma_1' \,|\, \sigma_2', t_1, t_2; u)$. (I use the same symbols ρ_1, $\rho_{1'}$ for both pairs only to avoid having to specify the arguments of the functions these symbols really represent.)

Now, $P(\sigma_1 \,|\, \sigma_2, t_1, t_2; u) > P(\sigma_1' \,|\, \sigma_2, t_1, t_2; u)$ because σ_2 is compatible with σ_1 as Bill Smith, the politician, is more likely to be running in an election than Bill Jones, the runner. This is also reinforced by u as the topic of conversation is the local election as mentioned in Section 7.1. In other words, the agents' common goal is to discuss the local election. On the other hand, $P(\sigma_1 \,|\, \sigma_2', t_1, t_2; u) < P(\sigma_1' \,|\, \sigma_2', t_1, t_2; u)$ because now σ_2' and σ_1 are incompatible as one involves running in a race and the other involves Bill Smith whereas σ_2' and σ_1' are compatible because Bill Jones is a runner. This is so despite the presence of u which might bolster the former probability because incompatible meanings are much worse than an irrelevant remark. The parse trees t_1 and t_2 do not play any role as they are unique. Taking the earlier numerical values, both players have to choose between $(\rho_1, \rho_{1'}) = (5/6, 1/6)$ and $(\rho_1, \rho_{1'}) = (1/4, 3/4)$. These numbers are determined by a variety of objective and subjective factors that are exogenous but the choice between them is endogenous.

This semantic lexical game g_1 is obtained by applying the semantic game map g_u to the word φ_1. That is, $g_u(\varphi_1) = g_1$. It is possible to construct g_1 entirely from the assumption that $\mathcal{A}$ is trying to imagine how $\mathcal{B}$ might interpret his uttering φ_1 as part of φ in u, from the Semantic Constraint which provides its possible interpretations, from his intention to convey one or the other meaning which sets the payoffs to higher or lower values based on whether $\mathcal{B}$'s interpretation gets it right or not, and from internally determined inequalities among the partly objective and partly subjective probabilities such as $P(\sigma_1 \,|\, \sigma_2, t_1, t_2; u) > P(\sigma_1' \,|\, \sigma_2, t_1, t_2; u)$. The map g_u is available to $\mathcal{A}$ without any additional assumptions and all he has to do is apply it to φ_1.

I reiterate that g_1 has been entertained only in $\mathcal{A}$'s head so far. $\mathcal{B}$ has as yet no idea what is brewing in $\mathcal{A}$'s mind. Later, when $\mathcal{A}$ identifies an optimal utterance and utters it, $\mathcal{B}$ will construct a very similar game by applying a similar map g_u to the utterance she hears. These games constructed by the two interlocutors may or may not be identical and when they are identical they may or may not have common knowledge of it. To mark the possibility that these games may diverge, I label them $g_u^{\mathcal{A}}(\varphi_1) = g_1^{\mathcal{A}}$ and $g_u^{\mathcal{B}}(\varphi_1) = g_1^{\mathcal{B}}$, the first called a speaker game and the second called an addressee game as mentioned in Section 3.1. I should clarify that the prior probabilities in $g_1^{\mathcal{A}}$ represent $\mathcal{A}$'s view of $\mathcal{B}$'s assignments whereas in $g_1^{\mathcal{B}}$ they represent $\mathcal{B}$'s own assignments. To avoid unnecessary encumbrances, I will just drop the superscripts and leave the disambiguation to the context. When I need to discuss them both together, I will reintroduce the labels.

As an aside, I mention that a reader may wonder how a speaker's private intention which is part of s_1 or $s_{1'}$ can ever become common knowledge. This can happen because both share the word that has been uttered and they share the two possible interpretations of the word because they share the Semantic Constraint. This leads them to share the same game tree with s_1 and $s_{1'}$ as its initial situations as just explained. And the addressee would know that in s_1 the speaker must be intending to convey the first interpretation and in $s_{1'}$ the second. That is, she is confronted with two possibilities and she can infer that in the first there must be a certain intention and in the second there must be a different corresponding intention. The same question could arise with seemingly private preferences. As explained above, both agents would come to have the common knowledge that the a payoffs are greater than the c payoffs because the former represent the correct interpretation and the latter the incorrect interpretation.

Clearly, then, g_1 presents a choice situation for both agents to consider, as yet only in $\mathcal{A}$'s thought. While he does not have any choice of utterance – he has to utter φ_1 – he does have a choice of probability distribution. This choice depends on the optimal interpretation determined for φ_2, σ_2 or σ_2', because $\mathcal{A}$ will want his choice to be compatible with this interpretation. Otherwise, he may choose one value and therefore one version of the game and $\mathcal{B}$ may choose another and, besides, the interpretations chosen for the parts of the whole utterance may not hang together. As the reader may anticipate, there will be a corresponding semantic lexical game $g_u(\varphi_2) = g_2$ that involves precisely a choice for $\mathcal{B}$ between σ_2 and σ_2'. So, the right solution for g_1 depends on the solution to g_2. But, as will be seen presently, the right solution for g_2 depends on the solution to g_1. In the same way, there will also be corresponding syntactic games $g_u'(\varphi_1) = g_1'$ and $g_u'(\varphi_2) = g_2'$ which involve just a trivial choice of trees t_1 and t_2 as these are the only possible parses of φ_1 and φ_2. In general, their solution will also influence the solution to g_1 and vice versa. Just as $\mathcal{A}$'s only real choice is between two prior probability distributions, so $\mathcal{B}$ has a choice of interpretation between σ_1 and σ_1' and also a choice between the same two probability distributions. She will also want to make her choice of distribution compatible with $\mathcal{A}$'s and with her own choice of interpretation in g_2 as just observed.

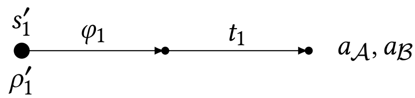


Figure 7.3: Syntactic lexical game g_1'

The syntactic lexical game g_1' is shown in Figure 7.3. In it, the only initial situation is s_1'. This corresponds to the initial situations s_1 and $s_{1'}$ in Figure 7.2. If there had been an ambiguity in the parse of φ_1, say another tree t_1', there would have been a second initial situation $s_{1'}'$ in Figure 7.3. Notice where the primes are placed, whether on the symbol s itself or on the subscript 1 or both. The prime on the symbol corresponds to the syntactic game and the prime on the subscript determines whether it is an alternative initial situation in the relevant game, semantic or syntactic.

In this game, there is no intention to convey the parse t_1 in s_1' as there was in the previous game g_1. In semantic games, there is an intention whether explicit or implicit. In syntactic games, neither the speaker nor the addressee are likely to be aware of the parse of the corresponding word or expression. All that happens is the initial situation is a part of the utterance situation u which includes a fact that the relevant parse tree is being transmitted or conveyed, just like a transfer of information from one hard disk to another.

The rest of the diagram should be straightforward to read. $\mathcal{A}$ has just the one choice of utterance φ_1 and $\mathcal{B}$ has just one choice of syntactic interpretation t_1. The payoffs received are the same as before because the interpretation is obviously correct with respect to what is being transmitted in s_1'.

The prior probability $\rho_1' = P(t_1 \mid x_1, x_2, y_2; u)$ where the x_1 is either σ_1 or σ_1', x_2 is either σ_2 or σ_2' as before, and y_2 is just t_2 as there is no other parse tree corresponding to φ_2. It is also influenced by the parameter u. Since there is just one alternative t_1 in this game, $\rho_1' = P(t_1 \mid x_1, x_2, y_2; u) = 1$ independent of what values the conditioning variables adopt.

This syntactic lexical game g_1' is obtained by applying the syntactic game map g_u' to the word φ_1. That is, $g_u'(\varphi_1) = g_1'$. Unlike the case of the semantic game g_1, g_1' is constructed entirely nonconsciously or subpersonally by $\mathcal{A}$. This is also often true for semantic games but in those instances there may be a kind of semiconsciousness present. The syntactic game g_1' uses the parsing alternatives provided by the Syntactic Constraint and other information analogous to the information used in constructing g_1. That is, the map g_u' is available to $\mathcal{A}$ just as g_u was and all that is required is for $\mathcal{A}$ to apply it to φ_1.

This game also presents a choice situation for both agents, as yet only in $\mathcal{A}$'s head. All the choices on both sides – the utterance and the prior probability on the speaker's side and the interpretation and the prior probability on the addressee's side – are trivial. So, the solution to this game is also trivial but it still does depend in an otiose way on the values of the conditioning variables x_1, x_2, and y_2 and so on the solutions to those games just as the solutions to those games depend on the solution to g_1'.

Now, we move to the next set of lexical games for the second word φ_2 = RAN shown in Figure 7.4 and Figure 7.5. These diagrams should be relatively straightforward to understand.

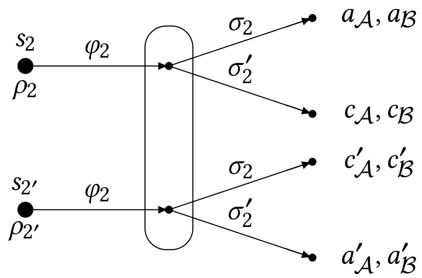


Figure 7.4: Semantic lexical game g_2

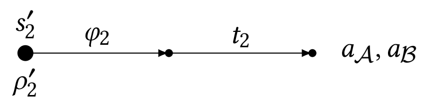


Figure 7.5: Syntactic lexical game g_2'

In s_2 in Figure 7.4, the speaker is conveying σ_2, although because φ_2 is semantically ambiguous, the addressee could potentially interpret the word as conveying σ_2' as well. This choice is represented by two branches labeled σ_2 and σ_2' emanating from the upper intermediate situation in the oval. The payoffs follow the same pattern as in g_1. In $s_{2'}$ in the lower half of the same game, a similar choice presents itself to the addressee but this time the payoffs are reversed because in this hypothetical situation $\mathcal{A}$ is conveying the other possible meaning σ_2'. Again, the primed a payoffs are greater than the corresponding primed c payoffs.

The prior probabilities for g_2 are $\rho_2 = P(\sigma_2 \mid x_1, y_1, y_2; u)$ and $\rho_{2'} = P(\sigma_2' \mid x_1, y_1, y_2; u)$ and these must sum to 1 as there are only two possible meanings the speaker could be conveying. Since these probabilities are also functions of the conditioning variables and parameter u, there are different *versions* of essentially the same game g_2 when x_1 takes on different values as y_1 and y_2 are just t_1 and t_2 and u is held fixed.

As I said above with g_1, $P(\sigma_2 \mid \sigma_1, t_1, t_2; u) > P(\sigma_2' \mid \sigma_1, t_1, t_2; u)$ because the presence of σ_1 is compatible with σ_2 as Bill Smith is more likely to be running in an election than running in a race. This is also reinforced by u as the topic or goal of conversation is the local election. On the other hand, $P(\sigma_2 \mid \sigma_1', t_1, t_2; u) < P(\sigma_2' \mid \sigma_1', t_1, t_2; u)$ because now σ_2 and σ_1' are incompatible as one involves running in an election and the other involves Bill Jones whereas σ_2' and σ_1' are compatible because Bill Jones is, in fact, a runner. This is so despite the presence of u which might bolster the former probability because incompatible meanings are much worse than an irrelevant statement. Once again, the parse trees do not play any role except to confirm all the probabilities as they are unique. If we assume the same numerical values as before, both players could be choosing between $(\rho_2, \rho_{2'}) = (5/6, 1/6)$, one version of g_2, and $(\rho_2, \rho_{2'}) = (1/4, 3/4)$, another version of g_2. Again, the actual numbers themselves are determined by a variety of objective and subjective factors that are exogenous but the choice between them is endogenous.

The same kinds of choices and payoffs occur in Figure 7.5 except that this time a trivial syntactic ambiguity has to be "resolved" as in g_1'. This time the only prior conditional probability is $\rho_2' = P(t_2 \mid x_1, x_2, y_1; u) = 1$ as before.

The games g_2 and g_2' are obtained by applying the same semantic and syntactic maps g_u and g_u' to φ_2. That is, $g_u(\varphi_2) = g_2$ and $g_u'(\varphi_2) = g_2'$. Again, as argued above, both maps are available to $\mathcal{A}$ and so the relevant games can be constructed without any further assumptions.

This completes the discussion of semantic and syntactic lexical games as constructed nonconsciously by $\mathcal{A}$. There are two semantic games and two syntactic games, each corresponding to the utterance of φ_1 or φ_2 in u. Moreover, as I have shown in detail, the games are *interdependent* because the solution to each of them depends on the solutions to the others via the prior probabilities which are conditioned by these solutions. In this example, the syntactic games were trivial and solvable immediately and so their dependence on other games was redundant. Later in Chapter 9, I will look at a more complex example with syntactic ambiguities as well. Because in general one cannot know which version of a semantic or syntactic game is going to be optimal without solving all the other lexical games, they have to be solved together. In this sense, these interlocking games are circular not only because each individual game involves an interdependence between speaker and addressee but also because no game can be solved without solving all the games simultaneously.

The next step is to consider the *products* of these lexical games to form phrasal and sentential games. In the very simple two-word sentence at hand, there are no phrasal games because we get the sentential games at once. I show the semantic sentential game $g_{12} = g_1 \otimes g_2$ in Figure 7.6.

There are several new things in this game product. First, the tree diagram is a "product" of the trees of g_1 and g_2.[7] Next, there are four initial situations $s_{12} = s_1 \cup s_2$, $s_{12'} = s_1 \cup s_{2'}$, $s_{1'2} = s_{1'} \cup s_2$, and $s_{1'2'} = s_{1'} \cup s_{2'}$. The latter situations s_1, $s_{1'}$, and s_2, $s_{2'}$ are the initial situations of the two semantic games g_1 and g_2 that are being multiplied. That is, the initial situations of the product are derived from the initial situations of the multiplicands as they should be and 2 x 2 = 4 such combinations are possible from the two initial situations of g_1 and the two initial situations of g_2.

I have abbreviated the sentence $\varphi = \varphi_1 \circ \varphi_2$ as $\varphi_1\varphi_2$ and the products $\sigma_1 \odot \sigma_2$, $\sigma_1 \odot \sigma_2'$, $\sigma_1' \odot \sigma_2$, and $\sigma_1' \odot \sigma_2'$ as $\sigma_1\sigma_2$, $\sigma_1\sigma_2'$, $\sigma_1'\sigma_2$, and $\sigma_1'\sigma_2'$, respectively. The speaker's single choice of utterance $\varphi_1\varphi_2$ is a concatenation of the individual utterances φ_1 and φ_2 from g_1 and g_2. And the addressee's possible interpretations are $\sigma_1\sigma_2$, $\sigma_1\sigma_2'$, $\sigma_1'\sigma_2$, and $\sigma_1'\sigma_2'$, where again the components are obtained from the respective games. The payoffs corresponding to the relevant branches in the multiplicands are added to give the payoffs in the relevant branch of the product. For example, the payoff for $\mathcal{A}$ in the topmost branch corresponding to the path s_{12}, $\varphi_1\varphi_2$, $\sigma_1\sigma_2$ is $a_{\mathcal{A}} + a_{\mathcal{A}}$, which is the sum of the $a_{\mathcal{A}}$ in g_1 corresponding to the path s_1, φ_1, σ_1 and the $a_{\mathcal{A}}$ in g_2 corresponding to the path s_2, φ_2, σ_2.

Finally, the four priors are $\rho_{12} = P(\sigma_1, \sigma_2 \mid y_1, y_2; u)$, $\rho_{12'} = P(\sigma_1, \sigma_2' \mid y_1, y_2; u)$, $\rho_{1'2} = P(\sigma_1', \sigma_2 \mid y_1, y_2; u)$, and $\rho_{1'2'} = P(\sigma_1', \sigma_2' \mid y_1, y_2; u)$ which must sum to 1 as they represent all the alternatives. Again, it should be easy to see how these probabilities are generated from the corresponding probabilities in the multiplicands g_1 and g_2 where the priors are $\rho_1 = P(\sigma_1 \mid x_2, y_1, y_2; u)$, $\rho_{1'} = P(\sigma_1' \mid x_2, y_1, y_2; u)$ and $\rho_2 = P(\sigma_2 \mid x_1, y_1, y_2; u)$, $\rho_{2'} = P(\sigma_2' \mid x_1, y_1, y_2; u)$. It is clear that $\rho_{12} > \rho_{1'2'}$ because both are conditioned by t_1, t_2, and u and because u is about the local

[7] This game tree product is, of course, different from the parse tree product described in Section 4.2.

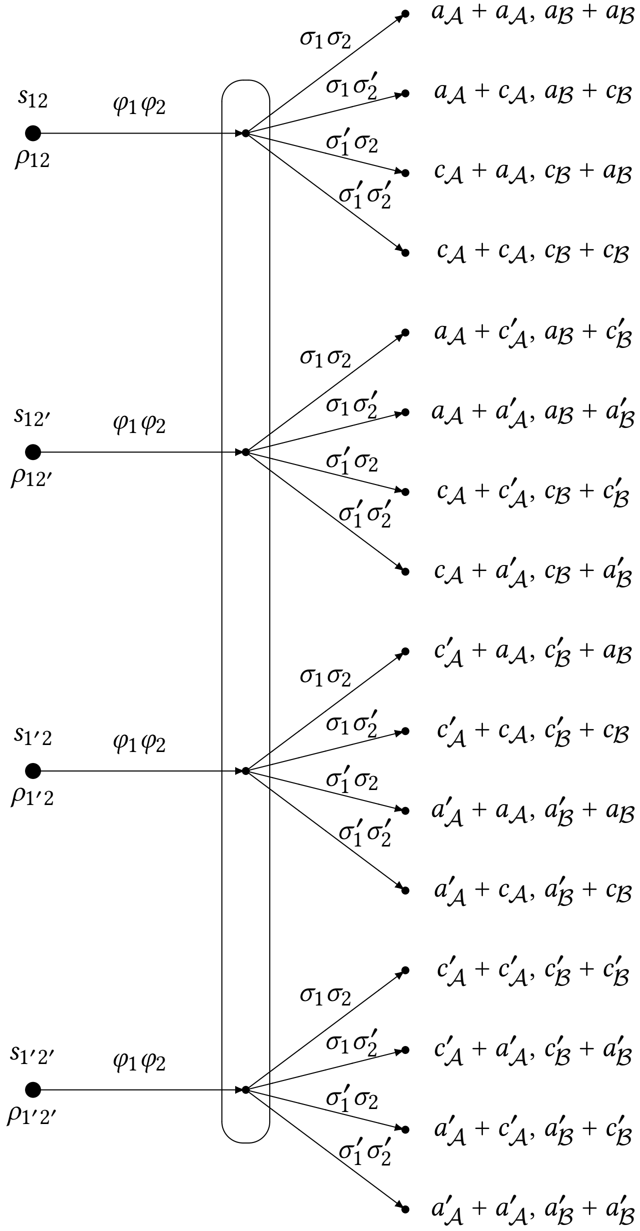


Figure 7.6: Semantic sentential game $g_{12} = g_1 \otimes g_2$

election it pushes up the former and pushes down the latter. The parse trees t_1 and t_2 are both unique and play no role. The "cross" probabilities $\rho_{12'}$ and $\rho_{1'2}$ are both very much lower than the other two because they represent incompatible interpretations, either Bill Smith the politician running in a race or Bill Jones the runner running in an election. So ρ_{12} is the highest among these four prior probabilities. A possible set of values could be $(\rho_{12}, \rho_{12'}, \rho_{1'2}, \rho_{1'2'}) = (0.5, 0.1, 0.1, 0.3)$. I will presently be showing how all these games are to be solved but I point out here that the prior probabilities play a key role.

The product game $g_{12} = g_1 \otimes g_2$ can also be obtained directly by applying the semantic map g_u to the sentence $\varphi_1 \circ \varphi_2$. When this is done, it becomes evident that $g_{12} \equiv g_u(\varphi_1 \circ \varphi_2) = g_u(\varphi_1) \otimes g_u(\varphi_2) \equiv g_1 \otimes g_2$.

The syntactic sentential game $g'_{12} = g'_1 \otimes g'_2$ is shown in Figure 7.7.

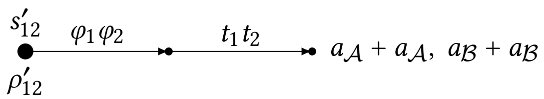


Figure 7.7: Syntactic sentential game $g'_{12} = g'_1 \otimes g'_2$

Each multiplicand has just one initial situation s'_1 or s'_2 and so there is just $1 = 1 \times 1$ initial situation $s'_{12} = s'_1 \cup s'_2$ in the product. The interpretation $t_1 t_2 = t_1 \star t_2 = t_{12}$ as shown earlier. The payoffs are the sum of the corresponding payoffs in the component games as before. The prior probability $\rho'_{12} = P(t_1, t_2 \mid x_1, x_2; u) = 1$ as the reader should now be able to work out. This is a trivial game and even though the prior probability is conditioned by other variables that take on different values depending on the optimal solution to the semantic games g_1, g_2, and g_{12}, it is always just 1 so the game is solvable immediately. As with the semantic sentential game, this syntactic product can be obtained directly by applying the syntactic map g'_u to $\varphi_1 \circ \varphi_2$ and so we have $g'_{12} \equiv g'_u(\varphi_1 \circ \varphi_2) = g'_u(\varphi_1) \otimes' g'_u(\varphi_2) \equiv g'_1 \otimes' g'_2$.

This is the Flow Constraint from the speaker's viewpoint. Recall that all the interlocking games constructed so far – g_1, g_2, g_{12} on the semantic side and g'_1, g'_2, g'_{12} on the syntactic side – actually carry a superscript $\mathcal{A}$ to mark this viewpoint. These are the games generated by the speaker when he considers the sentence $\varphi = \varphi_1\varphi_2 =$ BILL RAN as a possible thing to utter in order to convey $\sigma_1\sigma_2 =$ *Bill Smith stood (for election)* to $\mathcal{B}$ and thereby get across the whole content $\sigma =$ *Bill Smith ran in the local election* in order to evoke the response a, acceptance, by $\mathcal{B}$. This content $\sigma_1\sigma_2$, if indeed it is the content of an utterance of φ in u, something I have yet to show, would be called the *locutionary meaning* of the utterance because it comes directly from the words uttered in the situation u. As

I have said before, the sentence φ has no meaning without a context and does not have a conventional meaning at all. As the Semantic Constraint makes clear, only words have conventional meanings and they, together with the situation u, yield the locutionary meaning and, indeed, also the illocutionary meaning of the utterance which is the meaning that comes largely from the context u and only indirectly from the words uttered. This latter meaning, which includes things like free enrichment, modulation, figures of speech, implicature, and illocutionary force, will be addressed in Part IV.

I will postpone showing how these six locutionary games are solved because I first want to develop the complete model of locutionary communication. For the moment, just assume that their joint solution is in fact $\sigma_1\sigma_2$ rather than $\sigma_1\sigma_2'$, $\sigma_1'\sigma_2$, or $\sigma_1'\sigma_2'$, the other possibilities shown in Figure 7.6. I will label this locutionary meaning $\sigma^\ell = \sigma_1\sigma_2 =$ *Bill Smith stood (for election)*. The full illocutionary meaning will be labeled σ^ι so that the complete meaning of the utterance will be $\sigma = \sigma^\ell \odot \sigma^\iota =$ *Bill Smith ran in the local election*.

On the other hand, it is obvious that the equilibrium or optimal parse of the utterance is just $t = t_{12} = t_1t_2$ as indicated in Figure 7.7 as it is the only possibility. In more complex cases where syntactic ambiguities are present, the solution will be correspondingly more complex and syntax will also play a more active role in the overall determination of both locutionary and illocutionary contents and vice versa. Indeed, the syntactic and semantic (and phonetic) games all have to be solved together because they generally cannot be solved "locally" except when they are trivial. While the solution process involves a slightly elaborate description, it involves rather easy probabilistic computations that are straightforward to implement in our nonconscious (and semiconscious) neural processes.

The six interdependent *local* games that have been constructed so far can also be collected into a locutionary global game $LG_u(\varphi) = \{g_1, g_2, g_{12}, g_1', g_2', g_{12}'\}$ and we can say that when $\mathcal{A}$ considers how $\mathcal{B}$ might interpret an utterance of φ in u in his head he will construct $LG_u(\varphi)$. This interdependence of the games of partial information is another reason why they are not games of incomplete information (or signaling games). If f_u is a solution function that maps games into their solutions, we can also write, based on our assumption that the joint solution to the semantic locutionary games is $\sigma_1\sigma_2$ and to the syntactic locutionary games is t_1t_2, that $f_u[LG_u(\varphi)] = \{f_u(g_1), f_u(g_2), f_u(g_{12}), f_u(g_1'), f_u(g_2'), f_u(g_{12}')\} = \{\sigma_1, \sigma_2, \sigma_1\sigma_2, t_1, t_2, t_1t_2\}$. In other words, the solution to each local game, whether semantic or syntactic, is compatible with the solutions to all the other games because they must agree not only in the contents they deliver but also in the values of the conditioning variables that get set in the prior probability functions. For example, in the first lexical game g_1, the prior probability function

$\rho_1 = P(\sigma_1 \mid x_2, y_1, y_2; u) = P(\sigma_1 \mid \sigma_2, t_1, t_2; u)$ *in equilibrium*, which shows that the conditioning variables are σ_2, t_1, t_2 which all agree with the solutions to the other five games. Further, the equilibrium paths that will actually be followed in these six games will all begin with the actual situations, s_1, s_2, s_{12} in the three semantic games and s'_1, s'_2, s'_{12} in the syntactic games, because all the other initial situations are just counterfactual possibilities. And these paths will follow the branches that are consonant with the solutions listed above. If these paths are traced in all the trees then the payoffs actually delivered to the speaker would be $a_{\mathcal{A}}$, $a_{\mathcal{A}}$, $a_{\mathcal{A}} + a_{\mathcal{A}}$ in the three semantic games and $a_{\mathcal{A}}$, $a_{\mathcal{A}}$, $a_{\mathcal{A}} + a_{\mathcal{A}}$ in the three syntactic games. We can just take the maximum value $a_{\mathcal{A}} + a_{\mathcal{A}}$ of all these six numbers as the payoff that would be actually delivered to the speaker if he were to utter φ in u. Precisely what we do is somewhat arbitrary: the payoffs could also be added up to yield $8a_{\mathcal{A}}$ or just the lexical payoffs could be added to yield $4a_{\mathcal{A}}$. This is because of the affine nature of utilities mentioned in footnote 9 in Section 5.1.

There is one important matter that remains to be specified with respect to the products of games. I considered the products of two semantic games and of two syntactic games but why can't one multiply a semantic game with a syntactic game? After all, both yield *contents* and I have been saying that semantic and syntactic contents are to be treated uniformly. The answer is one can – by forming ordered pairs of possible semantic and syntactic contents. For example, the "pure" semantic game g_1 in Figure 7.2 would have (σ_1, t_e) in place of just σ_1 and (σ'_1, t_e) in place of just σ'_1 where t_e is just the empty tree. Likewise, the "pure" syntactic game g'_1 in Figure 7.3 would have $(\mathbf{1}, t_1)$ in place of just t_1 where $\mathbf{1}$ is just empty or no information. Then we can perform the product $g_1 \otimes g'_1 = g_{11'}$ which would have (σ_1, t_1) and (σ'_1, t_1) as the interpretive choices for $\mathcal{B}$. This shows that the locutionary global game actually contains many more local games. Indeed, it is not enough to consider just ordered *pairs* of contents but we need $2n$-tuples where n is the number of words in the sentence. This is because we can also form products like $g_1 \otimes g'_2 = g_{12'}$ as well as $g_{12} \otimes g'_{12} = g_{121'2'}$, the full mixed sentential product.

I will not bother to list all of them as there are too many permutations of the indices. In fact, the upper bound for the number of these games is $2^{2n} - 1 = 15$ when $n = 2$.[8] So one can see that the number of games grows quite rapidly

[8]It is instructive to work out this formula to check one's understanding of these games. In general, there will be $2n$ indices, 1, 2, 3, …, n for semantic lexical games and and $1'$, $2'$, $3'$, …, n' for syntactic lexical games. These indices form a set of $2n$ elements which has 2^{2n} subsets. And as there is no null subset involved, we subtract 1, giving us the formula $2^{2n} - 1$. The reason it is an upper bound is that there will be many zero products and the corresponding games would be discarded.

with n, the length of the sentence. I will show in the next chapter that the only product of semantic and syntactic games that matters is the full mixed sentential product so I will ignore the other mixed products. Remarkably, the only games needed to understand an utterance turn out to be the pure lexical games, a vast computational simplification resulting from our rationality.

7.3.4 Back to the Generation Game

Thus, our conclusion so far is that, based on the Content Selection Game induced by the Setting Game in u, $\mathcal{A}$ resolves to convey σ in u in order to evoke the response a from $\mathcal{B}$, and then, in the Generation Game that is induced by the Content Selection Game, casts about for a sentence that would do the job, finds φ as one such sentence by some assumed conversion process from σ to φ, constructs $LG_u(\varphi)$ in his head as described by the Phonetic, Syntactic, Semantic, and Flow Constraints, solves it with the map f_u, checks that the equilibrium content is compatible with σ, and receives a value $v_u^{\mathcal{A}}[LG_u(\varphi)] = a_{\mathcal{A}} + a_{\mathcal{A}}$. We say the *value* of $LG_u(\varphi)$ to $\mathcal{A}$ is $a_{\mathcal{A}} + a_{\mathcal{A}}$. This entire computation, or some approximate version of it, presumably takes only milliseconds to perform in a speaker's head and happens nonconsciously. In actual fact, it is not just $LG_u(\varphi)$ that is constructed but also $IG_u(\varphi)$, the global set of local illocutionary games corresponding to the illocutionary meanings of the utterance, whose description I have postponed to Part IV. In other words, it is the global game $G_u(\varphi) = LG_u(\varphi) \cup IG(\varphi)$ that the speaker identifies when he contemplates uttering φ and it is the global game's value $v_u^{\mathcal{A}}[G_u(\varphi)]$ that he computes. I realize this is a lot of terminology but once we have set it up, the discussion will flow very smoothly.

There is one other factor that enters into $\mathcal{A}$'s consideration of φ. That is the *cost* of uttering φ, $k_u^{\mathcal{A}}(\varphi)$, which is based on a variety of factors, mostly objective ones such as the length and complexity of the sentence and the effort required to mentally process it and to physically produce it but also subjective ones especially related to the consequences of making certain information explicit and leaving certain information implicit. These subjective costs have to do partly with aesthetic things like conversational style but also with how making certain things explicit may affect the relationship with the addressee in positive or negative ways depending on whether these explicit things are polite or impolite and how leaving certain things implicit may result in costly errors. The mental processing involves playing the global game and determining its value. The *net* value $\mathcal{A}$ would receive from uttering φ is then $v_u^{\mathcal{A}}(\varphi) = v_u^{\mathcal{A}}[G_u(\varphi)] - k_u^{\mathcal{A}}(\varphi)$.[9] This

[9]Here, I have used the same symbol $v_u^{\mathcal{A}}$ for a function of φ and for a function of $G_u(\varphi)$ to avoid multiplying symbols needlessly. Also, note that this symbol is different from $v_{\mathcal{A}}$, the payoff to $\mathcal{A}$ shown in the Content Selection Game in Figure 7.1.

finally completes my description of what happens when $\mathcal{A}$ imagines uttering φ in u. Essentially, he computes its net value $v_u^{\mathcal{A}}(\varphi)$ based on the value $v_u^{\mathcal{A}}[G_u(\varphi)]$ of the global game determined by the four Constraints and its cost $k_u^{\mathcal{A}}(\varphi)$. I will consider the details of costs in Section 8.4.

In the same way, $\mathcal{A}$ may compute the net values of other sentences he could utter in u to convey σ, possibly $\varphi' =$ BILL SMITH RAN and $\varphi'' =$ BILL RAN IN THE ELECTION. These net values would be determined by the global games $G_u(\varphi')$, $G_u(\varphi'')$ and the associated costs $k_u^{\mathcal{A}}(\varphi')$, $k_u^{\mathcal{A}}(\varphi'')$ of the two sentences. Each of these global games would involve the corresponding lexical, phrasal, and sentential semantic and syntactic speaker games (as well as the relevant illocutionary games) and their global solutions.

On the other hand, the limits of rationality mentioned earlier may make him abandon his search for more sentences if he feels sufficiently satisfied with the best net value he has computed so far. This is a trial and error process as becomes especially clear when one is writing rather than speaking as then there is ample scope to try out different sentences to express roughly the same content. Indeed, when writing is considered, it becomes obvious that there is an interdependence between the Content Selection Game and the Generation Game because a new sentence with a slightly improved content may suggest itself in the Generation Game and then the new content would have to be evaluated in an altered Content Selection Game.

In any case, he considers one or more sentences and this leads to the choice he faces in the Generation Game $GG_u(\sigma)$ as shown in Figure 7.8.

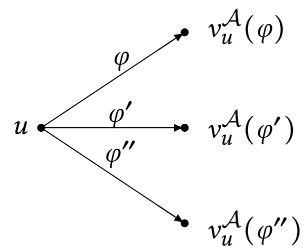


Figure 7.8: Partial Generation Game $GG_u(\sigma)$

This may appear superficially like a one-person decision problem but keep in mind that the net values are based on the values of the corresponding global games which involve both $\mathcal{A}$ and $\mathcal{B}$. $GG_u(\sigma)$ is so far just in $\mathcal{A}$'s head. His solving it involves choosing the sentence with the highest net value and, finally, publicly uttering it in u.

I will assume φ is in fact the best sentence to utter in u because it conveys the same content σ as the other sentences with a lower cost. Its cost is lower because it is shorter and simpler than the other sentences and the physical and mental effort involved is therefore lower (even though there may not be any illocutionary games required for the other sentences because all the information has been made explicit in them). Also, if the ongoing conversation is about the local election, then aesthetic considerations might imply that repeating the information about the election explicitly is inelegant. For example, if $\mathcal{B}$ has just referred to the election explicitly in her prior utterance to $\mathcal{A}$, then the election is already very prominent in the context. These sorts of things lead to the use of pronouns, for instance. There may be no very weighty relationship-related dimensions involved and no politeness considerations in this kind of purely informational and casual exchange. I will consider such psychosocial factors in Sections 8.6 and 18.2 and Chapter 11 as the first task is to lay out the model in its simplest form.

I end this section by reiterating the point made in Section 3.1 that sharply separating the Content Selection Game from the Generation Game is just a useful idealization. The two games are interdependent because identifying the optimal content to convey leads to the search for the optimal sentence to utter and vice versa. For example, in considering $GG_u(\sigma)$, $\mathcal{A}$ may opt for something more tentative and less assertive such as η = I THINK BILL RAN based on his own and his addressee's psychology as well as on their interaction so far. This would take him back to an altered Content Selection Game with a slightly different content corresponding to η and so on. And, indeed, bits of content and bits of language rather than entire sentences may go back and forth between the two games with alterations in both games as new possibilities arise. Speaking and writing are ultimately creative acts so there can be no complete closure to this process. That is part of the reason why I do not as yet have a model of how sentences are generated from contents: the map is partly open-ended as it depends on many factors including the speaker's mastery of the language.

To the best of my knowledge, most models in semantics, including my own earlier work, do not make the distinction between content selection and generation. Even computational studies of generation (e.g. Jurafsky & Martin 2009: Chapter 24) seem to omit the content selection problem and simply start with a given content. Thus, while the sharp separation referred to above is an idealization, the key point is to recognize that *two* choices, not one, are involved, the choice of content and the choice of appropriate sentence to convey the content, whatever the particular manner in which the two are effected, whether separately as I have shown or together via an ordered pair in a single more complex game.

7.4 The Interpretation Game

Once $\mathcal{A}$ has publicly uttered φ, $\mathcal{B}$ presumably hears (or reads) it. Based on her using the same four Constraints – Phonetic, Syntactic, Semantic, and Flow – that the speaker has used, this leads to the public emergence of a game between $\mathcal{A}$ and $\mathcal{B}$. There are, in fact, three games: one is the actual game $G_u(\varphi)$ that comes into existence and the other two $G_u^{\mathcal{A}}(\varphi)$, $G_u^{\mathcal{B}}(\varphi)$ are $\mathcal{A}$'s and $\mathcal{B}$'s models of their interaction. That is, the local games contained in the global games have three aspects and each of them derives from the game maps g_u, $g_u^{\mathcal{A}}$, and $g_u^{\mathcal{B}}$. The "objective" map g_u is not in either agent's private information but can be said to exist on account of the utterance. The construction of $G_u^{\mathcal{B}}(\varphi)$ is very similar to the construction of $G_u^{\mathcal{A}}(\varphi)$ above. Just as $\mathcal{A}$'s locutionary global game

$$LG_u^{\mathcal{A}}(\varphi) = \{g_1^{\mathcal{A}}, g_2^{\mathcal{A}}, g_{12}^{\mathcal{A}}, g_1'^{\mathcal{A}}, g_2'^{\mathcal{A}}, g_{12}'^{\mathcal{A}}\}$$

so $\mathcal{B}$'s locutionary global game is

$$LG_u^{\mathcal{B}}(\varphi) = \{g_1^{\mathcal{B}}, g_2^{\mathcal{B}}, g_{12}^{\mathcal{B}}, g_1'^{\mathcal{B}}, g_2'^{\mathcal{B}}, g_{12}'^{\mathcal{B}}\}$$

and the objective locutionary global game is

$$LG_u(\varphi) = \{g_1, g_2, g_{12}, g_1', g_2', g_{12}'\}.$$

$\mathcal{B}$ receives just this one sentence φ and does not need to consider other alternatives such as φ' or φ'' as $\mathcal{A}$ did in $GG_u(\sigma)$. So the Interpretation Game is just $UG_u(\varphi) = G_u^{\mathcal{B}}(\varphi)$.[10] Thus, there is an asymmetry between $\mathcal{A}$'s and $\mathcal{B}$'s effort in communication: the speaker generally has to work harder. Incidentally, this slight time lag between $G_u^{\mathcal{A}}(\varphi)$ being played in $\mathcal{A}$'s head and $G_u(\varphi)$ appearing publicly and $G_u^{\mathcal{B}}(\varphi)$ emerging in $\mathcal{B}$'s head after the utterance and the possibility of there not being common knowledge is a third reason why these games of partial information are not games of incomplete information (or signaling games).

The objective game $G_u(\varphi)$ is made up of those parts of $\mathcal{A}$'s and $\mathcal{B}$'s subjective games $G_u^{\mathcal{A}}(\varphi)$ and $G_u^{\mathcal{B}}(\varphi)$ that are *known* to each of them. For example, in each game in $G_u^{\mathcal{A}}(\varphi)$ there will be a part that involves $\mathcal{A}$'s own possible actions which would clearly be known to him. And there will be a part where he models $\mathcal{B}$'s possible actions which may or may not be accurately known to him. On the other hand, in each game in $G_u^{\mathcal{B}}(\varphi)$ there will be a part that involves $\mathcal{B}$'s own possible actions which would clearly be known to her. And there will be a part where she

[10] I use the letter U as a mnemonic for "understanding" as the letter I has already been used for the illocutionary global game $IG(\varphi)$.

models $\mathcal{A}$'s possible actions which may or may not be accurately known to her. $G_u(\varphi)$ will then consist of the parts of each interlocutor's subjective games that are known to them, $\mathcal{A}$'s own possible actions from $G_u^{\mathcal{A}}(\varphi)$ and $\mathcal{B}$'s own possible actions from $G_u^{\mathcal{B}}(\varphi)$. This kind of setup for interpretation is more general than Grice's understanding, for example, as it makes possible a formal modeling of miscommunication and weaker flows of information in a natural way.

If all goes well, $G_u(\varphi) = G_u^{\mathcal{A}}(\varphi) = G_u^{\mathcal{B}}(\varphi)$ and, indeed, they all become (nonconscious) common knowledge between $\mathcal{A}$ and $\mathcal{B}$. So far, I have described just the locutionary global game $LG_u^{\mathcal{A}}(\varphi)$ in the previous section and said that the illocutionary global game $IG_u^{\mathcal{A}}(\varphi)$ will be developed in Part IV. But it is the global game $G_u(\varphi) = LG_u(\varphi) \cup IG_u(\varphi)$ that unfolds publicly and $G_u^{\mathcal{B}}(\varphi) = LG_u^{\mathcal{B}}(\varphi) \cup IG_u^{\mathcal{B}}(\varphi)$ that is induced privately in $\mathcal{B}$. I will generally assume that the objective and subjective games are all identical and common knowledge as this is what happens when the communication succeeds. But there are situations where there is miscommunication or a weaker flow of information and this is then explained by the divergence among these three games or by the lack of common knowledge.

Prior to the utterance, $\mathcal{A}$ merely believed (nonconsciously) that the emergent game $G_u(\varphi)$ would be common knowledge. When this belief turns out to be informationally caused,[11] that is, when it is knowledge, then $G_u(\varphi)$ will be common knowledge between $\mathcal{A}$ and $\mathcal{B}$. This would mean that $\mathcal{B}$'s corresponding (nonconscious) belief that $G_u(\varphi)$ is common knowledge would also be informationally caused. In this sense, common knowledge of the global game is not something that is exogenously assumed as is usually the case in applications of game theory but it is something that issues from the utterance of φ in the ambient utterance situation u when communication succeeds, and fails to obtain when there is miscommunication or when there is a flow of information that is weaker than full communication. There could be a variety of factors responsible for miscommunication, either differences in the prior probability estimates each agent uses in each local game belonging to the global game or differences in the Phonetic, Syntactic, or Semantic Constraint outputs (e.g. the addressee might consider an additional word or parse or meaning or might fail to consider some possibility owing to differences in the agents' knowledge of the language) or other differences in their subjective games. Whatever the factor, common knowledge of the game will then not obtain. I have considered miscommunication in some detail earlier so I will not deal with it in this book.[12] In the case of weaker information flows, the content does get through but it does not become common knowledge.

[11] Here "information" is used in Dretske's sense of being a true proposition. See Section 2.1.

[12] See Parikh (2001: Chapter 9) and Parikh (2010: Section 5.3).

This happens when, for example, $\mathcal{A}$ suggests or hints something. I have also considered such weaker flows earlier though I will address them again briefly in Chapter 20.[13] Generally, I will just assume throughout that common knowledge does materialize.

Incidentally, common knowledge of $G_u(\varphi) = G_u^{\mathcal{A}}(\varphi) = G_u^{\mathcal{B}}(\varphi)$ is not so hard to come by because the utterance itself is generally public and therefore common knowledge, the language is sufficiently shared so that the Syntactic and Semantic Constraints are (nonconscious) common knowledge, and therefore the possible contents considered in the various local games are common knowledge, and the payoffs are determined endogenously from $\mathcal{A}$'s possible intentions which are given by the possible contents, which leaves just the prior probabilities. These are specified by a host of subjective and objective factors and need to be just roughly shared. However, they are the one item that can go awry and so when there is miscommunication it is more often than not because the probabilities were differently assessed by the interlocutors. As I said in Section 2.2, there are sufficient grounds for saying that often enough there will be a situation $s \subset u$ such that (a) $s \models G_u(\varphi)$, (b) $s \models \mathcal{A}$ knows s, and (c) $s \models \mathcal{B}$ knows s, which means that often enough $\mathcal{A}$ and $\mathcal{B}$ will have common knowledge of $G_u(\varphi)$, especially its locutionary part $LG_u(\varphi)$, and, consequently, (partial) communication will succeed.

As before, I will postpone discussing the solution process to the Interpretation Game. Indeed, the solution process is a joint one between $\mathcal{A}$ and $\mathcal{B}$ and is carried out against the background of $G_u(\varphi)$, the objective global game. Given the assumed identity of the three global games, we can say for the moment that their joint solution is $\sigma^\ell = \sigma_1\sigma_2$ = *Bill Smith stood (for election)* and $t = t_{12} = t_1 t_2 = [_{\mathrm{S}}[_{\mathrm{NP}}\,\mathrm{Bill}][_{\mathrm{VP}}\,\mathrm{ran}]]$ and the full meaning $\sigma = \sigma^\ell \odot \sigma^\iota$ = *Bill Smith ran in the local election*. In other words, both $\mathcal{A}$ and $\mathcal{B}$ derive the same semantic and syntactic content from the utterance.

After solving $UG_u(\varphi)$, $\mathcal{B}$ has understood the utterance. In a narrow sense of communication, this is all that is involved. But in a wider sense, she still has to respond to the content. When I defined communication in Section 5.3, it was the narrow sense that mattered as communication is not concerned with how the addressee responds to the utterance even though her possible responses in the interactive structure we have called the Content Selection Game influence the content and therefore the sentence $\mathcal{A}$ chooses to convey.

As I said in Parikh (1987b: 46), the interpretive act by $\mathcal{B}$ in the Interpretation Game is not observable by $\mathcal{A}$. As a result, given just this much, $\mathcal{A}$ cannot tell if

[13] See Parikh (2001: Section 6.5) and Parikh (2010: Section 5.10).

the communication was successful. It is only $\mathcal{B}$'s response in the Content Selection Game, if it is an observable act, that allows $\mathcal{A}$ to infer how $\mathcal{B}$ might have interpreted his utterance. Otherwise, he would need to wait for some public action such as an utterance by $\mathcal{B}$ in an ongoing conversation of the kind we are modeling or some other kind of perceptible action to infer how she might have understood him. Even with such observations, there will generally remain some uncertainty about $\mathcal{B}$'s precise interpretation just as $\mathcal{B}$ will remain unsure about precisely what $\mathcal{A}$ was conveying because there are often fundamental indeterminacies in communication as discussed in detail in Parikh (2010: Chapter 5). Such indeterminacies arise from a variety of sources as I pointed out there, from possibly different utterance situations the agents carve out (and concomitant differences in the games induced) as well as differences in their knowledge of the language that includes the phenomenon of vagueness to be discussed in Chapter 11. In other words, the divergences between what was conveyed and understood are related to uncertainties each agent has about the content in the other agent's mind.

I have tried to tread a fine line in the preceding paragraphs. On the one hand, common knowledge is, often enough, not that hard to come by, but, on the other, it also does often fall short of what is required for an exact identity between what is conveyed and what is understood. More of the locutionary content will generally be identical as opposed to the illocutionary content but even with the former there are differences that arise, especially on account of vagueness.

7.5 The Content Selection Game again

Based on the content σ that $\mathcal{B}$ infers from $G_u(\varphi)$, she constructs the Content Selection Game shown in Figure 7.1 in Section 7.2 for the first time. Until this point in time only $\mathcal{A}$ has played the game as a private matter. He has had to convert the optimal content σ into a public utterance of φ in u in the Generation Game to convey σ to $\mathcal{B}$. She then infers σ from the utterance in the Interpretation Game and so comes to construct the Content Selection Game. This is part of the circularity in communication. In general, as I remarked in Section 7.2, her model of the Content Selection Game may differ from $\mathcal{A}$'s and there will again be the same circumstance of an objective game and two subjective games. But, to simplify things, I assume common knowledge of a single game. In the trivial interaction at hand, $\mathcal{B}$ has just one action a, acceptance, to choose from, so she simply carries out that action. Even here, this response is invisible to $\mathcal{A}$ and so he has to wait for some other public action by $\mathcal{B}$ to infer not just the meaning she derived from his utterance but also whether she accepted this meaning or not. Of course, in this trivial game, $\mathcal{A}$ knows that his content will be accepted because that is the only available response.

When describing the Interpretation Game in the previous section, I had pointed to an asymmetry between $\mathcal{A}$ and $\mathcal{B}$. The speaker has to potentially consider multiple sentences in the Generation Game as indicated by Figure 7.8 but the addressee need consider just the single utterance selected by the speaker. This means the addressee never has to consider alternative *utterances* as is commonplace in Gricean pragmatics (e.g. Horn 1972; 1984, Levinson 1983; 2000). However, as will become clear in Part IV, she does have to consider alternative *contents* when the Content Selection Game presents more complex possibilities. The reason for this somewhat subtle difference is that while it is reasonable for addressees to reflect on *what* a speaker may be conveying, it is idle for them to concern themselves with *how* a speaker may be conveying it as many different ways of expressing the same content may exist. And, given the compositional nature of generation and interpretation, this can lead to an inefficient regress in processing, as I argue in Section 8.2. It is because much Gricean work is not fleshed out in as much detail as Equilibrium Semantics provides that it becomes possible to ignore this issue.

Once $\mathcal{B}$ has responded, the circle of communication is over and both agents return to the Setting Game where the whole enterprise began. All of this generally takes no more than a few seconds in thought and speech for $\mathcal{A}$ and comprehension and response for $\mathcal{B}$ although a writer may take more time to formulate his utterance and a reader may take more time to digest a sentence. All that has been accomplished here is that $\mathcal{A}$ has communicated σ to $\mathcal{B}$ and she has accepted his communication.

Below, I will look at two slightly more complex Content Selection Games where both agents have a real choice of action.

7.6 Back to the Setting Game

Once they are back in the Setting Game where they are discussing the local election and perhaps politics generally, $\mathcal{A}$'s wish to elicit a certain sort of response from $\mathcal{B}$ has been more or less fulfilled. They are now ready to begin the next round of their conversation where $\mathcal{B}$ might, on the basis of what she has just heard and the setting they are in, form her own wish to elicit a corresponding response from $\mathcal{A}$, leading to a new Content Selection Game. If we were interested in modeling dialogue, this is how it would have to be done, allowing for cross-references to the shared history the interlocutors build up. This shared history, and particularly the various objects the agents have referred to, become salient for future utterances in the dialogue via what is called an *information state* which maintains a record of the past contents conveyed. There are naturally three such states – one objective one and two subjective ones which may be more or less identical and common knowledge.

I have implicitly described a simpler kind of dialogue where each participant utters just one sentence at a time but there is no difficulty in extending such a picture to alternating *discourses* by each agent. It is possible to allow even more general settings where the agents might be performing other nonverbal actions such as gestures or interrupting the flow by a side exchange with some other party or by some other intervening acts. All such settings involve a sequence of utterance situations rather than a single one. Such sequences are called *discourse* or *dialogue* situations and are denoted by d.

I have so far not discussed the entire content conveyed in a communication, just the infon σ and the parse tree t, the latter being transmitted nonconsciously. What actually happens is that an utterance describes a situation c *as* supporting or containing a content σ. That is, what is conveyed is the *proposition* $p = (c \vDash \sigma)$ where c is called the described situation.[14] The content σ is tied to some situation that is also communicated. For example, if $\mathcal{A}$ and $\mathcal{B}$ are talking about New York politics generally, then the situation of New York politics will be the described situation. If they are discussing just the local election, then the situation involving the local election will be the described situation. I have said above that the full content $\sigma = \sigma^{\ell} \odot \sigma^{\iota}$ = *Bill Smith ran in the local election* and this implies that σ^{ι} = *in the local election* is an illocutionary content of the utterance. This in turn suggests that the described situation is some larger situation, either covering New York politics (c) or even politics generally (c') or possibly New York generally (c''). But I could equally have said that the described situation is one involving the local election (c_0) in which case there is no need for the illocutionary content σ^{ι} as it would be redundant. In other words, the described situation is generally quite indeterminate and is seldom fully shared by the agents. Further, whether something is an illocutionary content and occurs to the right of the turnstile in the proposition conveyed or whether it is just a part of the described situation and so occurs to the left of the turnstile is often indeterminate. That is, the proposition conveyed could be any of $c \vDash \sigma$ or $c' \vDash \sigma$ or $c'' \vDash \sigma$ or $c_0 \vDash \sigma^{\ell}$ or even other possibilities.

As I said in Section 2.1, it is agents who carve out situations and it is agents who determine the boundaries of what they are discussing. These boundaries are

[14] I have used c as a mnemonic for the situation supporting a content as the letter d is reserved for the discourse or dialogue situation.

generally fuzzy and differently identified by the two agents. These differences are in fact quite productive as they lead the interlocutors to pursue different possibilities in what they contribute to the conversation as different contents suggest themselves to each agent. If one agent thinks they are discussing politics generally (c'), he may shift the discussion to the presidential election; if the other agent thinks they are discussing New York generally (c''), she may shift the discussion to the excesses of Wall Street. As such, the described situation also keeps changing through a conversation, getting larger or smaller or going off in a different direction altogether.

It is propositions that are true or false (or indeterminate) so that if the described situation had for some reason been a local election in Mumbai (c_0') then the proposition $c_0' \models \sigma^\ell$ that Bill Smith ran in the local election in Mumbai would have been false. Both versions of the proposition $c \models \sigma$ and $c_0 \models \sigma^\ell$ may well be true. This raises the question how the described situation is even roughly determined. Where does it come from? It is partly inferred from u and the larger dialogue situation d, it partly issues from σ itself, and it is partly just creatively constructed by each agent. There is considerable latitude in where its boundaries are fixed and what matters to the communication is just whether it supports the content σ or not. I will say more about this when I look at how truth plays a role in meaning in Section 10.4.

For the time being, I just point out that the utterance and dialogue situation are different from the described situation although both are part of the context as I said in Section 2.2. This is a more general and, in my view, more accurate setup than the framework of a context set introduced by Stalnaker (1998) which plays both roles of utterance situation and described situation and consequently has fewer degrees of freedom. As discussed in some detail in Parikh (2010: 65–66), if d is made up of a sequence of utterances each taking place in utterance situations $u_1, u_2, \ldots$, then, just as u_1 contributes to constructing c_1, c_1 in turn contributes to modifying u_1 to establish the next utterance situation u_2, which modifies c_1 and generates the next described situation c_2, and so on.

7.7 The Communication Game

We have come full circle and we can now appreciate the explanation for $\mathcal{A}$'s utterance and $\mathcal{B}$'s response. First, there is the setting u in which everything takes place. Part of this setting is the Setting Game which, in this case, is just a casual conversation between $\mathcal{A}$ and $\mathcal{B}$ about a local election. This induces the Content Selection Game for $\mathcal{A}$, maybe because it is $\mathcal{A}$'s turn to contribute something to the dialogue. $\mathcal{A}$ solves this game and identifies the equilibrium content σ to convey to $\mathcal{B}$ in order to evoke some response a in her, something as minimal as reception or acceptance in this example. After this selection, $\mathcal{A}$ plays the induced Generation Game and converts σ into possible sentences he could utter in u to convey this content. One of these is φ. He evaluates φ by mentally building a model of the corresponding global game determined by the four Constraints, Phonetic, Syntactic, Semantic, and Flow, and computing its net value based on the global game's value and the cost of the utterance. After possibly evaluating other candidates as well, he chooses the optimal sentence and publicly utters it. Upon hearing the utterance, $\mathcal{B}$ builds her mental model of the global game called the Interpretation Game based on the same four Constraints and solves it. If the communication is to succeed, the three global games – the objective game and the two subjective models of it – have to be identical and common knowledge and $\mathcal{B}$ arrives at the content σ. Based on this, she plays the Content Selection Game and chooses her best response, just acceptance in this example. Finally, both return to the Setting Game for further rounds of the conversation.

Thus, $\mathcal{A}$, in attending to the Setting Game, starts with the Content Selection Game and then plays the Generation Game. $\mathcal{B}$ starts with the Interpretation Game and then plays the Content Selection Game. This double set of choices that both agents have to make – selecting a content and then the utterance for the speaker and choosing an interpretation and then an action for the addressee – are quite reasonable once one realizes that one choice in each set of choices would have been required even if telepathy were possible. However, in practice, all this happens nonconsciously and seamlessly, possibly going back and forth multiple times among the various games as small chunks rather than whole utterances are processed in real time.

This is more or less the full structure of communication. There can be many variations of it. It can be seen to be qualitatively quite rich but each part of it is relatively simple. That is, its building blocks are uncomplicated and easily understood but the interactions among them make the overall system quite complex. This is exactly what we should expect intuitively and, as will be seen below, the solution to all these games is also quite straightforward and readily implementable either in human neural structures or in artificial agents such as robots.

We can now go back to the chart that was displayed in Section 3.1.

<u>Utterance Situation</u>

Setting Game

→ $\mathcal{A}$'s wish to elicit some response from $\mathcal{B}$

→ Content Selection Game

→ $\mathcal{A}$'s equilibrium content

→ Generation Game

→ $\mathcal{A}$'s equilibrium utterance

→ Interpretation Game

→ $\mathcal{B}$'s equilibrium content

→ Content Selection Game

→ $\mathcal{B}$'s equilibrium response

→ Back to the Setting Game

It should now be very clear how this cycle of communication works for just a single utterance. It is useful to assign symbols to all the games involved, so we denote the Setting Game by SG_u and the Content Selection Game by CSG_u. The Generation Game already has the name $GG_u(\sigma)$ and the Interpretation Game has the name $UG_u(\varphi)$. In general, there will be other contents $\mathcal{A}$ will contemplate in CSG_u but it is just the optimal one σ that induces the corresponding Generation Game $GG_u(\sigma)$. Likewise, there may be multiple candidate sentences $\mathcal{A}$ may consider in $GG_u(\sigma)$ but it is only the optimal utterance of φ that induces the Interpretation Game $UG_u(\varphi)$. So we can leave these optimal arguments in there as part of the definition of the whole game called the Communication Game which has these four interlinked games as components. The Communication Game is then $\Gamma_u = (SG_u, CSG_u, GG_u(\sigma), UG_u(\varphi))$. But the content σ and sentence φ can also be suppressed if desired to yield just $\Gamma_u = (SG_u, CSG_u, GG_u, UG_u)$. Recall that the global games $G_u^{\mathcal{A}}(\varphi) = LG_u^{\mathcal{A}}(\varphi) \cup IG_u^{\mathcal{A}}(\varphi)$, $G_u^{\mathcal{B}}(\varphi) = LG_u^{\mathcal{B}}(\varphi) \cup IG_u^{\mathcal{B}}(\varphi)$ are part of $GG_u(\sigma)$ and $UG_u(\varphi)$, respectively. I have yet to describe the illocutionary game component of these global games which will be taken up in Part IV. There is also an objective global game $G_u(\varphi) = LG_u(\varphi) \cup IG_u(\varphi)$ that is induced by $\mathcal{A}$'s public utterance of φ which I have deliberately left out of the chart above and from the definition of Γ_u. It forms a kind of backdrop to the subjective games that the agents actually solve but what happens is determined by the objective game which is the actual game that gets played as the subjective games are just partial models of it. Of course, we will be largely concerned with situations where all three games are identical. If we wish, this missing objective game can be made explicit by defining $\Gamma_u = (SG_u, CSG_u, GG_u(\sigma), UG_u(\varphi), G_u(\varphi)) = (SG_u, CSG_u, GG_u, UG_u, G_u)$. Finally, each global game is made up of several local games that are either speaker games or addressee games or the corresponding objective games. These are called games of partial information and it is their joint solution that plays a crucial role in the process of communication.

At this stage, we are in a position to say that communication (in its wide sense) basically involves the playing of the Communication Game Γ_u in some situation u. It is as simple or as complex as that, whichever way you wish to see it.

8 Solving Communication Games

8.1 Solving Locutionary Global Games

Assuming common knowledge of the objective locutionary global game $LG_u(\varphi) = \{g_1, g_2, g_{12}, g_1', g_2', g_{12}'\}$ from Section 7.3 to keep things simple, I now look at how this game can be solved. When there is common knowledge, $LG_u^{\mathcal{A}}(\varphi) = LG_u^{\mathcal{B}}(\varphi) = LG_u(\varphi)$, so solving the objective game automatically implies a solution to the subjective games. As I said in that section, f_u is a solution function that maps games into their solutions and $f_u[LG_u(\varphi)] = \{f_u(g_1), f_u(g_2), f_u(g_{12}), f_u(g_1'), f_u(g_2'), f_u(g_{12}')\} = \{\sigma_1, \sigma_2, \sigma_1\sigma_2, t_1, t_2, t_1 t_2\}$. The solution to the syntactic games can be trivially obtained in this example so I will focus on the semantic games.

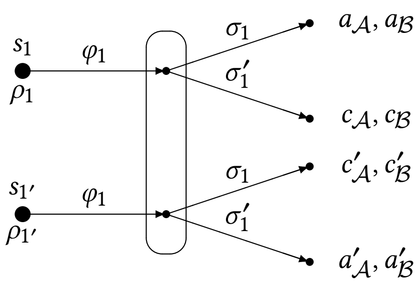


Figure 8.1: Semantic lexical game g_1

Recall the first local semantic game $g_1 = g_u(\varphi_1)$ shown again in Figure 8.1. In g_1, $\mathcal{A}$ utters φ_1 = BILL and $\mathcal{B}$ has to choose between σ_1 = *Bill Smith* or σ_1' = *Bill Jones*. The prior $\rho_1 = P(\sigma_1 \mid x_2, y_1, y_2; u)$ where the x_2 is a variable ranging across the possible corresponding meanings σ_2, σ_2', and the y_1, y_2 are variables standing for the possible corresponding parse trees t_1, t_2, respectively. Likewise, the other prior $\rho_{1'} = P(\sigma_1' \mid x_2, y_1, y_2; u)$. Thus, there are different *versions* of essentially the same game g_1 when x_2, y_1, or y_2 take on different values whenever this is possible. In addition to choosing utterances and interpretations, the agents also have to choose the particular values of the conditioning variables in the prior probabilities, that is, they have to choose between the two versions of the game. This is what makes all the local games interdependent.

In solving such games, the players have to identify all the possible strategies they could adopt and then choose the best one based on what the other player might do. A strategy is just a specification of what an agent would do at each of their information sets and of what version of the game they would choose. $\mathcal{A}$ has two information sets, one at s_1 and the other at $s_{1'}$, so he has to specify what he would do at both of these. However, he has just one choice of action, φ_1, in each set so there is no real choice. On the other hand, he can choose between the two versions of g_1 essentially by choosing $x_2 = \sigma_2$ or $x_2 = \sigma_2'$, as the other conditioning variables y_1 and y_2 take on just one value t_1 and t_2 respectively, and as the parameter u is fixed. So his two strategies are $(\varphi_1, \varphi_1, \sigma_2)$ and $(\varphi_1, \varphi_1, \sigma_2')$ where the φ_1 is repeated in each strategy because that is the action in both s_1 and $s_{1'}$. Since φ_1 is the only possible action, it can just be ignored, and we can simply say he has a choice between σ_2 and σ_2'. $\mathcal{B}$ has just one information set, the elongated oval, but she has two choices, σ_1 and σ_1', in that set and she has the same two versions of the game to choose from so her strategies are (σ_1, σ_2), (σ_1, σ_2'), (σ_1', σ_2), and (σ_1', σ_2'). A strategy profile is a combination of a strategy of $\mathcal{A}$'s and a strategy of $\mathcal{B}$'s. All of their strategies listed above are called *pure* strategies; *mixed* strategies involve a probability distribution on pure strategies. The description of a strategy here is a generalization of the usual definition because of the need to include the choice of version or of the values of the conditioning variables. This is another reason why such games are called games of *partial* information.

One way to solve such a game is to look for strategy profiles that no player will want to deviate from unilaterally. Such an equilibrium strategy, called a Nash equilibrium as we saw in Section 2.2,[1] will implicitly involve each agent taking account of what the other agent might do. As I said in Section 5.1, the key idea in decision-making is that an agent should maximize his expected utility modulo what the other agent does. So we have to compare the expected utilities of different strategies for each agent and choose the ones that neither will want to deviate from unilaterally. Because g_1 is part of the family of interdependent games $LG_u(\varphi) = \{g_1, g_2, g_{12}, g_1', g_2', g_{12}'\}$ and because each agent will want to choose the version of g_1 that the other chooses and that is compatible with the solutions to the other local games, there is an additional compatibility condition that is imposed on the solution that requires that both players play the same version

[1]Strictly speaking, the right solution concept is not quite a Nash equilibrium but a Perfect Bayesian equilibrium but, in the simple structure of a single partial information game, this nuance can be safely ignored. See Parikh (2010: Section 3.3.5).

of each game and that this version have conditioning variable values that agree with the choices made by $\mathcal{B}$ in the other games. For example, since there is a choice in g_1 between two values of the conditioning variable x_2, either $x_2 = \sigma_2$ or $x_2 = \sigma_2'$, the value chosen in equilibrium by both players must match with the interpretation chosen in g_2, where $\mathcal{B}$ has a choice between interpreting the utterance of φ_2 as σ_2 or σ_2' as shown in Figure 7.4. And such a requirement must be satisfied in all six interdependent games. This means that while it is correct to see each player as choosing the optimal version of each local game by specifying the values of the conditioning variables in the prior probability distribution, each of them will optimally want this choice to agree with the interpretive choices made by $\mathcal{B}$ in the other local games. So there are fewer degrees of freedom than might initially appear to be the case.

8.1.1 The two versions of g_1

To be completely clear about the choice of version in each agent's strategy, I show the versions explicitly in Figures 8.2 and 8.3. The prior probability functions ρ_1 and $\rho_{1'}$ have been replaced by the actual probabilities with different combinations of the conditioning variables.

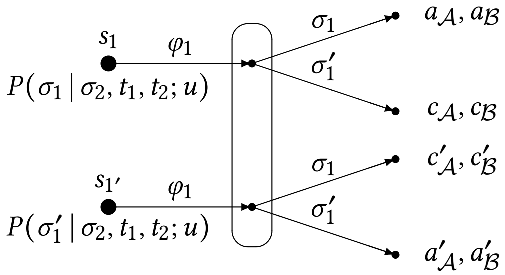


Figure 8.2: The first version of the semantic lexical game g_1

Consider Figure 8.2 which shows the first version of g_1. Assuming this first version is being played by both players as the compatibility condition requires, we investigate which strategy profile is best. This leaves just one real choice to be made: $\mathcal{B}$ has to decide between σ_1 and σ_1'. If $\mathcal{B}$ opts for σ_1 then $\mathcal{A}$'s expected utility will be $P(\sigma_1 \mid \sigma_2, t_1, t_2; u)a_{\mathcal{A}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u)c'_{\mathcal{A}}$ by tracing the paths followed by starting with s_1 through to the relevant terminal node resulting from the branch labeled σ_1 and the payoff $a_{\mathcal{A}}$ received there and by starting with $s_{1'}$ through to the relevant terminal node resulting from the branch labeled σ_1 and

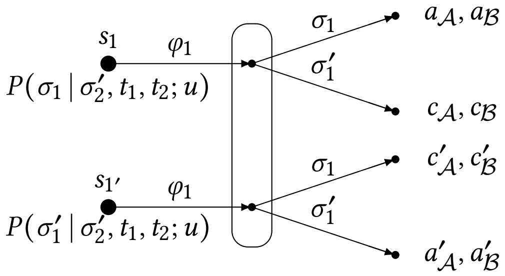


Figure 8.3: The second version of the semantic lexical game g_1

the payoff $c'_{\mathcal{A}}$ received there. If $\mathcal{B}$ opts for σ_1' then $\mathcal{A}$'s expected utility will be $P(\sigma_1 \mid \sigma_2, t_1, t_2; u)c_{\mathcal{A}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u)a'_{\mathcal{A}}$. $\mathcal{B}$ receives the same expected utilities with $\mathcal{A}$ replaced by $\mathcal{B}$ in the above expressions.

As I said on page 101 in Section 7.3, $P(\sigma_1 \mid \sigma_2, t_1, t_2; u) > P(\sigma_1' \mid \sigma_2, t_1, t_2; u)$ owing to the compatibility between the conditioning variables and parameter and the conditioned variable in the first probability as opposed to the second. If, in addition, it is assumed that the unprimed payoffs are identical to the corresponding primed payoffs for both players, that is, if $a_{\mathcal{A}} = a'_{\mathcal{A}}$, $c_{\mathcal{A}} = c'_{\mathcal{A}}$, $a_{\mathcal{B}} = a'_{\mathcal{B}}$, and $c_{\mathcal{B}} = c'_{\mathcal{B}}$[2] because these payoffs are endogenously generated by the agents trying to coordinate with $\mathcal{A}$'s intentions to convey σ_1 and σ_1' in s_1 and $s_{1'}$ and because there is no preference for any particular result in the ongoing conversation in u, it follows that $P(\sigma_1 \mid \sigma_2, t_1, t_2; u)a_{\mathcal{B}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u)c'_{\mathcal{B}} > P(\sigma_1 \mid \sigma_2, t_1, t_2; u)c_{\mathcal{B}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u)a'_{\mathcal{B}}$ and so $\mathcal{B}$'s σ_1 strategy is better than her σ_1' strategy.[3] In fact, if $\mathcal{B}$ chooses σ_1 in the first version of g_1 then *both* agents will receive a higher expected payoff than if she chooses σ_1'. So, with respect to the first version, (φ_1, σ_1) is a Nash equilibrium as neither player will want to deviate unilaterally from it, $\mathcal{A}$ because there is nothing to deviate to given his single choice of utterance φ_1 and $\mathcal{B}$ because σ_1' yields a lower expected utility to her. Therefore, all this part of the solution process involves is comparing two prior probabilities. Indeed, as I suggested earlier, their values could well be 5/6 and 1/6 and it would just be

[2]Actually, all that is required is that $a_{\mathcal{A}} - c_{\mathcal{A}} = a'_{\mathcal{A}} - c'_{\mathcal{A}}$ and $a_{\mathcal{B}} - c_{\mathcal{B}} = a'_{\mathcal{B}} - c'_{\mathcal{B}}$ as can be seen from transposing terms in the inequality $P(\sigma_1 \mid \sigma_2, t_1, t_2; u)a_{\mathcal{A}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u)c'_{\mathcal{A}} > P(\sigma_1 \mid \sigma_2, t_1, t_2; u)c_{\mathcal{A}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u)a'_{\mathcal{A}}$ and in the corresponding inequality for $\mathcal{B}$. This essentially makes the game a coordination game of the kind considered in Section 2.2.

[3]As was discussed on page 99, the a payoffs are always greater than the corresponding c payoffs, primed and unprimed, for both players because they represent the correct interpretation with respect to what $\mathcal{A}$ is conveying.

a matter of seeing which one is larger. This computation is quite undemanding, either for the human brain or for an artificial agent.

Exactly the same reasoning shows that in the second version of g_1 it is σ_1' that gives the higher expected utility to $\mathcal{B}$ (and to $\mathcal{A}$ as well) because $P(\sigma_1 \mid \sigma_2', t_1, t_2; u) < P(\sigma_1' \mid \sigma_2', t_1, t_2; u)$ as was established earlier. So (φ_1, σ_1') will be a Nash equilibrium in this version as neither player will want to deviate unilaterally from it. In fact, I had suggested that these probabilities could be 1/4 and 3/4, respectively.

So, depending on which version is chosen by $\mathcal{A}$ and $\mathcal{B}$, either σ_1 or σ_1' will be the equilibrium choice for $\mathcal{B}$ (and φ_1 is the only possible choice for $\mathcal{A}$). A more formal way of putting this is that the equilibrium of g_1 will be either $(\varphi_1, \sigma_1, \sigma_2)$ or $(\varphi_1, \sigma_1', \sigma_2')$ where the choice of the conditioning variable between σ_2 and σ_2' has been made explicit. Of course, the compatibility condition mentioned above means that this latter choice will be fixed in solving g_2. This makes the interdependence among the games crystal clear. I repeat that the trees t_1 and t_2 do not play any real role in this example as they are the only choices for the parses of φ_1 and φ_2. However, if we wish to make their role explicit as well, we can say that the choice of equilibrium in g_1 is between $(\varphi_1, \sigma_1, \sigma_2, t_1, t_2)$ and $(\varphi_1, \sigma_1', \sigma_2', t_1, t_2)$. Since only the real choices need to be mentioned, it is possible to drop φ_1, t_1 and t_2 from these Nash equilibria and write just (σ_1, σ_2) and (σ_1', σ_2'). Here, the first component of the equilibrium, σ_1 or σ_1', is selected in g_1 and the second component of the equilibrium, σ_2 or σ_2', is chosen in g_2 rather than g_1 owing to the compatibility condition. A choice of σ_2 in g_2 favors a choice of σ_1 in g_1 and a choice of σ_2' in g_2 favors a choice of σ_1' in g_1.

As the reader might anticipate, a similar scenario obtains with the two versions of g_2. In one version, σ_2 will be the best option for $\mathcal{B}$ based on the inequality $P(\sigma_2 \mid \sigma_1, t_1, t_2; u) > P(\sigma_2' \mid \sigma_1, t_1, t_2; u)$ and, in the other version, σ_2' will be the best option based on the inequality $P(\sigma_2 \mid \sigma_1', t_1, t_2; u) < P(\sigma_2' \mid \sigma_1', t_1, t_2; u)$. This time, by similar reasoning, we can conclude that (σ_2, σ_1) and (σ_2', σ_1') are the two Nash equilibria the players have to choose between. Here, the first component of the equilibrium, σ_2 or σ_2', is selected in g_2 and the second component of the equilibrium, σ_1 or σ_1', is chosen in g_1 rather than g_2 owing to the compatibility condition again. A choice of σ_1 in g_1 favors a choice of σ_2 in g_2 and a choice of σ_1' in g_1 favors a choice of σ_2' in g_2.

8.1.2 Looking for Pareto-Nash equilibria

These observations allow us to say that σ_1 and σ_2 are compatible with each other and σ_1' and σ_2' are compatible with each other. An alternative way of putting this is that the first versions of g_1 and g_2 are compatible with each other and the second versions of g_1 and g_2 are compatible with each other. In yet other words, we have two Nash equilibria in g_1 and we have two Nash equilibria in g_2 with the unprimed contents mutually implying each other and the primed contents mutually implying each other. How do we choose between these two sets of compatible pairs? One way to eliminate one pair of Nash equilibria is by bringing in the Pareto criterion and looking for what I have called Pareto-Nash equilibria in Section 2.2, something Spence (2001) also recommends in a different context.[4] The Pareto criterion states that a Nash equilibrium is Pareto-efficient if it is at least as good as the other Nash equilibrium for both players and is strictly better for at least one player. For (σ_1, σ_2) to be the unique Pareto-Nash equilibrium of g_1, we require that:

[4] Another possibility is to consider what are called risk-dominant Nash equilibria but they yield the same result under the assumptions we have made and are less intuitive. See Harsanyi & Selten (1988) and Parikh (2010: Section 3.3.5). More recently, so-called *iterated best response* reasoning has been considered by Franke (2009) and Jäger (2012). While such reasoning can occur in games where agents consciously reason about one another, it is highly implausible in the context of communication which takes only milliseconds. Besides, as Jäger communicated to me, it is NP-complete. As we will soon see, while it is necessary to go through the Pareto-Nash calculations that I am describing, their final result is dramatically simple to compute and also does not involve each agent consciously reasoning about the other explicitly. In particular, while Pareto-Nash equilibrium is also NP-complete, in our context of communication, its complexity class is just P.

Some theorists feel queasy about employing the Pareto criterion or even allowing agents to choose among Nash equilibria on the ground that agents can choose between strategies but cannot choose between strategy profiles, since the latter also include the other agent's choice. I have discussed this and related issues at some length in Parikh (2001: footnote 9, pages 40–41) and in Parikh (2010: Section 3.3.5) but here are some further observations. The central worry seems to be that some element of cooperation between the agents is required for the agents to choose Pareto-Nash equilibria in coordination games of the kind we are concerned with here. This may well be true. Earlier, I have suggested that such cooperation could result from one aspect of Grice's Cooperation Principle, which is being assumed in any case.

Agents can nevertheless sometimes choose different equilibria, and when they do, there is miscommunication. I discuss this kind of error in Section 8.4 in some detail. Often, it will lead to clarification requests, as Benz (2012) points out, and then the right equilibrium gets explicitly selected. It is unlikely, however, that both agents will get stuck jointly in a suboptimal equilibrium because the speaker knows what he intends to convey and his payoffs and probabilities will reflect this and make his choice optimal.

Given the alternatives – risk-dominance or epistemic reasoning of the iterated best response kind – assuming an element of cooperation in the choice of equilibria which, as we shall soon see, amounts to no more than choosing the maximal options, seems rather innocuous. Many game theorists such as Spence are not unduly troubled by this. I had mentioned in Parikh (2006a: 108–109) that the calculations in van Rooij (2004) showing the inadequacy of the Pareto criterion were flawed, as kindly conveyed to me by van Rooij himself.

either

$$P(\sigma_1 \mid \sigma_2, t_1, t_2; u) a_{\mathcal{A}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u) c_{\mathcal{A}}' \\ > P(\sigma_1 \mid \sigma_2', t_1, t_2; u) c_{\mathcal{A}} + P(\sigma_1' \mid \sigma_2', t_1, t_2; u) a_{\mathcal{A}}'$$

and

$$P(\sigma_1 \mid \sigma_2, t_1, t_2; u) a_{\mathcal{B}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u) c_{\mathcal{B}}' \\ \geq P(\sigma_1 \mid \sigma_2', t_1, t_2; u) c_{\mathcal{B}} + P(\sigma_1' \mid \sigma_2', t_1, t_2; u) a_{\mathcal{B}}'$$

or

$$P(\sigma_1 \mid \sigma_2, t_1, t_2; u) a_{\mathcal{A}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u) c_{\mathcal{A}}' \\ \geq P(\sigma_1 \mid \sigma_2', t_1, t_2; u) c_{\mathcal{A}} + P(\sigma_1' \mid \sigma_2', t_1, t_2; u) a_{\mathcal{A}}'$$

and

$$P(\sigma_1 \mid \sigma_2, t_1, t_2; u) a_{\mathcal{B}} + P(\sigma_1' \mid \sigma_2, t_1, t_2; u) c_{\mathcal{B}}' \\ > P(\sigma_1 \mid \sigma_2', t_1, t_2; u) c_{\mathcal{B}} + P(\sigma_1' \mid \sigma_2', t_1, t_2; u) a_{\mathcal{B}}'$$

Likewise, for (σ_2, σ_1) to be a Pareto-Nash equilibrium of g_2, we require that:
either

$$P(\sigma_2 \mid \sigma_1, t_1, t_2; u) a_{\mathcal{A}} + P(\sigma_2' \mid \sigma_1, t_1, t_2; u) c_{\mathcal{A}}' \\ > P(\sigma_2 \mid \sigma_1', t_1, t_2; u) c_{\mathcal{A}} + P(\sigma_2' \mid \sigma_1', t_1, t_2; u) a_{\mathcal{A}}'$$

and

$$P(\sigma_2 \mid \sigma_1, t_1, t_2; u) a_{\mathcal{B}} + P(\sigma_2' \mid \sigma_1, t_1, t_2; u) c_{\mathcal{B}}' \\ \geq P(\sigma_2 \mid \sigma_1', t_1, t_2; u) c_{\mathcal{B}} + P(\sigma_2' \mid \sigma_1', t_1, t_2; u) a_{\mathcal{B}}'$$

or

$$P(\sigma_2 \mid \sigma_1, t_1, t_2; u) a_{\mathcal{A}} + P(\sigma_2' \mid \sigma_1, t_1, t_2; u) c_{\mathcal{A}}' \\ \geq P(\sigma_2 \mid \sigma_1', t_1, t_2; u) c_{\mathcal{A}} + P(\sigma_2' \mid \sigma_1', t_1, t_2; u) a_{\mathcal{A}}'$$

and

$$P(\sigma_2 \mid \sigma_1, t_1, t_2; u) a_{\mathcal{B}} + P(\sigma_2' \mid \sigma_1, t_1, t_2; u) c_{\mathcal{B}}' \\ > P(\sigma_2 \mid \sigma_1', t_1, t_2; u) c_{\mathcal{B}} + P(\sigma_2' \mid \sigma_1', t_1, t_2; u) a_{\mathcal{B}}'$$

There are so many symbols in these inequalities that it is hard to decipher what they amount to. It is convenient to let $P(\sigma_1, \sigma_2 \mid t_1, t_2; u) = j$, $P(\sigma_1, \sigma_2' \mid t_1, t_2; u) = k$, $P(\sigma_1', \sigma_2 \mid t_1, t_2; u) = l$, and $P(\sigma_1', \sigma_2' \mid t_1, t_2; u) = m$ in order to see the probabilities more clearly. This implies:

$$\begin{aligned} P(\sigma_1 \mid \sigma_2, t_1, t_2; u) &= \frac{P(\sigma_1, \sigma_2 \mid t_1, t_2; u)}{P(\sigma_2 \mid t_1, t_2; u)} = \frac{P(\sigma_1, \sigma_2 \mid t_1, t_2; u)}{P(\sigma_1, \sigma_2 \mid t_1, t_2; u) + P(\sigma_1', \sigma_2 \mid t_1, t_2; u)} \\ &= \frac{j}{j+l} \end{aligned}$$

$$\begin{aligned} P(\sigma_1 \mid \sigma_2', t_1, t_2; u) &= \frac{P(\sigma_1, \sigma_2' \mid t_1, t_2; u)}{P(\sigma_2' \mid t_1, t_2; u)} = \frac{P(\sigma_1, \sigma_2' \mid t_1, t_2; u)}{P(\sigma_1, \sigma_2' \mid t_1, t_2; u) + P(\sigma_1', \sigma_2' \mid t_1, t_2; u)} \\ &= \frac{k}{k+m} \end{aligned}$$

$$\begin{aligned} P(\sigma_1' \mid \sigma_2, t_1, t_2; u) &= \frac{P(\sigma_1', \sigma_2 \mid t_1, t_2; u)}{P(\sigma_2 \mid t_1, t_2; u)} = \frac{P(\sigma_1', \sigma_2 \mid t_1, t_2; u)}{P(\sigma_1, \sigma_2 \mid t_1, t_2; u) + P(\sigma_1', \sigma_2 \mid t_1, t_2; u)} \\ &= \frac{l}{j+l} \end{aligned}$$

$$\begin{aligned} P(\sigma_1' \mid \sigma_2', t_1, t_2; u) &= \frac{P(\sigma_1', \sigma_2' \mid t_1, t_2; u)}{P(\sigma_2' \mid t_1, t_2; u)} = \frac{P(\sigma_1', \sigma_2' \mid t_1, t_2; u)}{P(\sigma_1, \sigma_2' \mid t_1, t_2; u) + P(\sigma_1', \sigma_2' \mid t_1, t_2; u)} \\ &= \frac{m}{k+m} \end{aligned}$$

$$\begin{aligned} P(\sigma_2 \mid \sigma_1, t_1, t_2; u) &= \frac{P(\sigma_1, \sigma_2 \mid t_1, t_2; u)}{P(\sigma_1 \mid t_1, t_2; u)} = \frac{P(\sigma_1, \sigma_2 \mid t_1, t_2; u)}{P(\sigma_1, \sigma_2 \mid t_1, t_2; u) + P(\sigma_1, \sigma_2' \mid t_1, t_2; u)} \\ &= \frac{j}{j+k} \end{aligned}$$

$$\begin{aligned} P(\sigma_2 \mid \sigma_1', t_1, t_2; u) &= \frac{P(\sigma_1', \sigma_2 \mid t_1, t_2; u)}{P(\sigma_1' \mid t_1, t_2; u)} = \frac{P(\sigma_1', \sigma_2 \mid t_1, t_2; u)}{P(\sigma_1', \sigma_2 \mid t_1, t_2; u) + P(\sigma_1', \sigma_2' \mid t_1, t_2; u)} \\ &= \frac{l}{l+m} \end{aligned}$$

$$P(\sigma_2' \mid \sigma_1, t_1, t_2; u) = \frac{P(\sigma_1, \sigma_2' \mid t_1, t_2; u)}{P(\sigma_1 \mid t_1, t_2; u)} = \frac{P(\sigma_1, \sigma_2' \mid t_1, t_2; u)}{P(\sigma_1, \sigma_2 \mid t_1, t_2; u) + P(\sigma_1, \sigma_2' \mid t_1, t_2; u)}$$
$$= \frac{k}{j+k}$$

$$P(\sigma_2' \mid \sigma_1', t_1, t_2; u) = \frac{P(\sigma_1', \sigma_2' \mid t_1, t_2; u)}{P(\sigma_1' \mid t_1, t_2; u)} = \frac{P(\sigma_1', \sigma_2' \mid t_1, t_2; u)}{P(\sigma_1', \sigma_2 \mid t_1, t_2; u) + P(\sigma_1', \sigma_2' \mid t_1, t_2; u)}$$
$$= \frac{m}{l+m}$$

Now we can substitute these simpler and more transparently interrelated fractions into the inequalities above. The first set of inequalities becomes:
either

$$\frac{j}{j+l} a_{\mathcal{A}} + \frac{l}{j+l} c_{\mathcal{A}}' > \frac{k}{k+m} c_{\mathcal{A}} + \frac{m}{k+m} a_{\mathcal{A}}'$$

and

$$\frac{j}{j+l} a_{\mathcal{B}} + \frac{l}{j+l} c_{\mathcal{B}}' \geq \frac{k}{k+m} c_{\mathcal{B}} + \frac{m}{k+m} a_{\mathcal{B}}'$$

or

$$\frac{j}{j+l} a_{\mathcal{A}} + \frac{l}{j+l} c_{\mathcal{A}}' \geq \frac{k}{k+m} c_{\mathcal{A}} + \frac{m}{k+m} a_{\mathcal{A}}'$$

and

$$\frac{j}{j+l} a_{\mathcal{B}} + \frac{l}{j+l} c_{\mathcal{B}}' > \frac{k}{k+m} c_{\mathcal{B}} + \frac{m}{k+m} a_{\mathcal{B}}'$$

The second set of inequalities becomes:
either

$$\frac{j}{j+k} a_{\mathcal{A}} + \frac{k}{j+k} c_{\mathcal{A}}' > \frac{l}{l+m} c_{\mathcal{A}} + \frac{m}{l+m} a_{\mathcal{A}}'$$

and

$$\frac{j}{j+k} a_{\mathcal{B}} + \frac{k}{j+k} c_{\mathcal{B}}' \geq \frac{l}{l+m} c_{\mathcal{B}} + \frac{m}{l+m} a_{\mathcal{B}}'$$

or

$$\frac{j}{j+k} a_{\mathcal{A}} + \frac{k}{j+k} c_{\mathcal{A}}' \geq \frac{l}{l+m} c_{\mathcal{A}} + \frac{m}{l+m} a_{\mathcal{A}}'$$

and

$$\frac{j}{j+k} a_{\mathcal{B}} + \frac{k}{j+k} c_{\mathcal{B}}' > \frac{l}{l+m} c_{\mathcal{B}} + \frac{m}{l+m} a_{\mathcal{B}}'$$

Recalling that $a_{\mathcal{A}} = a'_{\mathcal{A}} > c'_{\mathcal{A}} = c_{\mathcal{A}}$ and $a_{\mathcal{B}} = a'_{\mathcal{B}} > c'_{\mathcal{B}} = c_{\mathcal{B}}$, cross-multiplying and simplifying, and combining the two sets of inequalities, we get:

$$(jk > lm \text{ and } jl > km) \text{ for } \mathcal{A} \text{ and } (jk \geq lm \text{ and } jl \geq km) \text{ for } \mathcal{B}$$

or

$$(jk \geq lm \text{ and } jl \geq km) \text{ for } \mathcal{A} \text{ and } (jk > lm \text{ and } jl > km) \text{ for } \mathcal{B}$$

This further simplifies to:

$$jk > lm \quad \text{and} \quad jl > km \quad \text{for both agents} \tag{5}$$

Now Equation 5 can be simplified in two different ways. First, multiplying the two inequalities, we get:

$$j^2 kl > m^2 kl$$

which implies:

$$j > m$$

As $j = P(\sigma_1, \sigma_2 \mid t_1, t_2; u)$ and $m = P(\sigma'_1, \sigma'_2 \mid t_1, t_2; u)$, this tells us that for (σ_1, σ_2) to be a Pareto-Nash equilibrium of g_1 and for (σ_2, σ_1) to simultaneously be a Pareto-Nash equilibrium of g_2, we must have:

$$P(\sigma_1, \sigma_2 \mid t_1, t_2; u) > P(\sigma'_1, \sigma'_2 \mid t_1, t_2; u)$$

As already noted above, $P(\sigma_1 \mid \sigma_2, t_1, t_2; u) > P(\sigma'_1 \mid \sigma_2, t_1, t_2; u)$ and $P(\sigma_1 \mid \sigma'_2, t_1, t_2; u) < P(\sigma'_1 \mid \sigma'_2, t_1, t_2; u)$. This implies:

$$P(\sigma_1, \sigma_2 \mid t_1, t_2; u) > P(\sigma'_1, \sigma_2 \mid t_1, t_2; u)$$

and

$$P(\sigma_1, \sigma_2 \mid t_1, t_2; u) > P(\sigma_1, \sigma'_2 \mid t_1, t_2; u)$$

So $P(\sigma_1, \sigma_2 \mid t_1, t_2; u)$ is the maximum among

$$\left\{P(\sigma_1, \sigma_2 \mid t_1, t_2; u), P(\sigma'_1, \sigma_2 \mid t_1, t_2; u), P(\sigma_1, \sigma'_2 \mid t_1, t_2; u), P(\sigma'_1, \sigma'_2 \mid t_1, t_2; u)\right\}.$$

Looking back to Figure 7.6, this simply means that for (σ_1, σ_2) to be a Pareto-Nash equilibrium of g_1 and for (σ_2, σ_1) to be a Pareto-Nash equilibrium for g_2, all that is required is that the prior probability $\rho_{12} = P(\sigma_1, \sigma_2 \mid t_1, t_2; u)$ be the maximum among the four priors in the product game g_{12}. This result could also be obtained by solving g_{12} directly for its Nash equilibrium which can be seen to be $\sigma_1\sigma_2$ by similar computations. Importantly, in examples with syntactic ambiguity we would need to consider the full mixed semantic-syntactic sentential product mentioned on page 110 and compute its Nash equilibrium.[5]

Keep in mind that this is the result when these pairs form the unique Pareto-Nash equilibria of their respective games. There is no guarantee that this will always happen. It is quite possible, in a particular locutionary game, to have $P(\sigma_1, \sigma_2 \mid t_1, t_2; u) = P(\sigma_1', \sigma_2' \mid t_1, t_2; u)$. In such cases, the best that can be said is that *both* sets of solutions are plausible contents of the communication. This can certainly happen, and does, for example, with puns. But, in the communication at hand, it can safely be asserted that:

$$P(\sigma_1, \sigma_2 \mid t_1, t_2; u) > P(\sigma_1', \sigma_2' \mid t_1, t_2; u)$$

and that $P(\sigma_1, \sigma_2 \mid t_1, t_2; u)$ is the *strict* maximum of

$$\left\{P(\sigma_1, \sigma_2 \mid t_1, t_2; u), P(\sigma_1', \sigma_2 \mid t_1, t_2; u), P(\sigma_1, \sigma_2' \mid t_1, t_2; u), P(\sigma_1', \sigma_2' \mid t_1, t_2; u)\right\}$$

given the nature of u. This is because the implicit shared goal of u is to discuss the local election and so it gives a bigger boost to $P(\sigma_1, \sigma_2 \mid t_1, t_2; u)$ than to $P(\sigma_1', \sigma_2' \mid t_1, t_2; u)$. In other words, when it is warranted by u, the Pareto-Nash inequalities will naturally yield a strict or unique maximum, and when it is not warranted, they will not.

An alternative manipulation of Equation 5 implies:

$$jk + jm > jm + lm \quad \text{and} \quad jl + jm > jm + km$$

and so:

$$\frac{j}{j+l} > \frac{m}{k+m} \quad \text{and} \quad \frac{j}{j+k} > \frac{m}{l+m}$$

which is just:

$$P(\sigma_1 \mid \sigma_2, t_1, t_2; u) > P(\sigma_1' \mid \sigma_2', t_1, t_2; u) \text{ and } P(\sigma_2 \mid \sigma_1, t_1, t_2; u) > P(\sigma_2' \mid \sigma_1', t_1, t_2; u)$$

Therefore $P(\sigma_1 \mid \sigma_2, t_1, t_2; u)$ is the maximum among the four priors

$$\left\{P(\sigma_1 \mid \sigma_2, t_1, t_2; u), P(\sigma_1' \mid \sigma_2, t_1, t_2; u), P(\sigma_1 \mid \sigma_2', t_1, t_2; u), P(\sigma_1' \mid \sigma_2', t_1, t_2; u)\right\}$$

[5] See Chapter 9.

in the two versions of g_1 and $P(\sigma_2 \mid \sigma_1, t_1, t_2; u)$ is simultaneously the maximum among the four priors

$$\left\{P(\sigma_2 \mid \sigma_1, t_1, t_2; u), P(\sigma_2' \mid \sigma_1, t_1, t_2; u), P(\sigma_2 \mid \sigma_1', t_1, t_2; u), P(\sigma_2' \mid \sigma_1', t_1, t_2; u)\right\}$$

in the two versions of g_2.

8.1.3 A theorem

Thus, $LG_u(\varphi) = \{g_1, g_2, g_{12}, g_1', g_2', g_{12}'\}$ can be solved in two equivalent ways. We can confine our attention to just the lexical games and select the maximum priors across all versions in each game, ensuring that the priors are compatible, that is, that the conditioning variable in one is a conditioned variable in the other and vice versa. Or we can choose the maximum priors in the full mixed semantic-syntactic sentential product game. In the simple example at hand, the parse trees t_1 and t_2 did not play any real role so we were able to manage with just the semantic sentential product, but when they do, then the full mixed semantic-syntactic sentential product needs to be accessed. I record this important observation below.[6]

Theorem 8.1. The locutionary global game induced by an utterance can be solved either by solving just the lexical games in it or by solving just the full mixed sentential game or by solving appropriate intermediate phrasal games.

From both a psycholinguistic and computational viewpoint as well, all that has to be done is to work with either the conditional probabilities of the lexical games or the joint probabilities of the product games and select the corresponding maximum values. This gives us the solution to the locutionary global game in a completely rigorous and foundational way assuming nothing more than the rationality of agents.

The foregoing calculations show that there is a certain kind of overdetermination at work. First, the conditions for $\mathcal{A}$'s best actions yield the same results as the conditions for $\mathcal{B}$'s best actions so they are, in a sense, superfluous. And we know that $\mathcal{A}$ makes his choice of probability distributions or versions based on $\mathcal{B}$'s choices in the other games. However, this *is* a nontrivial decision and it is $\mathcal{A}$ who decides to let it depend on $\mathcal{B}$'s choices in the other local games. The results

[6] Important statements are displayed as theorems rather than facts or propositions even though their proofs are not particularly complex. I do not actually give proofs of any of the theorems but I hope the example considered makes them plausible.

therefore show that $\mathcal{A}$'s making his choices compatible with $\mathcal{B}$'s is in fact optimal from $\mathcal{A}$'s viewpoint. Incidentally, the same decision procedure in her choice of probability distribution or version works optimally for $\mathcal{B}$ as well.

8.1.4 The compact form

The information contained in the locutionary global game $LG_u(\varphi) = \{g_1, g_2, g_{12}, g_1', g_2', g_{12}'\}$ is scattered across six separate local games. But these games are highly interconnected and have a reciprocity through their prior probabilities. Is there a way to capture these dependencies in a more compact way? The answer is yes, as determined by the game theorist Ennio Stacchetti during a private discussion about $LG_u(\varphi)$. He came up with a single representation that compactly marshals all the information dispersed among the six local games, assuming that $\mathcal{A}$'s and $\mathcal{B}$'s choice of version or probability distribution is compatible with $\mathcal{B}$'s choices in the other games. I call the former representation the *distributed* form of $LG_u(\varphi)$ and Stacchetti's representation the *compact* form of $LG_u(\varphi)$. The existence of the compact form shows that it is right to think of $LG_u(\varphi)$ as a single game. I have tried to motivate this a little via the various calculations above. The compact form is shown in Figure 8.4.[7]

t_1, t_2	σ_2	σ_2'
σ_1	$P(\sigma_1 \mid \sigma_2, t_1, t_2; u), P(\sigma_2 \mid \sigma_1, t_1, t_2; u), 1, 1$	$P(\sigma_1 \mid \sigma_2', t_1, t_2; u), P(\sigma_2' \mid \sigma_1, t_1, t_2; u), 1, 1$
σ_1'	$P(\sigma_1' \mid \sigma_2, t_1, t_2; u), P(\sigma_2 \mid \sigma_1', t_1, t_2; u), 1, 1$	$P(\sigma_1' \mid \sigma_2', t_1, t_2; u), P(\sigma_2' \mid \sigma_1', t_1, t_2; u), 1, 1$

Figure 8.4: The compact form of $LG_u(\varphi)$

In this new, more abstract representation of $LG_u(\varphi)$, the players are no longer $\mathcal{A}$ and $\mathcal{B}$ but may be said to be the four lexical variables x_1, x_2, y_1, y_2 corresponding to the semantic and syntactic contents of the two words BILL and RAN in the sentence φ uttered in u. (Alternatively, the words could be said to be the players but each word would be taken as having two avatars, one semantic and one syntactic.) Readers new to game theory may feel that the players in a game have to

[7]This representation can also be referred to as a normal or strategic form. It is in fact both, a compact form and a normal form, the first because of its relation to the distributed form and the second because it is itself related to a so-called extensive form which is different from the distributed form. In other words, the distributed form can be translated into a single extensive form as well. Thus, we have the distributed form ~ the compact form ≡ the normal form ~ the extensive form.

be *agents* as ordinarily conceived but it is possible to allow games to have a more abstract kind of agency with more abstract kinds of players as we have here.

Recall that x_1 stood for either σ_1 or σ_1' and these are the two strategies available to the player x_1. Similarly, x_2 has the two strategies σ_2 and σ_2'. These two sets of strategies are shown on the left and on top of the payoff matrix. The other two players y_1 and y_2 have just single strategies t_1 and t_2 so they are shown in the upper left corner of Figure 8.4 in order to avoid having to present a four-dimensional matrix with the third dimension labeled t_1 and the fourth t_2. Instead of a four-dimensional matrix, multiple two-dimensional matrices can also represent the strategies of these players were they to have some real choice. In general, for a sentence with n words there would be a $2n$-dimensional matrix because there would be n semantic variables and n syntactic variables. The payoffs for the various combinations of strategies are shown in the cells of the matrix. For example, in the first cell, we have the vector $(P(\sigma_1 \mid \sigma_2, t_1, t_2; u), P(\sigma_2 \mid \sigma_1, t_1, t_2; u), 1, 1)$ which indicates that the payoff to x_1 is the first component, the payoff to x_2 is the second component, and the payoffs to y_1 and y_2 are just 1 and 1 because the relevant conditional probabilities $P(t_1 \mid \sigma_1, \sigma_2, t_2; u) = 1$ as shown in Figure 7.3 and $P(t_2 \mid \sigma_1, \sigma_2, t_1; u) = 1$ as shown in Figure 7.5.

The somewhat complicated mass of symbols in Figure 8.4 can be simplified if we drop the players y_1, y_2 as they are just silent bystanders and make the substitution of letters for conditional probabilities I had suggested above. The result is shown in Figure 8.5.

	σ_2	σ_2'
σ_1	$\frac{j}{j+l}, \frac{j}{j+k}$	$\frac{k}{k+m}, \frac{k}{j+k}$
σ_1'	$\frac{l}{j+l}, \frac{l}{l+m}$	$\frac{m}{k+m}, \frac{m}{l+m}$

Figure 8.5: The compact form of $LG_u(\varphi)$ with simpler symbols

This is now a familiar sort of game with a payoff matrix for two players. From the distribution of payoffs, it can be immediately seen that there must be at least one pure strategy Nash equilibrium. This is because a Nash equilibrium implies that neither player would want to deviate unilaterally from it. If no such Nash equilibrium existed, there would have to be a cycle with strategy profile (σ_1, σ_2) dominated by (σ_1', σ_2) for the first player (that is, the first player would deviate unilaterally from its strategy σ_1 to another strategy σ_1'), which in turn would be dominated by (σ_1', σ_2') for the second player, which would itself be dominated by (σ_1, σ_2') for the first player again, which would finally be dominated by (σ_1, σ_2) for the second player. In other words, there would be no "resting place" or Nash equilibrium. This would happen if the following inequalities among payoffs held, where the symbol "$\prec_i$" for $i = 1, 2$ stands for "dominated by for player i":

$$
\begin{array}{lll}
(\sigma_1, \sigma_2) \prec_1 (\sigma_1', \sigma_2) & \text{implies} & \frac{j}{j+l} < \frac{l}{j+l} \\
(\sigma_1', \sigma_2) \prec_2 (\sigma_1', \sigma_2') & \text{implies} & \frac{l}{l+m} < \frac{m}{l+m} \\
(\sigma_1', \sigma_2') \prec_1 (\sigma_1, \sigma_2') & \text{implies} & \frac{m}{k+m} < \frac{k}{k+m} \\
(\sigma_1, \sigma_2') \prec_2 (\sigma_1, \sigma_2) & \text{implies} & \frac{k}{j+k} < \frac{j}{j+k}
\end{array}
$$

all of which together imply $j < l < m < k < j$, a contradiction. This means there cannot be such a cycle. It should be clear by symmetry considerations that such an argument will apply to any compact form. I record this below.

Theorem 8.2. Every compact form of a locutionary global game has at least one Nash equilibrium in pure strategies.

For (σ_1, σ_2) to be a Nash equilibrium, we must have:

$$\frac{j}{j+l} \geq \frac{l}{j+l}$$

and

$$\frac{j}{j+k} \geq \frac{k}{j+k}$$

These inequalities imply $j \geq l$ and $j \geq k$. For (σ_1', σ_2') to be a Nash equilibrium, we must have:

$$\frac{m}{k+m} \geq \frac{k}{k+m}$$

and

$$\frac{m}{l+m} \geq \frac{l}{l+m}$$

These inequalities imply $m \geq k$ and $m \geq l$. If both (σ_1, σ_2) and (σ_1', σ_2') are Nash equilibria – as we saw earlier – then in order for the former to be the unique Pareto-Nash equilibrium, we must have:

either

$$\frac{j}{j+l} > \frac{m}{k+m}$$

and

$$\frac{j}{j+k} \geq \frac{m}{l+m}$$

or

$$\frac{j}{j+l} \geq \frac{m}{k+m}$$

and

$$\frac{j}{j+k} > \frac{m}{l+m}$$

This implies either $jk > lm$ and $jl \geq km$ or $jk \geq lm$ and $jl > km$, both of which imply $j > m$, exactly as we had when we considered the distributed form. But notice how much easier it was to obtain the result.

So far, I have considered just the four lexical variables x_1, x_2, y_1, y_2 derived from the words in the sentence as players but there are, in fact, potentially phrasal and sentential variables too. These are $x_{12} = x_1 \odot x_2$ and $y_{12} = y_1 \star y_2$. The four possible strategies of x_{12} are $\sigma_1\sigma_2$, $\sigma_1\sigma_2'$, $\sigma_1'\sigma_2$, and $\sigma_1'\sigma_2'$ and the single strategy of y_{12} is t_1t_2. But since these choices of the sentential variables are in fact determined by the choices of the four lexical players, these sentential players have no *real* choice beyond the lexical choices. And, indeed, this is reflected in the fact that the payoffs the semantic sentential variables receive from the four strategy choices – $P(\sigma_1, \sigma_2 \,|\, t_1, t_2; u) = j$, $P(\sigma_1, \sigma_2' \,|\, t_1, t_2; u) = k$, $P(\sigma_1', \sigma_2 \,|\, t_1, t_2; u) = l$, and $P(\sigma_1', \sigma_2' \,|\, t_1, t_2; u) = m$ – are also ordered by the requirements we have displayed above for various equilibria. For example, as we just saw, if (σ_1, σ_2) is a Pareto-Nash equilibrium of the compact form, then $P(\sigma_1, \sigma_2 \,|\, t_1, t_2; u) = j$ is automatically the maximum among

$$\Big\{P(\sigma_1, \sigma_2 \,|\, t_1, t_2; u), P(\sigma_1, \sigma_2' \,|\, t_1, t_2; u), P(\sigma_1', \sigma_2 \,|\, t_1, t_2; u), P(\sigma_1', \sigma_2' \,|\, t_1, t_2; u)\Big\} = \{j, k, l, m\}.$$

This leads to an insight that is parallel to Theorem 8.1 above.

Theorem 8.3. The compact form of a locutionary global game contains all the information required to determine its equilibrium. There is no need to consider nonlexical players.

Theorems 8.2 and 8.3 immediately imply the next observation.

Theorem 8.4. Every locutionary global game has at least one Nash equilibrium in pure strategies.

This result does not imply that mixed strategies may never appear as equilibria. When there is more than one Nash equilibrium in pure strategies, there may not be a unique Pareto-Nash equilibrium in pure strategies, in which case a unique mixed strategy equilibrium may be optimal.

8.1.5 The main theorems

In general, a sentence φ can be expressed as $\varphi_1 \circ \varphi_2 \circ \ldots \circ \varphi_n = \varphi_1 \varphi_2 \ldots \varphi_n$ where each φ_i is a word and there are n words in the sentence. If φ is uttered in some utterance situation u, then each word φ_i has a range of possible referential meanings given by the Semantic Constraint. The variable x_i stands for these possible meanings of φ_i and all such x_i together form the *meaning vector* $x = (x_1, x_2, \ldots, x_n)$. Likewise, the possible parse trees y_j of each word φ_j can be collected in the *parse vector* $y = (y_1, y_2, \ldots, y_n)$. Here, the range of each variable y_j is given by the Syntactic Constraint.

The meaning and parse vectors x and y together form the *content vector* $z = (x, y)$ with $2n$ components. This vector ranges over all the possible lexical meanings and parses of φ uttered in u. Note that $y_j = z_{n+j}$. Also, $z_{-k} = (z_1, z_2, \ldots, z_{k-1}, z_{k+1}, \ldots, z_{2n})$. That is, z_k is dropped from z in z_{-k}. Finally, $z^\star$ represents the equilibrium meaning and parse of all the lexical items.

We are now ready to state the central result of this section based on all the calculations above.

Theorem 8.5. Given an utterance and its locutionary global game (with symmetric payoffs), its lexical meanings and parses are given by the following equivalent characterizations:

(6) $$z^\star = \operatorname*{argmax}_z P(z_k \mid z_{-k}; u) = \operatorname*{argmax}_z P(z; u), \quad k = 1, \ldots, 2n$$

When there is more than one solution, each solution is given an equal probability.[8] The meaning and parse of the whole utterance can be obtained by computing the products of the lexical meanings and parses.

Equation 6 is called the Fundamental Equation of Equilibrium Linguistics. It provides a constraint on the beliefs of speakers and addressees in communication. The term $\operatorname{argmax}_z P(z_k \mid z_{-k}; u)$ after the first equality sign and the term

[8] By Theorem 8.4, a solution always exists.

$\text{argmax}_z\, P(z;u)$ after the second equality sign refer implicitly to the lexical games and to the full mixed semantic-syntactic sentential product game in the locutionary global game. Equation 6 improves upon the equation in Parikh (2010: Section 7.4) where the equivalence between the two ways of calculating $z^\star$ remained implicit. In principle, all a computing agent has to do is to run through the finite number of alternatives one by one until one or more vectors $z^\star$ is found to solve the system. It would be very easy to extend it to the Phonetic Constraint and to phonetic contents. At the start of this section, I mentioned the solution function f_u which allows us to write $f_u[LG_u(\varphi)] = \{f_u(g_1), f_u(g_2), f_u(g_{12}), f_u(g_1'), f_u(g_2'), f_u(g_{12}')\} = \{\sigma_1, \sigma_2, \sigma_1\sigma_2, t_1, t_2, t_1 t_2\}$. This function is an equivalent way of expressing $z^\star$.

The Fundamental Equation treats semantic, syntactic, (and phonetic) contents in a completely homogeneous manner. This is what justifies calling them all *contents* of the utterance. It also shows how the pipeline view of meaning – first phonetics, then syntax, then semantics, and last pragmatics – is completely transcended and replaced by a circular system of simultaneous equations. As I will show in Section 10.2, this is one way in which Equilibrium Linguistics generalizes Frege's principle of compositionality because Frege assumed that semantics reflects syntax but not the other way around. In Equilibrium Linguistics, all three contents – semantics, syntax, and phonetics – reflect one another. I do not show this right away as it is better to first complete the analysis fully and then discuss the philosophical consequences that emerge.

I had assumed earlier that $a_{\mathcal{A}} = a'_{\mathcal{A}}$, $c_{\mathcal{A}} = c'_{\mathcal{A}}$, $a_{\mathcal{B}} = a'_{\mathcal{B}}$, and $c_{\mathcal{B}} = c'_{\mathcal{B}}$, making the games coordination games. This is what allowed us to essentially ignore the payoffs and this is why we got the pure probabilistic result above. However, in general, these payoffs may not be symmetric because different outcomes corresponding to correct or incorrect interpretations may be valued differently by each agent. It is not difficult to bring the payoffs back in by simply scaling the conditional probabilities by appropriate factors in accordance with the inequalities for Pareto-Nash equilibria on page 133 as shown in Figure 8.6.

	σ_2	σ_2'
σ_1	$\frac{j}{j+l}a_{\mathcal{B}} + \frac{l}{j+l}c'_{\mathcal{B}}, \frac{j}{j+k}a_{\mathcal{B}} + \frac{k}{j+k}c'_{\mathcal{B}}$	$\frac{k}{k+m}a_{\mathcal{B}} + \frac{m}{k+m}c'_{\mathcal{B}}, \frac{k}{j+k}a'_{\mathcal{B}} + \frac{j}{j+k}c_{\mathcal{B}}$
σ_1'	$\frac{l}{j+l}a'_{\mathcal{B}} + \frac{j}{j+l}c_{\mathcal{B}}, \frac{l}{l+m}a_{\mathcal{B}} + \frac{m}{l+m}c'_{\mathcal{B}}$	$\frac{m}{k+m}a'_{\mathcal{B}} + \frac{k}{k+m}c_{\mathcal{B}}, \frac{m}{l+m}a'_{\mathcal{B}} + \frac{l}{l+m}c_{\mathcal{B}}$

Figure 8.6: The scaled compact form of $LG_u(\varphi)$

I will not analyze this scaled form of the locutionary game but just state the corresponding equation that results. Figure 8.6 reflects $\mathcal{B}$'s payoffs rather than $\mathcal{A}$'s though a corresponding matrix exists for $\mathcal{A}$ even though he has no choice of utterance. Let $V_k^{\mathcal{B}}(z, z_k')$ be $\mathcal{B}$'s payoffs in the lexical game g_k (or $g'_{(k-n)}$ if $k > n$ and the game is syntactic rather than semantic) and $V^{\mathcal{B}}(z, z')$ be her payoffs in the full mixed sentential game: these payoffs vary with different values of the vectors z and z', both of which represent full independent content vectors for φ. This notation compactly captures all the different individual payoffs such as $a_{\mathcal{B}}$, $a'_{\mathcal{B}}$, $c_{\mathcal{B}}$, and $c'_{\mathcal{B}}$ and need not be symmetric, even allowing for variations across the different lexical games.

Theorem 8.6. Given a locutionary global game, the lexical meanings and parses of an utterance are given by the following equivalent characterizations:

$$
\begin{aligned}
z^\star &= \underset{z_{-k}, z_k'}{\operatorname{argmax}} \sum_{z_k} P(z_k \mid z_{-k}; u) V_k^{\mathcal{B}}(z, z_k') \\
&= \underset{z'}{\operatorname{argmax}} \sum_{z} P(z; u) V^{\mathcal{B}}(z, z'), \quad k = 1, \ldots, 2n
\end{aligned}
\tag{7}
$$

When there is more than one solution, each solution is given an equal probability. The meaning and parse of the whole utterance can be obtained by computing the products of the lexical meanings and parses.

Equation 7 may be called the Fundamental Equation of Equilibrium Linguistics with Payoffs. It expresses a more general result of our analysis and can also be put in matrix form. Generally Equation 6 suffices although in certain contexts as pointed out in Section 12.2 we have to resort to Equation 7.

This completes my discussion of how locutionary global games are solved. The description has taken a few pages but the results are strikingly simple and in accord with our intuitions. All that needs to be done is to find the interpretations with the highest probabilities in a *compatible* manner where by "compatible" I mean the property that $\operatorname{argmax}_z P(z_k \mid z_{-k}; u) = \operatorname{argmax}_z P(z; u)$ as described by Equation 6. One aspect of this – identifying the *most likely* contents – is certainly common in both human and artificial contexts. But the second aspect – maximizing these probabilities in a *compatible* manner – is less evident. That communication involves semantic, syntactic, and phonetic contents mutually determining one another via such an intricate probabilistic structure is remarkable.

Just to be clear that not every probability distribution is compatible, Figure 8.7 shows a distribution $P(z_1, z_2)$ with two variables z_1, z_2 and three values in the range of each variable that is not.

z_1, z_2	σ_2	σ_2'	σ_2''
σ_1	0.20	0.15	0.10
σ_1'	0.15	0.15	0.00
σ_1''	0.10	0.00	0.15

Figure 8.7: An incompatible distribution

For example, z_1 could range over the contents σ_1, σ_1', σ_1'' corresponding to the three rows and z_2 could range over the contents σ_2, σ_2', σ_2'' corresponding to the three columns. For this distribution,

$$\operatorname*{argmax}_z P(z) = (\sigma_1, \sigma_2) \text{ but } \operatorname*{argmax}_z P(z_k \mid z_{-k}) = (\sigma_1'', \sigma_2''),$$

as can be checked by inspection. That is, the maximum of the joint distribution occurs at the first row and first column and the maximum of the conditional distributions occurs at the third row and third column. Such a distribution is ruled out by Theorem 8.5.

Why do compatible probability distributions arise in communication? Recall from page 100 in Section 7.3 that each prior probability is *not* the conditional probability of some content but is rather the conditional probability of a speaker's *conveying* that content. That is, these prior distributions govern the *intentions* and *actions* of rational agents, of what they would and would not choose to convey or interpret if they were conveying or interpreting certain other things as well. So compatibility describes the two-sided *logic* of rational intention and action. It is co-extensive with rationality itself in the context of communication. Earlier, I had mentioned that in many settings agents are not perfectly rational but because the structures involved in locutionary global games are relatively simple and because the payoffs are also assumed to be symmetric, there is little scope for limited agents to falter. Even when scaled compact forms as shown in Figure 8.6 are used to derive the more general Equation 7, there is no reason for rationality to fail.

Equally important, these probability distributions have nothing to do with language per se. The same property would hold in any communication involving images, gestures, and other symbol systems. It may even apply beyond communication to appropriately linked subsystems of a larger system, as I conjecture in Parikh (2010: Section 7.1.4).

Theorem 8.5 provides a theoretical foundation for and generalization of various observations in both psycholinguistics and statistical natural language processing, as I discuss in Chapter 12. On the other hand, to the best of my knowledge, the philosophy of language and theoretical linguistics and artificial intelligence do not appear to have entertained anything close to such ideas about how locutionary meanings and other contents might be derived from first principles. They usually just say such meanings are "conventional" and turn their attention to deriving implicatures using some version of Grice's conversational maxims. The problem of rigorously deriving content is a *philosophical* and *theoretical* problem, not just some detail to be relegated to empirical investigation.

Where do all these probabilities come from? For human interpretation, they would come partly from the grammar and partly from the utterance situation and wider discourse situation and would be based on both objective and subjective factors.[9] People are generally quite good at rough probability estimates (as in figuring whether it is likely to rain by looking at the sky) and all they need to determine here is whether certain probabilities are higher than others. No precise numerical estimates are required. For an artificial agent, these probabilities would come from the grammar and from suitable corpora.

To a certain degree, all these probabilities rely on background knowledge of one kind or another, some of which may be present in the encyclopedic knowledge associated with the conventional meaning of a word and some of which may be present in one or more knowledge or belief bases in the agent's head. Some linguists may feel that the explanation has just been pushed back to some unknown probabilities but the problem of determining such probabilities is common to all action, not just communicative action, as should be clear from the *very* wide applicability and, indeed, applications of decision and game theory. Rather than say, as Chomsky has chosen to, that this makes the entire realm of human choice and behavior a mystery, one has to approach it in the usual way of science, which is to first divide the problem into two parts and then to tackle each separately. In the case of communication, I have shown how the problem of deriving locutionary meaning can be reduced to some probabilistic computations. Now, as computer scientists frequently do in limited domains, distributed and interconnected knowledge bases can be set up and it would become very clear how

[9] See Mante et al. (2013) for some recent work on how contextual information is processed by the brain involving a recurrent neural network model and experiments with monkeys. See also http://engineering.stanford.edu/news/stanford-researchers-surprised-find-how-neural-circuits-zero-specific-information-needed-decisi. These findings, which are still in their infancy, are relevant for all the games in the book as they all make use of contextually derived probabilities.

such probabilities can emerge from such data. If an agent knows that black clouds make rain likely and he sees black clouds, it follows that the probability of rain is high. That is all. Indeed, it is not even necessary to build an actual knowledge or belief base in the case of human communication; it suffices merely to assume that it exists and that neuroscience will, in time, enable us to understand how it figures in decision-making. Building satisfactory conversational agents based on Equilibrium Semantics does require constructing and *scaling* the knowledge base sufficiently but I think it is possible to say that the problem of determining locutionary meaning has been solved. I touch upon this point briefly again in Section 12.2.

An alternative response to this issue of explaining how such parameters arise is to say that the complaint misunderstands the scientific problem. In physics, for example, the solution to the problem of projectile motion is divided into two parts: the relevant equation of motion and the initial or boundary conditions (e.g. the initial velocity). The scientific problem is determining the equation of motion. In different situations, the initial or boundary conditions will be determinable with more or less precision but no one thinks it is reasonable to say that the problem has just been pushed back to some unknown parameters.

I reproduce a paragraph from my dissertation, *Language and Strategic Inference* (1987b: 5):

> We will argue for this thesis by developing a detailed account of one strategic inference in isolation. Any complete utterance, that is, any utterance that attempts to express a proposition, will involve many separate acts and strategic inferences. For example, part of a communication will typically involve a referential act, an act of referring to some object, and the communication of this reference to the addressee. Each bit of information communicated will require its own strategic inference(s). Thus, any complete utterance involves an entire system of simultaneous strategic inferences. These inferences have to be simultaneous in general because they codetermine each other in general. For example, an utterance of "Bill has the book" will require inferring the designata of each of the four words in the sentence, (not to mention its internal structure), in order to determine the proposition expressed. No word has any particular priority in this determination. That is, there may be interactions among the various strategic inferences. And the embedding circumstances play a vital role in each inference. Mathematically, this amounts to a system of simultaneous equations.

Earlier, I didn't know how to work out this idea. Now, several years later, I do.

As the Interpretation Game $UG_u(\varphi) = G_u^{\mathcal{B}}(\varphi) = LG_u^{\mathcal{B}}(\varphi) \cup IG_u^{\mathcal{B}}(\varphi) = LG_u(\varphi) \cup IG_u(\varphi) = G_u(\varphi)$ owing to our assumption of equality of the subjective and objective games and common knowledge of them, we have solved the locutionary part of the Interpretation Game the addressee faces. Its illocutionary part will be addressed in Part IV. The Generation Game is just a little more complicated as multiple sentences may need to be evaluated and their costs taken into account.

8.2 My former partial information games

Readers familiar with my earlier books will have noticed that a game like the one in Figure 8.1 would ordinarily have been shown as in Figure 8.8.

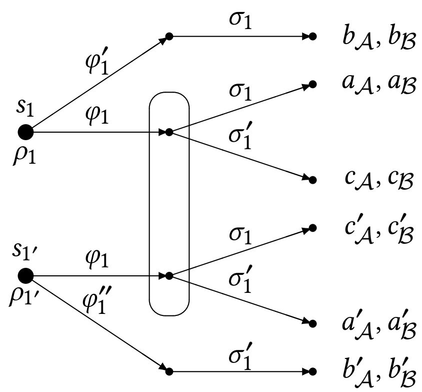


Figure 8.8: Semantic lexical game g_1

In this game, there are two more branches with alternative utterances φ_1' at s_1 and φ_1'' at $s_{1'}$. These are both *unambiguous* with respect to the ambiguity in φ_1 = BILL. So φ_1' = BILL SMITH and φ_1'' = BILL JONES represent one set of possibilities for these alternatives. As they are unambiguous, there is only one interpretation of each possible, either σ_1 or σ_1'. As they are costlier but involve correct interpretations, the resulting payoffs $b_{\mathcal{A}}, b_{\mathcal{B}}$ and $b_{\mathcal{A}}', b_{\mathcal{B}}'$ are somewhat lower than the corresponding a payoffs and higher than the corresponding c payoffs.

The games I have drawn in this and the previous chapter do not have such alternative utterances for two reasons. When deriving locutionary meaning, if we allow an alternative like φ_1' = BILL SMITH, then this alternative itself needs to be interpreted word by word, leading to a possible regress and further disambiguations, which is unrealistic. In Parikh (2010: 88 and 120–121), I suggested dealing with such alternatives *holophrastically* but this is unnecessary. Alternative *contents* are anyway considered by both agents in the Content Selection Game and the speaker *does* occasionally consider more than one alternative sentence (or subsentential expression) in the Generation Game so further alternatives *within* the partial information game are not required. Secondly, such costs are best considered *outside* the locutionary global game $LG_u(\varphi)$ but within the Generation Game GG_u. This makes it possible to include more sources of cost.

This dropping of alternative utterances may seem superficially like a small change but it results in *very* different solution processes for these new games of partial information, as readers familiar with the earlier solution processes will no doubt have realized. This is what makes possible the two parts of the Fundamental Equation 6 and all the theorems in Section 8.1. Earlier, such finer calculations of equilibria were not possible.

Another consequence of this truncated form of the game without the alternatives is that such alternative utterances no longer need to be part of the definition of speaker meaning. The subjective global game $G_u^{\mathcal{A}}(\varphi)$ in Definition 5.1 in Section 5.3 contains just the sentence φ.

Others using game theory in this field (e.g. Benz et al. 2006), largely following the form of my earlier games of partial information above, have also used this idea of alternative utterances, apparently without realizing that these alternatives would themselves need to be evaluated using the same methods that are being applied to the main sentence, which could lead to an undesirable regress. To avoid this, they would once again have to follow my suggestion of locating alternative contents in the Content Selection Game and alternative sentences in the Generation Game outside the local games of partial information.

8.3 An interesting complication

While the results of Section 8.1 are quite powerful, they depend on a condition that has been left implicit.[10] Consider an utterance of the following sentence:

(8) The boy saw the girl with a telescope. (μ)

As is well known, the prepositional phrase WITH A TELESCOPE can attach either to the verb SAW (i.e. to the verb phrase SAW THE GIRL) or to the noun GIRL and so μ is syntactically ambiguous. But there is no syntactic ambiguity at the level of lexical categories if they are expressed as in an ordinary CFG. The product of

[10] I did this deliberately to simplify the discussion as will become clear in this section.

the optimal lexical parses obtained via Theorem 8.5 would then always yield *two* sentential parses even though the situation u in which μ is uttered may suffice to disambiguate between them.

This incomplete disambiguation can be completed by adding sufficient detail to the lexical categories so that the preposition WITH has two distinct representations rather than just one. These representations may be described as *grammatical roles*: when the preposition attaches to the verb SAW (or the verb phrase SAW THE GIRL), the role is that of an *instrument*, and when the preposition attaches to the noun GIRL, the role is that of *accompaniment*. Such roles and other refinements of lexical categories arise naturally in feature-based grammars of the kind described in Shieber (1986) and Steedman (2001).

Here are some examples of feature structures from Shieber (1986: Chapter 3):

$$\begin{bmatrix}\text{cat:} & \text{NP}\end{bmatrix}$$

$$\begin{bmatrix}\text{cat:} & \text{NP} \\ \text{agreement:} & \begin{bmatrix}\text{number:} & \text{singular} \\ \text{person:} & \text{third}\end{bmatrix}\end{bmatrix}$$

Here the first feature structure is just our plain vanilla category of a noun phrase. The second one shows the added feature of *agreement*. The value of this second feature is itself a feature structure with features *number* and *person* and corresponding values *singular* and *third*. This means the noun phrase is singular and third person and would have to combine with an appropriate category and feature structure to form a larger expression. The first feature structure above carries less information than the second about the noun phrase in question and is said to subsume the latter, leading to a natural lattice structure. The second, more informative feature structure also shows how feature structures can possess a nesting relationship.

In our example of the two roles for the preposition WITH, we could define the following two different feature structures for such words:

$$\begin{bmatrix}\text{role:} & \text{instrument}\end{bmatrix}$$

$$\begin{bmatrix}\text{role:} & \text{accompaniment}\end{bmatrix}$$

These different structures would differentiate syntactically between the two uses of the preposition at the *lexical* level as required.

Feature structures can be and often are as finely articulated at the lexical level as required. This feature-based way of describing a context-free grammar then allows all syntactic ambiguities to be represented at the lexical level and so the "lexical computations" of Theorem 8.5 – $z^\star = \mathrm{argmax}_z\, P(z_k \mid z_{-k}; u)$ – suffice to give us exactly one solution when warranted.

Even when the sentence is indefinitely long, as in $\mathcal{A}_1$ SAW THAT $\mathcal{A}_2$ SAW THAT ... THAT $\mathcal{A}_n$ SAW THE GIRL WITH A TELESCOPE for any $n \in \mathbb{N}$, the features can be recursively defined to allow an indefinite degree of differentiation at the lexical level for different attachments of the preposition WITH so that the lexical computations of Theorem 8.5 can still go through as required. In other words, all that is needed for the earlier results to hold is a sufficiently detailed feature-based representation of the grammar G. This is the condition that was left implicit in Section 8.1.

What happens when G is not feature-based (or, in other words, the feature structures are minimal) and there is a lack of differentiation at the lexical level for sentences like μ and similar sentences? In that case, the *compatibility* of the probability distribution as specified above will have to be generalized so that it applies to nonlexical levels as well.

Assume there is an utterance of a sentence φ as before. $K = \{1, 2, \ldots, n, n + 1, \ldots, 2n\}$ is the set of semantic and syntactic indices of z_k which stand for the *possible* lexical meanings and parses of φ uttered in u. Let the index set $K_\varphi \subset K$ be a *variable* set that ranges over the indices of the possible meanings and parses of φ at a lexical, phrasal, and sentential level. For example, in a simple sentence such as "John loves Mary," $K_\varphi \in \{\{1\}, \{2\}, \{3\}, \{2, 3\}, \{1, 2, 3\}, \{4\}, \{5\}, \{6\}, \{5, 6\}, \{4, 5, 6\}\}$ where the indices {1}, {2}, {3} correspond to the three semantic indices of the three words in the sentence, the set $\{2, 3\}$ corresponds to the semantic indices of the phrase LOVES MARY, the set $\{1, 2, 3\}$ corresponds to the semantic indices of the whole sentence, and the indices {4}, {5}, {6} correspond to the three syntactic indices of the three words in the sentence, the set $\{5, 6\}$ corresponds to the syntactic indices of the phrase LOVES MARY, the set $\{4, 5, 6\}$ corresponds to the syntactic indices of the whole sentence. Clearly, the range of variation of K_φ will be determined by each possible parse of φ and is thus determined by the grammar G. If we write $K_G^l(\varphi)$ for this range for the lth parse out of L possible parses (e.g. $K_G^1(\varphi) = \{\{1\}, \{2\}, \{3\}, \{2, 3\}, \{1, 2, 3\}, \{4\}, \{5\}, \{6\}, \{5, 6\}, \{4, 5, 6\}\}$ for the simple example where $L = 1$ possible parse), we can then say $K_\varphi \in K_G^l(\varphi)$ for the lth of L possible parses. That is, the index set K_φ is always a member of a particular range of variation $K_G^l(\varphi)$ based on which parse is being considered.

Define $z_{K_\varphi} = \{z_k \mid k \in K_\varphi\}$ and define $z_{K'_\varphi}$ to be its complement with respect to K. Again, for the simple example above, if $K_\varphi = \{2, 3\}$ then $z_{K_\varphi} = \{z_2, z_3\}$ and

$z_{K'_\varphi} = \{z_1, z_4, z_5, z_6\}$. In this case, $P(z_{K_\varphi} \mid z_{K'_\varphi}; u) = P(z_2, z_3 \mid z_1, z_4, z_5, z_6; u)$. Likewise, if $K_\varphi = \{4, 5, 6\}$ then $z_{K_\varphi} = \{z_4, z_5, z_6\}$, $z_{K'_\varphi} = \{z_1, z_2, z_3\}$, and $P(z_{K_\varphi} \mid z_{K'_\varphi}; u) = P(z_4, z_5, z_6 \mid z_1, z_2, z_3; u)$.

While the two paragraphs above are notationally cumbersome, the basic idea is very simple. Earlier, we found just the lexical level adequate when the grammar was feature-based. When it is not, we need to access the indices at a phrasal and sentential level as well so that the computations of Theorem 8.5 can be extended to nonlexical levels in a way that provides a generalized notion of compatibility. What the more general statement of Theorem 8.7 requires is to run through the relevant indices at phrasal and sentential levels as well and the indices that are relevant at such levels are the ones that are dictated by the possible parses being considered.

Theorem 8.7. Given an utterance and its locutionary global game (with symmetric payoffs), when the relevant grammar is not feature-based, its lexical meanings and parses are given by the following equivalent characterizations:

(9) $$z^\star = \underset{z}{\operatorname{argmax}}\, P(z_{K_\varphi} \mid z_{K'_\varphi}; u) = \underset{z}{\operatorname{argmax}}\, P(z; u), \quad K_\varphi \in K^l_G(\varphi),\ l \in L$$

When there is more than one solution, each solution is given an equal probability.[11] The meaning and parse of the whole utterance can be obtained by appropriately computing the products of the lexical, phrasal, or sentential meanings and parses.

I urge the reader to mentally run through the more complex computations involved with this extended idea of compatibility. Notice that the term

$$\underset{z}{\operatorname{argmax}}\, P(z_{K_\varphi} \mid z_{K'_\varphi}; u)$$

in Equation 9 does not involve any mixed semantic-syntactic products whereas the term $\operatorname{argmax}_z P(z; u)$ does so that there is still a reduction in computational effort by the equality of the two. If we call indefinitely articulated feature-based grammars complex and ordinary context-free grammars simple and if we call the lexical computations of Theorem 8.5 simple and the generalized computations of Theorem 8.7 complex, then there is a nice and intuitive inverse relationship between grammars and computations that follows from these two theorems. Let the grammar of a language be G and the corresponding computations of optimal meanings and parses be κ.

[11] Again, by Theorem 8.4, a solution always exists.

Theorem 8.8. G and κ are inversely related in their complexity: when G is simple, κ is complex, and when G is complex, κ is simple.

For the rest of the book, I will assume we have a sufficiently articulated feature-based grammar G and that the computations κ are therefore simple. In Chapter 9, I will attend to an example of syntactic ambiguity in detail, something I have not addressed so far but there all the ambiguities will be present at the lexical level itself so that issues of the kind raised by sentences like μ will not arise. For the present, we have to complete our discussion of our primary example φ = BILL RAN and I return to this below.

8.4 Solving Generation Games

As we saw in Section 7.3 and Figure 7.8, the speaker may consider more than one sentence and therefore more than one locutionary global game. Each such game has a value $v_u^{\mathcal{A}}[LG_u(\varphi)] = a_{\mathcal{A}} + a_{\mathcal{A}}$. Likewise, each corresponding illocutionary game will have some value and it is the total value $v_u^{\mathcal{A}}[G_u(\varphi)]$ that is computed. Each global game $G_u(\varphi)$ comes with a cost $k_u^{\mathcal{A}}(\varphi)$ and so the net value to $\mathcal{A}$, $v_u^{\mathcal{A}}(\varphi) = v_u^{\mathcal{A}}[G_u(\varphi)] - k_u^{\mathcal{A}}(\varphi)$, is determined. Then, the sentence with the highest net value is chosen.

The cost $k_u^{\mathcal{A}}(\varphi)$ is a crucial factor that controls what is actually uttered.[12] Incidentally, considerations of cost are entirely absent in Lewis (1969) and also in so-called cheap talk games in economics (e.g. Crawford & Sobel 1982; Farrell 1987; 1993; Farrell & Rabin 1996; Kartik 2009; Blume & Board 2013) and this is just one reason why they are unsuitable for modeling linguistic communication.[13] Cost

[12] I had first pointed this out in Parikh (1987b).

[13] The setup in Crawford and Sobel is that there is a sender and a receiver. The sender observes a state which is equally likely to be either 0 or 1 (although the original account involves a continuous state space). He then reports the state to the receiver either truthfully or not. The receiver observes the report and makes a decision which can be any number. The receiver prefers decisions that are closest to the value of the state and the sender prefers decisions that are closest to the state plus a bias $b > 0$.

There are two ways to view such a setup. The shared language between sender and receiver either does not constrain interpretation at all or contains expressions that have only single conventional meanings but then allows any interpretation.

Consider the first option. Suppose the sender observes a 0 and sends a 0. The receiver can interpret this expression "0" as either 0 or 1. This means there is *no* shared conventional meaning that the expression has that constrains its possible interpretations the way the word "bank" limits its possible locutionary interpretations to either *financial institution* or *land alongside a river*. In other words, the expression "0" is implicitly taken to be conventionally meaningless. That is, the whole language (in this case, just "0" and "1") has no conventional meanings and is based on a variety of factors, mostly objective ones such as the length and complexity of the sentence and the effort required to mentally process it and to physically produce it but also subjective ones especially related to the consequences of making certain information explicit and leaving certain information implicit.

It is possible that the primary decision is between what to make explicit or leave implicit and objective factors such as length, complexity, and mental and physical effort play a relatively smaller role. Conversational style, habits of speech, special contexts of speech and writing (e.g. poetry and fiction and nonfiction), and other aesthetic aspects are one factor that affect this primary decision and different communities and individuals may follow different patterns in how much is made explicit. This is a matter for sociolinguists and psycholinguists to dig deeper into.

A second dimension involves the maintaining or altering of relationships as pointed out by Pinker et al. (2008) based on the Politeness Theory of Brown & Levinson (1987) and referred to in Section 5.2.[14] This factor is extremely important and plays a role not only in how some content is communicated but even in

its expressions can convey anything at all. As a result, whenever the sender has to convey a content, he can use any expression in the whole language.

Consider the second option for viewing the setup. Here, the expression "0" always has only the single conventional meaning 0 but this conventional meaning can be a springboard for any further interpretation, 0 or 1. Therefore, any expression can again be used to convey any content *via* a single conventional meaning and the whole language is again always available.

Whichever way the setup is viewed, such models do not investigate how a particular meaning arises from potentially ambiguous expressions that have a fixed set of conventional meanings. As Myerson (1989: 265) rightly observes, such models treat (literal) meaning as endogenous to the setup rather than exogenous. The Semantic Constraint (i.e. the Conventional and Referential Constraints) mentioned in Section 7.3 which allows an expression (e.g. "bank") to be conventionally ambiguous is therefore missed altogether. It is this constraint that *restricts* the possible interpretations that are then disambiguated by the Flow Constraint. This is why the whole language is never available to the speaker when he wishes to convey a particular content.

The real question in linguistic communication is not about the conditions under which "bank" will be treated as having a single meaning rather than having no shared meaning (the first option) or as conveying some meaning other than its single conventional meaning (the second option) but is about the conditions under which one of its shared conventional meanings (e.g. either *financial institution* or *land alongside a river*) will prevail. See also Section 3.2 on the utility of conventional meanings. I believe more recent work such as Dewatripont & Tirole (2005) that considers costly communication has the same limitations.

[14] Pinker et al. study this problem from a similar angle but use a different, somewhat ad hoc setup in their explanations. They seem to conclude that indirect speech is used only when there is a risk of some kind of penalty (as in overtly or covertly bribing a police officer) or some kind of awkwardness in the relationship (as in directly or indirectly bribing the maître d' of a restaurant) or an alteration of a relationship (as in overtly or covertly making a sexual overture).

the choice of what content to communicate. In other words, it affects both the Content Selection Game and the Generation Game and I will be examining how it enters into content selection in Sections 8.6 and 18.2 and Chapter 11. In Generation Games, if a sentence is too explicit or too implicit, it will incur a cost and this may lead to its rejection owing to more suitable sentences. The utterance situation u obviously plays a pivotal role in this evaluation.

For example, suppose Bill Smith is not just a mutual acquaintance of the two interlocutors but is also someone $\mathcal{B}$ holds in high regard. Then in order to please $\mathcal{B}$, $\mathcal{A}$ may choose to be more deferential and utter φ' = BILL SMITH RAN rather than φ = BILL RAN. This would be so even though φ' is slightly longer and possibly stylistically too formal. This is because the higher cost due to greater length and formality is offset by the reduction in cost due to $\mathcal{A}$'s managing to please $\mathcal{B}$ by his show of deference. Overall, $k_u^{\mathcal{A}}(\varphi) > k_u^{\mathcal{A}}(\varphi')$ and therefore $v_u^{\mathcal{A}}(\varphi) < v_u^{\mathcal{A}}(\varphi')$ because $v_u^{\mathcal{A}}[G_u(\varphi)] = v_u^{\mathcal{A}}[G_u(\varphi')]$ in this instance, and so φ' would be uttered, not φ. This shows how factors like maintaining relationships and politeness toward one's interlocutors may alter the choice of sentence to convey even the same content σ. Indeed, a great deal about the use and even existence of honorifics and other such behavior can be explained in just this way. The example just discussed involves greater explicitness but greater implicitness may also be warranted in other situations where something like, say, a sexual overture has to be handled delicately or indirectly.

There is at least a third important dimension to what is made explicit. This is the avoidance of error in contexts where error can be costly. If a sentence leaves too much implicit, the addressee may infer the wrong content and this may lead to an undesired action and outcome. In the example at hand, it may not matter too much to $\mathcal{A}$ if $\mathcal{B}$ interpreted φ = BILL RAN as being about Bill Jones instead of Bill Smith or about running in a race instead of in the local election or both because they are involved in a casual conversation and misunderstandings could be easily

They overlook the fact that indirect speech is also used for the most mundane reasons: to reduce effort. And, contrary to their point about the role of common knowledge in influencing indirect speech (i.e. a direct statement becomes common knowledge whereas indirect speech doesn't), a great deal of indirect speech does in fact become common knowledge: "can you pass the salt?" or "would you like to join me for coffee?" uttered in appropriate contexts. It is the cancelability or defeasibility of indirect speech that makes people resort to it when, for example, they make a sexual overture because it allows an escape route. One can deny the overture even though it has become common knowledge. Of course, there are many cases where an overture may be merely suggested rather than communicated, in which case it does not become common knowledge. In either case, the overture can be canceled and this is what makes it relatively safe. All of this depends on (a "thick description" of) the context. Once again, see Gilbert Ryle's 1971 university lectures and especially Clifford Geertz (1973).

corrected. But suppose that it did matter to $\mathcal{A}$. Then he might prefer to utter alternative sentences such as φ' = BILL SMITH RAN or φ'' = BILL RAN IN THE ELECTION or even φ''' = BILL SMITH RAN IN THE ELECTION. How might such a preference for a longer sentence requiring more physical effort, however small, be determined? That is, how does the speaker decide when mistakes matter and when they don't?

In circumstances where errors count, the overall cost $k_u^{\mathcal{A}}(\varphi)$ of a shorter, less explicit sentence will be relatively high despite its shorter length and so its net value $v_u^{\mathcal{A}}(\varphi) = v_u^{\mathcal{A}}[G_u(\varphi)] - k_u^{\mathcal{A}}(\varphi)$ will go down. This higher cost component in $k_u^{\mathcal{A}}(\varphi)$ arises by considering the wrong actions the addressee might pursue by inferring an erroneous content. In other words, it is possible to look at the Content Selection Game in Section 7.2, see what payoff is delivered in it if the intended action a is pursued by the addressee (e.g. $v_{\mathcal{A}}$), and compare this payoff with the diminished payoff that would be realized if an undesired action a'' – something like acceptance of an erroneous interpretation – were chosen (e.g. $v''_{\mathcal{A}}$).[15] The difference between the two (i.e. $v_{\mathcal{A}} - v''_{\mathcal{A}}$) would give one the cost of error. This error would occur with some probability, say q, and so the new cost $K_u^{\mathcal{A}}(\varphi) = (1-q)k_u^{\mathcal{A}}(\varphi) + q[k_u^{\mathcal{A}}(\varphi) + (v_{\mathcal{A}} - v''_{\mathcal{A}})] = k_u^{\mathcal{A}}(\varphi) + q(v_{\mathcal{A}} - v''_{\mathcal{A}}) > k_u^{\mathcal{A}}(\varphi)$ when $q > 0$ and $v_{\mathcal{A}} - v''_{\mathcal{A}} > 0$. This is just an average of the old cost without error and the old cost with error and is greater than the old cost without error. This higher cost could tip the balance in favor of a longer, more complex sentence that is more explicit, as there would be a negligible possibility of error with such sentences.

This makes it clear that mistakes count when $v_{\mathcal{A}} - v''_{\mathcal{A}} \gg 0$ or when $q \gg 0$ or both. In the example as described, $v_{\mathcal{A}} - v''_{\mathcal{A}} \not\gg 0$ because a misunderstanding would be just mildly annoying and can be easily rectified. Also, $q \not\gg 0$ because, as I showed, the different locutionary games g_1 and g_2 reinforce each other's equilibrium meanings (viz. σ_1 = *Bill Smith* for g_1 and σ_2 = *ran in an election* for g_2) which makes the likelihood of error small. As neither factor is significant, the overall cost $K_u^{\mathcal{A}}(\varphi) = k_u^{\mathcal{A}}(\varphi) + q(v_{\mathcal{A}} - v''_{\mathcal{A}}) \approx k_u^{\mathcal{A}}(\varphi)$ and so the speaker does not need to reconsider the short sentence φ = BILL RAN. Notice that the actual calculation involved is extremely simple as just the product $q(v_{\mathcal{A}} - v''_{\mathcal{A}})$ has to be evaluated and added to the cost if it is significant. Again, this is well within the psycholinguistic grasp of the speaker and can be carried out in milliseconds without materially altering the time for a person to generate a sentence.

There are situations where the added cost will be significant. To take Benz's (2012) example, if there are two doctors that John might be seeing and he has to be picked up, it may not be enough to say "Please pick up John at the doctor's" as the identity of the doctor has been left implicit. This is so even though

[15] As should be clear, this undesired action a'' has not been explicitly represented in Figure 7.1.

the speaker may be more likely to be referring to a particular doctor and this is common knowledge. While it is possible for the addressee to infer which doctor is intended in such circumstances, the cost of error is high because going to the wrong location to pick up John is very undesirable (i.e. $v_{\mathcal{A}} - v''_{\mathcal{A}} \gg 0$). The probability of error q will depend on how much higher the relevant prior probability corresponding to the intended doctor is than the one corresponding to the unintended doctor in the locutionary game corresponding to the noun phrase "the doctor's." If the former is close to 1, q will be small, but if it is close to 0.5, q will be large. Whatever the value of q, the cost of error will be high owing to $v_{\mathcal{A}} - v''_{\mathcal{A}} \gg 0$ and so the speaker will usually take the trouble to spell out which doctor is intended explicitly. This means that the speaker will instead utter something like "Please pick up John at Dr. X's" which is perhaps a little costlier from a cognitive point of view as something more specific has to be processed, which takes more effort.

To be sure, agents are finite and partially rational and so they may not always carry out the additional calculations on every occasion where they are required. As ambiguity is rife in natural language, there could be errors in interpreting practically every word of an utterance and certainly this would be far too burdensome computationally. Perhaps some possibilities of error are processed at a more conscious level (e.g. those directly related to the addressee's action in the Content Selection Game which in turn is related to the conversational goal) and it is only such potential errors that are averted. So, even though I went through the calculations above for possible misinterpretations with φ = BILL RAN, the ambiguities in these words *in the circumstances described* are likely to be handled nonconsciously by both agents and so no error-related calculations are likely to be carried out at all. In the case of the two doctors, the error calculations are likely as this is a possibility that would be semiconscious if not fully conscious. Such issues can be tested empirically to determine when speakers speak more explicitly and when not.

If some possible misinterpretation is inadvertently missed by the speaker or if mistaken estimates for $v_{\mathcal{A}} - v''_{\mathcal{A}}$ and q are used accidentally, the addressee might notice the possible misinterpretation in the Interpretation Game and ask for a clarification. Such clarifications are not a regular part of the original communication as Benz suggests: they are new utterances in their own right and can take a variety of forms and so cannot be included as mere "feedback" as they need to be disambiguated and interpreted themselves. Why are clarifications triggered? There could be a number of causes: the addressee may infer an unexpected meaning (e.g. *Bill Jones ran in an election*) because she selects the wrong prior probabilities owing to a miscommunication, or she needs to be certain in order to carry out her action in the Content Selection Game because the locutionary game results in a mixed strategy solution or because an error would be costly (i.e. $v_{\mathcal{B}} - v''_{\mathcal{B}} \gg 0$ or her probability of error $q_{\mathcal{B}} \gg 0$ or both).

Benz also misses the other two possibilities: *neither* party may notice potential misinterpretations and the communication may go through smoothly as intended (such as going to the right location to pick up John even though it was left implicit) or an undesirable outcome can occur (such as going to the wrong location to pick up John).[16] Benz overcompensates for errors and attempts to build a model that eliminates them completely. But this is unrealistic because mistakes can and do occur. Curiously, he does not apply his own "error model" to look at how speakers might be more explicit but confines his attention to whether "clarification requests" are generated. The key lack in his model is that there is no estimate of the *magnitude* of error as a trigger either for more explicit sentences or for clarifications, there is just the *intuitive* recognition that errors can occur because communication is probabilistic. There is, in other words, no actual model of error in his error model. This leads to the prediction that practically every utterance will result in "clarification requests" from addressees, even when they are intuitively not warranted. Nevertheless, Benz's observation about the possibility of error in communication, first noted by Zaefferer (1977), is important, even though both Zaefferer and Benz focus exclusively on interpretation whereas its most important impact is on how it compels speakers to be relatively more explicit in some aspects of their utterances in certain circumstances.[17]

[16] The real-life case of Derek Bentley involving an armed robbery in the UK in 1952 where Bentley may have uttered the ambiguous "Let him have it, Chris" to his accomplice is cited by Carston (2008) in a different context. The utterance is ambiguous between *shoot the police officer* and *give the police officer your gun*. See http://en.wikipedia.org/wiki/Derek_Bentley_case. It is not known whether Bentley actually uttered this sentence and, if he did, what he might have meant.

[17] There are other issues with Benz's (2012) model but I cannot go into them here as it would require a detailed description of his model and would be a digression. One point is worth mentioning: he overlooks the fact that the prior probabilities in a game of partial information have to do with what the speaker intends to convey and not directly with the probability of the content as such. In the example of the doctor, the addressee may have to impute priors of 0.5 each (roughly, if fuzzy or interval probabilities are allowed) for the two doctors as there might be no way of telling which one is intended on a *particular* occasion regardless of the fact that John frequents one of them more than the other. This latter fact can inform the probability estimate as part of the overall situation (see Section 10.4 for related considerations) but cannot be identified directly with the probability estimate. If after everything is considered an equiprobable prior distribution is selected, it would lead to a mixed strategy solution and the addressee would not know with certainty where to go to pick up John. That is, the payoff the speaker would receive from the locutionary game for the implicit utterance of "Please pick up John at the doctor's" would be reduced and this would further diminish the net value of the sentence $v_u^{\mathcal{A}}(\varphi) = v_u^{\mathcal{A}}[G_u(\varphi)] - k_u^{\mathcal{A}}(\varphi)$ thereby favoring a more explicit sentence with a pure strategy solution and higher overall net value. In such situations, not only does the cost go up but the value of the game goes down, both leading to a lower net value.

This completes my discussion of how Generation Games are solved. As I have said in my earlier books, a comprehensive theory of cost is required but it is doubtful it can be easily erected as the relevant costs have many sources in general. Cost enters both in the choice of sentence which occurs in the Generation Game *outside* the global games for each sentence and it enters also within the global games via Theorem 8.6 that allows for non-symmetric payoffs. However, in particular contexts it is usually straightforward to know what costs to assign as is evident from the example of the doctor above as well as from the examples discussed in Section 12.2. So it could be said that Equilibrium Semantics does offer a more or less complete locutionary theory modulo costs which can be assigned on a case by case basis just as with the prior probabilities. In any case, while the task is difficult, it is not insurmountable. I have identified practically all the major sources of cost in the foregoing and it is just a matter of tackling them one by one. I have also made plausible when costs are likely to be relatively high or low, which means it is just a matter of assigning appropriate magnitudes to them so that they can all be suitably combined.

I believe that treating natural language communication as cheap talk has been a red herring. Also, without costs, everything would be expressed explicitly, which is absurd. It is possibly because economists have been more concerned with eliminating superfluous equilibria in economic settings that they have been led to ignore some of the issues involved in linguistic communication. We could now consider this motivating problem of economics – how to oust unwanted equilibria in economic games – and see how these aspects of language can help their goals. How, for example, do the right choices emerge in the job market signaling situation studied by Spence (1973) when costly but beneficial interviews are part of the game?

However, my concern is with language, and language is truly rich not just in the quantitative sense assumed in cheap talk games but especially in the logical and rhetorical devices it affords that enable us to persuade one another of our beliefs and desires. But these resources do not come cheap, they come with costs and benefits.

I repeat that I have looked at just the speaker's key decision about what to make explicit and what to leave implicit and have not modeled exactly how the actual words are selected together with their order and structure and pronunciation.

8.5 Solving Communication Games

Having solved the locutionary global game $LG(\varphi)$ and therefore the Interpretation Game $UG_u(\varphi) = UG_u$, and the Generation Game $GG(\sigma) = GG_u$, this leaves just the Content Selection Game CSG_u in the overall Communication Game $\Gamma_u = (SG_u, CSG_u, GG_u(\sigma), UG_u(\varphi), G_u(\varphi)) = (SG_u, CSG_u, GG_u, UG_u, G_u)$.[18] The Setting Game SG_u does not need to be solved explicitly as the content conveyed by the speaker and the action taken by the addressee will automatically contribute to its solution. In the example at hand, as mentioned in Section 7.5, CSG_u is trivial as there is just one action a that $\mathcal{B}$ can select in response to the interpretation $\sigma = \sigma^\ell \odot \sigma^\iota =$ *Bill Smith ran in the local election* derived from the global game. The parse tree t figures only in the latter and, as I said in Section 3.2, the purpose grammatical structure serves is to facilitate communication by making disambiguation easier.

It is worth going back to the chart displayed in Sections 3.1 and 7.7.

<u>Utterance Situation</u>

Setting Game

→ $\mathcal{A}$'s wish to elicit some response from $\mathcal{B}$

→ Content Selection Game

→ $\mathcal{A}$'s equilibrium content

→ Generation Game

→ $\mathcal{A}$'s equilibrium utterance

→ Interpretation Game

→ $\mathcal{B}$'s equilibrium content

→ Content Selection Game

→ $\mathcal{B}$'s equilibrium response

→ Back to the Setting Game

Now, all the lines above involving various equilibria have been filled in. In other words, we have solved the Communication Game Γ_u completely. The example considered was a very simple one but it allowed us to see all the moving parts involved in micro-semantics.

As I promised in Section 7.2, I will now look at a slightly more complex Content Selection Game.

[18] I remind the reader again that the illocutionary global game component of the global game will be addressed in Part IV.

8.6 An expanded Content Selection Game

In Section 7.2, I mentioned the possibility that $\mathcal{B}$ could have a choice between accepting and rejecting what $\mathcal{A}$ conveys. This leads to another branch issuing from the intermediate node in Figure 7.1 labeled a' and corresponding payoffs $v'_{\mathcal{A}}$, $v'_{\mathcal{B}}$. If $v_{\mathcal{A}} = v'_{\mathcal{A}}$, it means $\mathcal{A}$ does not care whether $\mathcal{B}$ accepts or rejects his communication. In this case, there is nothing more to be done except that, once $\mathcal{B}$ has interpreted $\mathcal{A}$'s utterance as before, she will have a choice between a and a' and will choose the action that yields the greater payoff, either $v_{\mathcal{B}}$ or $v'_{\mathcal{B}}$. For example, if $\mathcal{B}$ is inclined to be skeptical about $\mathcal{A}$'s information, she may have $v'_{\mathcal{B}} > v_{\mathcal{B}}$ and may reject σ. Now, if $v_{\mathcal{A}} > v'_{\mathcal{A}}$ rather than $v_{\mathcal{A}} = v'_{\mathcal{A}}$, then $\mathcal{A}$ would care about $\mathcal{B}$'s acceptance and may strive to be more persuasive in imparting his content. One way for him to do this is to convey some additional supporting content that, say, he read σ in a newspaper, which might give $\mathcal{B}$ a reason to believe him.[19] Let σ plus this additional supporting information be σ'. This leads to a new action for $\mathcal{A}$ represented by a new branch issuing from s labeled σ'. This makes the overall tree structure of the game more complex, as shown in Figure 8.9.

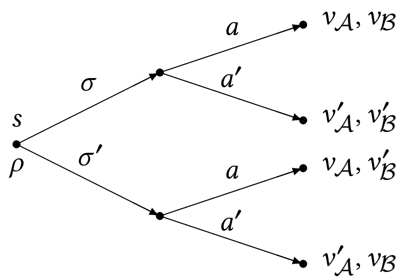


Figure 8.9: An expanded Content Selection Game

In this game, as assumed in the previous paragraph, $v_{\mathcal{A}} > v'_{\mathcal{A}}$ but $v'_{\mathcal{B}} > v_{\mathcal{B}}$ so there is a certain degree of conflict between $\mathcal{A}$ and $\mathcal{B}$. If $\mathcal{A}$ chooses σ, $\mathcal{B}$ would prefer a' or rejection, and if $\mathcal{A}$ chooses the costlier σ', $\mathcal{B}$ would choose a or acceptance. In other words, the distribution of the payoffs – notice that $v'_{\mathcal{B}}$ and $v_{\mathcal{B}}$ have been switched in the lower half of the tree and $\mathcal{A}$ would like $\mathcal{B}$ to choose a with either choice of content – is such that $\mathcal{B}$'s equilibrium action is a' in the

[19] See Cialdini (2006), for example.

upper half of the tree and a in the lower half of the tree, so $\mathcal{A}$ is compelled to convey σ' rather than σ. That is, the equilibrium path in this game is (σ', a) where the first member of the pair is $\mathcal{A}$'s equilibrium action and the second is $\mathcal{B}$'s. This means $\mathcal{A}$ is willing to convey more information in order to win $\mathcal{B}$'s acceptance and this additional information comes at an additional cost because it has either to be converted into a second sentence in the Generation Game or somehow conveyed as an implicature. As set up, the situation u does not make it easy to convey this information as an implicature – additional assumptions would be required – and so a longer sentence than φ = BILL RAN or a second sentence would have to be uttered by $\mathcal{A}$. It can safely be assumed that this additional cost of processing and speaking a little more is smaller than $v_{\mathcal{A}} - v'_{\mathcal{A}}$, the gain in utility $\mathcal{A}$ receives from acceptance by $\mathcal{B}$. Even though σ and σ' are invisible, they have been made perceptible via the linkage of the Content Selection Game with the Generation Game and the Interpretation Game.

This example shows how and why conflict, that is, slightly misaligned payoffs, and related features of the Content Selection Game can lead a speaker to utter a longer and costlier sentence or two. I pointed out in Sections 3.1 and 5.2 that the maxims of Quantity and Manner do not work as Grice had envisaged because a speaker may be driven to convey additional information beyond what is strictly required and utter more than the most perspicuous choice of sentence in order to succeed at his endeavor to evoke a desired response from the addressee.

Content Selection Games can get even more complex and I will describe two such examples, one in Chapter 11 and the other in Section 18.2. As can be seen, even this example introduced strongly evaluative dimensions of action of the kind broached by Taylor (1985/1999) and discussed in Section 6.1 because matters involving the relationship between $\mathcal{A}$ and $\mathcal{B}$ come into play. Indeed, just as many potentially conflicting factors ranging over aesthetic, relational, and error aspects were all assimilated into a single cost $k_u^{\mathcal{A}}(\varphi)$ in Section 8.4, so a very wide spectrum of cooperative and conflictual relationship-related elements all enter into the payoffs and prior probabilities of the Content Selection Game. That is, a very wide array of relationships is packed into relatively few numerical slots. Such relational features can be purely psychological or sociological or even anthropological.

For example, the reason why $\mathcal{B}$ is inclined to be skeptical about $\mathcal{A}$'s information could be almost anything: $\mathcal{A}$ is a known fibber or $\mathcal{B}$ is overly cautious or does not wish to be easily influenced or $\mathcal{A}$ has a vested interest in the information or they are themselves both members of certain political parties or whatever. But all these possibilities lead just to $\mathcal{B}$'s skepticism and to $v'_{\mathcal{B}} > v_{\mathcal{B}}$, that is all. This shows that the utterance situation can easily encompass much richer and more realistic and more nuanced scenarios of the kind we encounter every day.

What the apparatus of a Communication Game does, in and through the Setting Game, Content Selection Game, Generation Game, and Interpretation Game, is to offer a way to *reduce* the great diversity of human and social phenomena to a few numbers and inequalities *within a game-theoretic structure* and thereby make this variety scientifically manageable, as also argued in Chapter 6. For some, like Taylor, this might appear reductionist and unacceptable but no *relevant* aspect of the situation seems left out as far as predicting and explaining communicative behavior is concerned. If a concrete feel for the situation is desired, then one does have to return to the qualitative details of the situation. This is precisely the interplay between abstract and concrete inquiry that is a necessary feature of the social sciences as mentioned in Section 2.1.

9 An example with syntactic ambiguity

9.1 The example

In the simple example I have just finished analyzing, there was no syntactic ambiguity and so the syntactic games of partial information were all trivial. In order to fully justify Theorem 8.5 in Section 8.1, I now consider an example with both semantic and syntactic ambiguity. I will focus exclusively on the locutionary global game LG_u in the overall Communication Game $\Gamma_u = (SG_u, CSG_u, GG_u, UG_u, G_u)$.

Suppose $\mathcal{A}$ is a reporter for the *New York Times* and writes the headline "Fed raises interest" and $\mathcal{B}$ is a reader who reads the headline the next morning. Here φ is the whole sentence and φ_i its ith word. The situation u includes the general goings-on of the week and also related financial news.[1]

Intuitively, there are three possible meanings of this utterance, of which I will consider two:

(10) *the Federal Reserve augments the interest rate*

(11) *the Federal Reserve's increments arouse curiosity*

The second meaning seems improbable but is nevertheless a real possibility. The third possible meaning – *the Federal Reserve augments curiosity* – will be ignored here to keep things simple. If it is assumed as part of u that there have been no recent increases in the interest rate, the intended meaning of the reporter's utterance would be the first one above. As before, the utterance and its interpretation involve four constraints:

- Phonetic Constraint
- Syntactic Constraint
- Semantic Constraint
- Flow Constraint

[1] I will be using the same symbols as before to avoid multiplying symbols unnecessarily.

In general, the Phonetic Constraint generates all the possible words associated with the speech wave corresponding to φ in u, the Syntactic Constraint generates all the possible parse trees of φ in u, and the Semantic Constraint generates all the possible meanings associated with φ in u. The Flow Constraint then disambiguates all these ambiguities simultaneously and identifies the equilibrium or intended content, that is, the words uttered, their parse, and their meaning.

As the example involves a written sentence, the Phonetic Constraint can be omitted. There would be a corresponding Graphic Constraint that involves recognizing the characters, but for our purposes it can just be assumed that the words in the sentence φ are immediately available to $\mathcal{B}$.

9.1.1 The Syntactic Constraint

The two possible parses of the whole sentence generated by an appropriate background grammar G or algebraic system of trees $(\mathcal{T}, \star)$ are:[2]

- $[_{S}[_{NP}[_{N}\,\text{Fed}]][_{VP}[_{V}\,\text{raises}][_{NP}[_{N}\,\text{interest}]]]]$
- $[_{S}[_{NP}[_{N}\,\text{Fed}][_{N}\,\text{raises}]][_{VP}[_{V}\,\text{interest}]]]$

It is the first parse above that is optimal as it corresponds to the first meaning above. Thus, FED has two trees, RAISES has two trees, and INTEREST has two trees. The Syntactic Constraint gives us:

(12) FED:

- $\varphi_1 \longrightarrow [_{NP}[_{N}\,\text{Fed}]] = t_1$
- $\varphi_1 \longrightarrow [_{N}\,\text{Fed}] = t_1'$

FED is a noun in both cases but in the first case it forms a noun phrase by itself and in the second it forms a compound noun phrase with RAISES. In the first case, it can be chained one level up to the NP level owing to a rule such as NP → N whereas, in the second, no chaining is possible because the branch arising from RAISES is encountered owing to a rule such as NP → N N.

(13) RAISES:

- $\varphi_2 \longrightarrow [_{V}\,\text{raises}] = t_2$
- $\varphi_2 \longrightarrow [_{N}\,\text{raises}] = t_2'$

[2]The grammar and algebraic system are naturally different from the ones in Sections 4.2 and 7.3. Also this sentence has a syntactic ambiguity at the lexical level itself so we do not need recourse to the feature-based grammmars of Section 8.3.

(14) INTEREST:

- $\varphi_3 \longrightarrow [_{NP}[_{N}\,\text{interest}]] = t_3$
- $\varphi_3 \longrightarrow [_{VP}[_{V}\,\text{interest}]] = t_3'$

The Syntactic Constraint uses the elementary trees in $(\mathcal{T}, \star)$ to derive the possible parses of each word, phrase, and the whole sentence. Earlier, there was no syntactic ambiguity so there was just one syntactic content per word and for the whole sentence; now, there are two syntactic contents for each word.

9.1.2 The Semantic Constraint

The Semantic Constraint consists of two subconstraints, the Conventional Constraint and the Referential Constraint. The first maps words into their conventional meanings and the second maps each of these conventional meanings into their potential referents relative to u.

Each word in φ has multiple conventional meanings. I will assume the first word FED has just the one relevant conventional meaning and the latter two words have just the two relevant conventional meanings each.[3] Accordingly, the Semantic Constraint gives us:

(15) FED:

- Referential Use: $\varphi_1 \longrightarrow P^{\varphi_1} \xrightarrow{u}$ *the Federal Reserve* $= \sigma_1$

(16) RAISES:

- Predicative Use: $\varphi_2 \longrightarrow P_1^{\varphi_2} \xrightarrow{u}$ *augments* $= \sigma_2$
- Referential Use: $\varphi_2 \longrightarrow P_2^{\varphi_2} \xrightarrow{u}$ *increments*[4] $= \sigma_2'$

(17) INTEREST:

- Referential Use: $\varphi_3 \longrightarrow P_1^{\varphi_3} \xrightarrow{u}$ *interest rate* $= \sigma_3$
- Predicative Use: $\varphi_3 \longrightarrow P_2^{\varphi_3} \xrightarrow{u}$ *arouse curiosity* $= \sigma_3'$

The intermediate conventional meanings are denoted by the various properties $P_j^{\varphi_i}$.

[3] This is just to keep the discussion simple without any loss of generality.

[4] This corresponds to a noun, not a verb.

9.1.2.1 The Flow Constraint

The Flow Constraint gives us six lexical games, three semantic and three syntactic, corresponding to the three words.

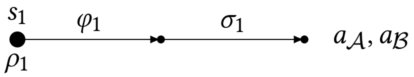


Figure 9.1: Semantic lexical game g_1

In Figure 9.1 showing the first semantic lexical game g_1, s_1 is the initial situation for the writer $\mathcal{A}$ though it also includes the later facts about potential readers, including $\mathcal{B}$. The only action available to $\mathcal{A}$ in s_1 is uttering φ_1 = FED and, since there is only one possible meaning for FED, the addressee has just one possible interpretation σ_1 as specified by the Semantic Constraint. When that action is taken, it leads to the terminal situation at which two payoffs $a_{\mathcal{A}}$, $a_{\mathcal{B}}$ are awarded to the two interlocutors.

The initial symbol ρ_1 represents the prior conditional probability that $\mathcal{A}$ is conveying σ_1 *given* that he is conveying the other meanings and parse trees of the rest of the sentence in the utterance situation u as we have seen before. In other words, $\rho_1 = P(\sigma_1 \mid x_2, x_3, y_1, y_2, y_3; u) = 1$ where the x_i are variables standing for the possible corresponding meanings σ_2, σ'_2 and σ_3, σ'_3, and the y_i are variables standing for the possible corresponding parse trees t_1, t'_1 and t_2, t'_2 and t_3, t'_3. The situation parameter u, which is *not* a random variable, also influences the probability. Again, ρ_1 is a *function* of the conditioning variables and parameter but because there is just a single probability, its value is always 1 independent of the values of the variables.

As before, this semantic lexical game g_1 is obtained by applying the semantic game map g_u to the word φ_1. That is, $g_u(\varphi_1) = g_1$.

The syntactic game in Figure 9.2 is now a little more complex than before as there are two possible contents rather than just one. Notice how all the primes on the symbols work, especially symbols such as $s'_{1'}$ and $\rho'_{1'}$ where there is a prime on the symbol and on the subscript. There is no easy way to avoid this and once it is accepted, all the other symbols that arise flow naturally out of these conventions.

In s'_1 in Figure 9.2, the speaker is assumed to be conveying t_1, although because φ_1 = FED is syntactically ambiguous between t_1 and t'_1, the addressee could

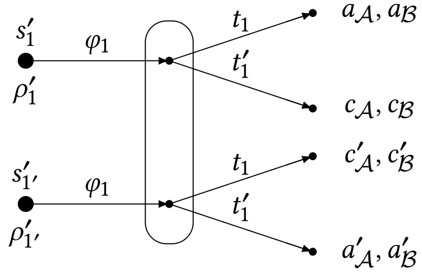


Figure 9.2: Syntactic lexical game g_1'

potentially interpret the word as having the parse t_1' as well. This choice is represented by two branches emanating from the upper intermediate situation inside the information set, the elongated oval, rather than just one as we had in g_1. The payoffs for the two agents are $a_{\mathcal{A}}$ and $a_{\mathcal{B}}$ for making the right choice t_1 as this is what $\mathcal{A}$ is conveying in s_1', and they are $c_{\mathcal{A}}$ and $c_{\mathcal{B}}$ for making the wrong choice t_1' as this is not what $\mathcal{A}$ is conveying in s_1'. The a payoffs are therefore again greater than the corresponding c payoffs as they represent the correct interpretation by $\mathcal{B}$.

In $s_{1'}'$ in the lower half of g_1', a similar choice presents itself to the addressee but this time the payoffs are reversed because in this *hypothetical* situation $\mathcal{A}$ is conveying the other possible parse t_1'. Again, the primed a payoffs are greater than the corresponding primed c payoffs.

The prior probabilities are $\rho_1' = P(t_1 \mid x_1, x_2, x_3, y_2, y_3; u)$ and $\rho_{1'}' = P(t_1' \mid x_1, x_2, x_3, y_2, y_3; u)$ and these must sum to 1 as there are only two possible parse trees the speaker could be conveying. Since these probabilities are also functions of the conditioning variables and parameter u, there are different *versions* of essentially the same game g_1' when, say, x_3 or y_2 take on different values. These versions differ only in the numerical values of the prior conditioned probabilities; in all other respects, they are identical.

This syntactic lexical game g_1' is obtained by applying the syntactic game map g_u' to the word φ_1. That is, $g_u'(\varphi_1) = g_1'$.

As I have remarked before, Equilibrium Linguistics sees both meanings and parse trees as different types of content communicated in an utterance and they are treated analogously. All the possible meanings and all the possible parses of an utterance codetermine one another to yield the equilibrium meaning and parse. Not only does semantics reflect syntax as Frege believed, but syntax also reflects semantics. Now, I show the next set of lexical games for the second word.

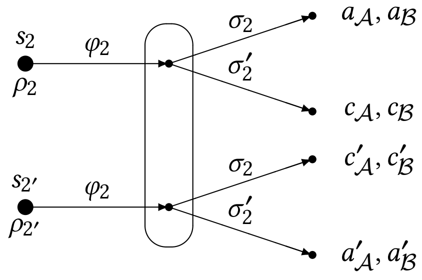


Figure 9.3: Semantic lexical game g_2

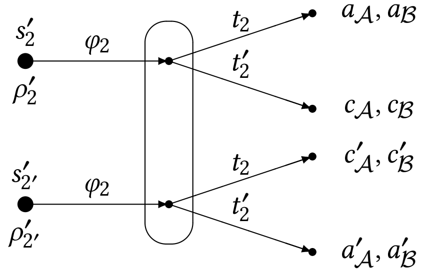


Figure 9.4: Syntactic lexical game g_2'

It should be relatively straightforward to interpret Figures 9.3 and 9.4. In s_2 in Figure 9.3, the speaker is conveying σ_2, although because φ_2 = RAISES is semantically ambiguous as mentioned above, the addressee could potentially interpret the word as conveying σ_2' as well. As one would expect from the foregoing, this choice is represented by two branches emanating from the intermediate situation rather than just one as we had in g_1. The payoffs follow the same pattern as in g_1'.

In $s_{2'}$ in the lower half of the same game, a similar choice presents itself to the addressee but this time the payoffs are reversed because in this hypothetical situation $\mathcal{A}$ is conveying the other possible meaning σ_2'. Again, the primed a payoffs are greater than the corresponding primed c payoffs.

The prior probabilities by analogy with the earlier ones are $\rho_2 = P(\sigma_2 \mid x_1, x_3, y_1, y_2, y_3; u)$ and $\rho_{2'} = P(\sigma_2' \mid x_1, x_3, y_1, y_2, y_3; u)$ and these must sum to 1 as there are only two possible meanings the speaker could be conveying. Since these probabilities are also functions of the conditioning variables and parameter u, there are different *versions* of essentially the same game g_2 when, say, x_3 or y_2 take on different values.

The same kinds of choices and payoffs occur in Figure 9.4 except that this time a syntactic ambiguity has to be resolved as in g'_1. Now the prior conditional probabilities are $\rho'_2 = P(t_2 \mid x_1, x_2, x_3, y_1, y_3; u)$ and $\rho'_{2'} = P(t'_2 \mid x_1, x_2, x_3, y_1, y_3; u)$ which must again sum to 1.

The games g_2 and g'_2 are obtained by applying the maps g_u and g'_u to φ_2. That is, $g_u(\varphi_2) = g_2$ and $g'_u(\varphi_2) = g'_2$.

The third set of lexical games is shown in Figures 9.5 and 9.6.

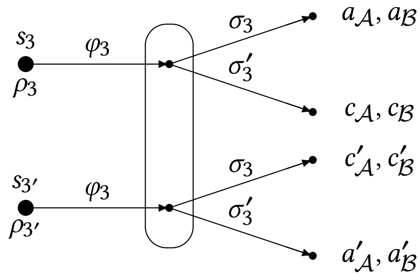


Figure 9.5: Semantic lexical game g_3

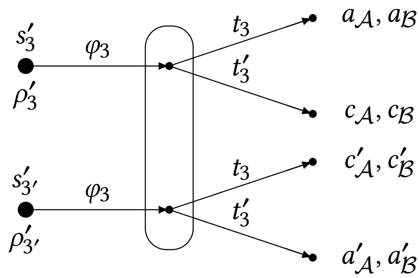


Figure 9.6: Syntactic lexical game g'_3

By now, it should be very easy to interpret the diagrams. Again, φ_3 = INTEREST has two possible meanings and parse trees as discussed above and so the game trees look similar to the ones for φ_2. Just for clarity, I mention that the prior conditional probablities for g_3 and g'_3 are $\rho_3 = P(\sigma_3 \mid x_1, x_2, y_1, y_2, y_3; u)$, $\rho_{3'} = P(\sigma'_3 \mid x_1, x_2, y_1, y_2, y_3; u)$ and $\rho'_3 = P(t_3 \mid x_1, x_2, x_3, y_1, y_2; u)$, $\rho'_{3'} = P(t'_3 \mid x_1, x_2, x_3, y_1, y_2; u)$, respectively, both sets of which must sum to 1.

This completes the discussion of semantic and syntactic lexical games. The next step is to look at the phrasal games that arise as a product of these lexical games. There are two nontrivial phrases, one in each parse, the verb phrase $[_{VP}[_{V}\text{raises}][_{NP}[_{N}\text{interest}]]]$ in the first parse and the noun phrase $[_{NP}[_{N}\text{Fed}]$ $[_{N}\text{raises}]]$ in the second parse shown above. I will start with the semantic phrasal game for the second phrase as it is the simplest. It is shown in Figure 9.7.

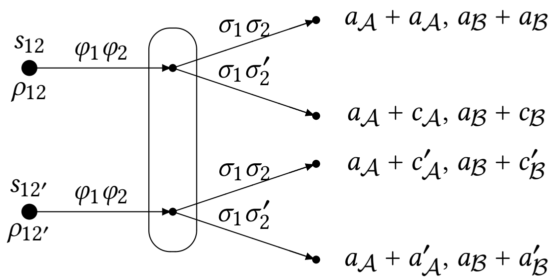


Figure 9.7: Semantic phrasal game g_{12}

As before, the tree diagram is a "product" of the trees of g_1 and g_2.[5] Next, there are two initial situations $s_{12} = s_1 \cup s_2$ and $s_{12'} = s_1 \cup s_{2'}$. The latter situations s_1, s_2, and $s_{2'}$ are the initial situations of the two games g_1 and g_2 that are being multiplied. In other words, the initial situations of the product are derived from the initial situations of the multiplicands as they should be and $1 \times 2 = 2$ such combinations are possible from one initial situation of g_1 and two initial situations of g_2.

The speaker's utterance $\varphi_1\varphi_2$ is a concatenation of the individual utterances φ_1 and φ_2 from g_1 and g_2. And the addressee's possible actions are $\sigma_1\sigma_2$ and $\sigma_1\sigma_2'$, where again the components are obtained from the respective games. As before, the payoffs corresponding to the relevant branches in the multiplicands are added to give the payoffs in the relevant branch of the product. For example, the payoff for $\mathcal{A}$ in the topmost branch corresponding to the path s_{12}, $\varphi_1\varphi_2$, $\sigma_1\sigma_2$ is $a_{\mathcal{A}} + a_{\mathcal{A}}$, which is the sum of the $a_{\mathcal{A}}$ in g_1 corresponding to the path s_1, φ_1, σ_1 and the sum of the $a_{\mathcal{A}}$ in g_2 corresponding to the path s_2, φ_2, σ_2.

Finally, the priors $\rho_{12} = P(\sigma_1, \sigma_2 \mid x_3, y_1, y_2, y_3; u)$ and $\rho_{12'} = P(\sigma_1, \sigma_2' \mid x_3, y_1, y_2, y_3; u)$ which must sum to 1. Again, it should be easy to see how these probabilities are generated from the corresponding probabilities in the multiplicands

[5]The game tree product is, of course, different from the parse tree product described earlier in Section 4.2.

g_1 and g_2 where the priors are $\rho_1 = P(\sigma_1 \mid x_2, x_3, y_1, y_2, y_3; u)$ and $\rho_2 = P(\sigma_2 \mid x_1, x_3, y_1, y_2, y_3; u)$, $\rho_{2'} = P(\sigma_2' \mid x_1, x_3, y_1, y_2, y_3; u)$.

The product game $g_{12} = g_1 \otimes g_2$ can also be obtained directly by applying the semantic map g_u to the phrase $\varphi_1 \circ \varphi_2$. In other words, $g_{12} \equiv g_u(\varphi_1 \circ \varphi_2) = g_u(\varphi_1) \otimes g_u(\varphi_2) \equiv g_1 \otimes g_2$.

Turning to the corresponding syntactic game, recall that the two trees of FED are $t_1 = [_{\mathrm{NP}}[_{\mathrm{N}}\,\mathrm{Fed}]]$ and $t_1' = [_{\mathrm{N}}\,\mathrm{Fed}]$, and the two trees of RAISES are $t_2 = [_{\mathrm{VP}}[_{\mathrm{V}}\,\mathrm{raises}]]$ and $t_2' = [_{\mathrm{N}}\,\mathrm{raises}]$. This means there are four potential parses of the phrase $\varphi_1\varphi_2$:

$$
\begin{aligned}
t_1 \star t_2 &= [_{\mathrm{NP}}[_{\mathrm{N}}\,\mathrm{Fed}]] \star [_{\mathrm{V}}\,\mathrm{raises}] = t_0 \\
t_1 \star t_2' &= [_{\mathrm{NP}}[_{\mathrm{N}}\,\mathrm{Fed}]] \star [_{\mathrm{N}}\,\mathrm{raises}] = t_0 \\
t_1' \star t_2 &= [_{\mathrm{N}}\,\mathrm{Fed}] \star [_{\mathrm{V}}\,\mathrm{raises}] = t_0 \\
t_1' \star t_2' &= [_{\mathrm{N}}\,\mathrm{Fed}] \star [_{\mathrm{N}}\,\mathrm{raises}] = [_{\mathrm{NP}}[_{\mathrm{N}}\,\mathrm{Fed}][_{\mathrm{N}}\,\mathrm{raises}]] = t_{1'2'}
\end{aligned}
$$

The only nonzero product corresponding to FED RAISES is the compound noun under an assumed grammatical rule like NP → N N or, equivalently, a tree such as $[_{\mathrm{NP}}[_{\mathrm{N}}\ ][_{\mathrm{N}}\ ]]$ which would left multiply t_1' and then t_2' as follows:

$$
\begin{aligned}
t_1' \star t_2' &= [_{\mathrm{N}}\,\mathrm{Fed}] \star [_{\mathrm{N}}\,\mathrm{raises}] \\
&= ([_{\mathrm{NP}}[_{\mathrm{N}}\ ][_{\mathrm{N}}\ ]] \triangleleft [_{\mathrm{N}}\,\mathrm{Fed}]) \triangleleft [_{\mathrm{N}}\,\mathrm{raises}] \\
&= [_{\mathrm{NP}}[_{\mathrm{N}}\,\mathrm{Fed}][_{\mathrm{N}}\ ]] \triangleleft [_{\mathrm{N}}\,\mathrm{raises}] \\
&= [_{\mathrm{NP}}[_{\mathrm{N}}\,\mathrm{Fed}][_{\mathrm{N}}\,\mathrm{raises}]] \\
&= t_{1'2'}
\end{aligned}
$$

Notice how the primes on $t_{1'2'}$ work: we need to prime both subscripts 1 and 2 as the corresponding subscripts on both the multiplicands are primed. A product such as $t_1 \star t_2' = t_{12'} = t_0$ and likewise with the other products as there are presumably no rules that correspond to such products in G.

The resulting syntactic phrasal game is shown in Figure 9.8.

This product should be straightforward to interpret. It follows the same pattern as Figure 9.7. For example, there are four initial nodes as there are two initial nodes s_1', $s_{1'}'$ in g_1' and two initial nodes s_2', $s_{2'}'$ in g_2' and $2 \times 2 = 4$. These situations are the corresponding unions of the initial situations of the multiplicands, and so on with the rest of the entities. We have again $g_{12}' \equiv g_u'(\varphi_1 \circ \varphi_2) = g_u'(\varphi_1) \otimes' g_u'(\varphi_2) \equiv g_1' \otimes' g_2'$.

There is a similar semantic phrasal game $g_{23} \equiv g_u(\varphi_2 \circ \varphi_3) = g_u(\varphi_2) \otimes g_u(\varphi_3) \equiv g_2 \otimes g_3$ with $2 \times 2 = 4$ initial nodes as can be seen from Figures 9.3 and 9.5. Likewise,

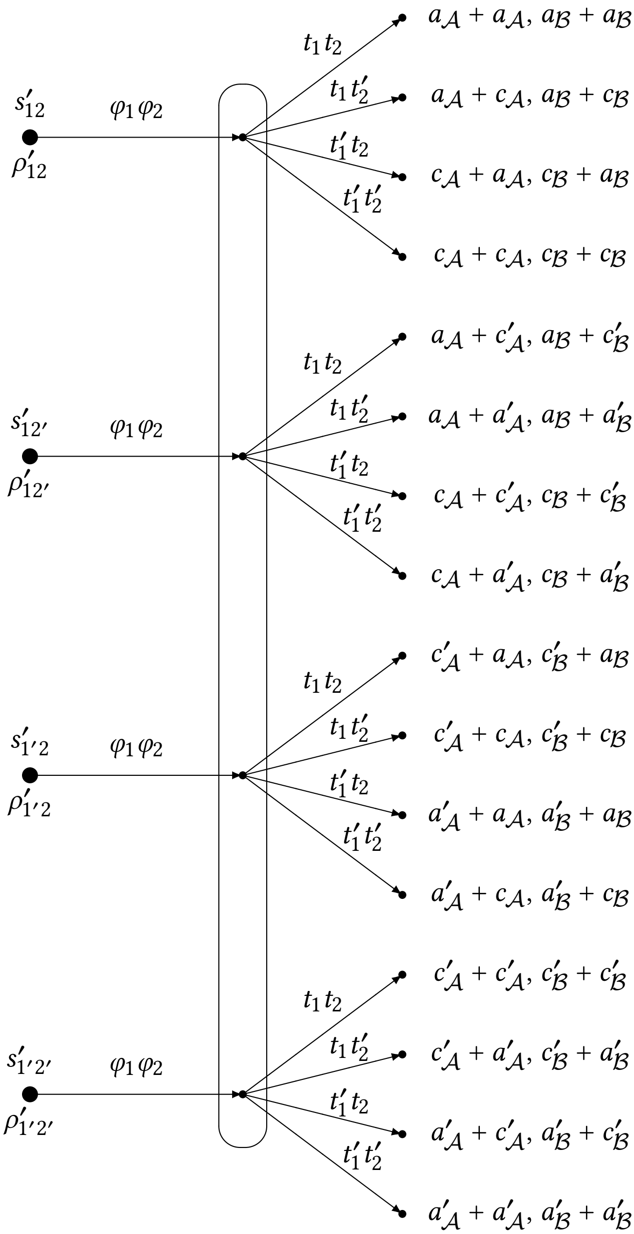


Figure 9.8: Syntactic phrasal game g'_{12}

there is a four initial node syntactic phrasal game $g'_{23} \equiv g'_u(\varphi_2 \circ \varphi_3) = g'_u(\varphi_2) \otimes' g'_u(\varphi_3) \equiv g'_2 \otimes' g'_3$ as can be seen from Figures 9.4 and 9.6. In the case of the latter game, it should be possible to see by inspection that the only nonzero parse tree product is $t_2 \star t_3$ corresponding to the verb phrase $[_{\mathrm{VP}}[_{\mathrm{V}}$ raises$][_{\mathrm{NP}}[_{\mathrm{N}}$ interest$]]]$. I will not tarry to draw these games.

The next two games are the semantic and syntactic sentential games corresponding to the sentence $\varphi_1\varphi_2\varphi_3$. Both games can be obtained in either of two ways. For the syntactic sentential game, we stipulate that $g'_{123} = g'_1 \otimes' g'_{23} = g'_{12} \otimes' g'_3$ despite the lack of associativity of $\star$. This is done by mandating that when forming a three-term product such as $(t_1 \star t_2) \star t_3$ or $t_1 \star (t_2 \star t_3)$, which have different results, an intermediate nonzero grouping should be favored and the other one dropped. Since $(t_1 \star t_2) = t_0$ but $(t_2 \star t_3) \neq t_0$, it is the latter that should be retained in place of the former (even if the final three-term product proves to be zero). If this were not stipulated, there would be two distinct syntactic sentential games which is undesirable because then the two possible parses of the sentence $t_1 \star (t_2 \star t_3)$ and $(t'_1 \star t'_2) \star t'_3$ cannot be directly compared in the same game. This turns out to be felicitous for other reasons too as the set of syntactic games remains associative with respect to $\otimes'$. The reader is urged to write out the sixteen possible parse tree products of the three lexical trees and see how eight of them are discarded in forming the syntactic game. This stipulation is just part of the overall way in which the product of syntactic games is defined, that is all.

In exactly the same way, for the semantic sentential game too, we stipulate that $g_{123} = g_1 \otimes g_{23} = g_{12} \otimes g_3$ despite the lack of associativity of $\odot$. This game also has eight initial nodes just like the syntactic sentential game, and eight zero products equal to the zero $\mathbf{0}$ of the infon lattice introduced in Section 2.1 are discarded. As I have not described this operation in any detail in this book, I will simply say the matter is analogous to what happens in the case of the syntactic sentential game.

Finally, we would form the full mixed semantic-syntactic sentential product as before by multiplying the semantic sentential product with the syntactic sentential product. This is straightforward to do and I leave it to the reader to confirm that it would have as many as $8 \times 8 = 64$ initial nodes and, likewise, 64 interpretive branches emerging from each of the 64 intermediate nodes in the information set for $\mathcal{B}$. This confirms that these full mixed sentential product games grow rapidly in size as also mentioned on page 110 in Section 7.3. Recall that we do not need a Pareto-Nash equilibrium for these full mixed products; just the Nash equilibrium will do, and if there is more than one, we just assign them equal probabilities.

We can collect all these interdependent games as before into the locutionary global game $LG(\varphi) = \{g_1, g_2, g_3, g_1', g_2', g_3'\}$ where the product games have been dropped as all the required information is contained in the lexical games as indicated by Theorem 8.1 in Section 8.1. I have chosen to describe the distributed form of $LG(\varphi)$ rather than the compact form which would involve a $2n = 2 \times 3 = 6$ dimensional matrix as there are $n = 3$ words in the sentence uttered. Such a high-dimensional matrix can be represented by several two-dimensional matrices.

All that remains is to show that the global Pareto-Nash equilibrium of all these games yields the intended meaning $\sigma = \sigma_1(\sigma_2\sigma_3)$ and parse $t = t_1(t_2t_3)$ for the utterance. This can be seen informally by noting which prior probabilities are higher in each of the lexical games and ensuring the corresponding solutions are compatible with one another as indicated by Theorem 8.5 in Section 8.1. The fact that this is easy to do by inspection is reassuring because the calculations would happen in milliseconds in the heads of the writer and reader of the sentence as only a few probabilities have to be compared in a compatible manner.

Assuming just basic facts about ontology, the language, grammar, and partial rationality, we have been able to compute the intended parse and meaning of the utterance in the presence of both syntactic and semantic ambiguities. This approach can be extended to any utterance in principle. Aspects of syntax such as *movement* (e.g. *wh*-movement) require some small adjustments that are straightforward to devise.

The full content conveyed by $\mathcal{A}$ is the proposition $p = (c \vDash \sigma)$ where c is the described situation as before and could be something like the (financial) news of the day. As noted earlier, the latter's boundaries are somewhat indeterminate.

I leave it to the reader to see how the solution to the locutionary global game fits into the Generation Game on the speaker's side and into the Interpretation Game on the addressee's side and to make appropriate assumptions to construct a Content Selection Game for the example, which could be assumed to be trivial.

9.2 Locutionary meaning

The locutionary meaning of an utterance is the meaning that comes "directly" from the words in the relevant utterance situation. As such, finding it involves just disambiguating between alternative referential meanings of words that have homonymous or polysemous conventional meanings and fixing the reference of pronouns. It does *not* involve other phenomena such as free enrichment, modulation, implicature, and direct and indirect illocutionary force. All of the latter belong to the illocutionary meaning of an utterance, although modulation straddles locutionary and illocutionary meaning as will be seen in Chapter 17. I will say more about this distinction between locutionary and illocutionary meaning and the crosscutting distinction between literal meaning and implicature at the end of Part IV.

In the foregoing, I have shown in a very thorough manner how lexical and structural disambiguations work. This can be extended to fixing pronoun reference in a straightforward way as that problem is also just one of disambiguation among alternative possible referents. All aspects of communication involving alternative possibilities, whether of a locutionary or illocutionary kind, are just problems of disambiguation and the Flow Constraint treats them all uniformly. The only difference that arises is in how the possibilities are generated. In the case of fixing pronoun references, if the pronouns are anaphoric, the possible referents come from the sentence or the discourse, and if they are deictic, they are available directly in the utterance situation u or in induced resource situations r_u that depend on u. The rudimentary model described in Clark & Parikh (2007) can easily be extended to cover all such cases so I will not say more about this problem here (except briefly in Section 18.2) and will take the problem of locutionary communication as fully solved.

It is fitting to first solve this problem completely in a detailed way and then look at some general philosophical results that concern the system of locutionary communication and, in particular, games of partial information. I now turn to this latter task.

10 Universality, Frege's principles, indeterminacy, and truth

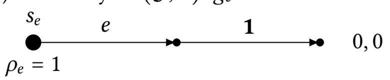
10.1 The universality of games of partial information

The games of partial information shown so far were formed with respect to particular examples but it is possible to construct such games more generally by the maps g_u and g'_u on utterances to obtain two monoidal systems $(\mathcal{G}, \otimes)$ and $(\mathcal{G}', \otimes')$ with zeroes as mentioned in Section 4.3.[1] The solution to locutionary global games can then be compactly described by the commutative diagram in Figure 10.1.

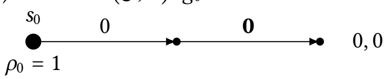
$\mathcal{C}_u$ and $\mathcal{C}'_u$ are the semantic and syntactic content maps and they can be decomposed into two sets of maps (g_u, g'_u) and (f_u, f'_u) where the latter represent functions from the relevant games to their solutions. In general, families of interdependent games corresponding to a whole utterance will need to be considered

[1]The identity of $(\mathcal{G}, \otimes)$ is the empty game g_e, which is just the trivial game with a single initial node (with prior probability 1) and a branch labeled with e, the empty string, issuing from it, a further branch labeled $\mathbf{1}$, the identity of $(\mathcal{I}, \odot)$, issuing from the node that ends the first branch, and any terminal payoff, preferably 0. The reader should check that for any game $g \in \mathcal{G}$, we get $g \otimes g_e = g_e \otimes g = g$ and that $g_u(e) = g_e$ and $f_u(g_e) = \mathbf{1}$. There is also a similar zero game g_0 with a single branch labeled 0, the zero of $(\mathcal{L}, \circ)$, and a further branch labeled $\mathbf{0}$, the zero of $(\mathcal{I}, \odot)$, such that $g \otimes g_0 = g_0 \otimes g = g_0$ for all $g \in \mathcal{G}$ and that $g_u(0) = g_0$ and $f_u(g_0) = \mathbf{0}$. Pictorially:

(i) Identity of $(\mathcal{G}, \otimes)$: g_e



(ii) Zero of $(\mathcal{G}, \otimes)$: g_0



The identity g'_e and zero g'_0 of $(\mathcal{G}', \otimes')$ can be defined analogously by replacing $\mathbf{1}$ with t_e, the identity of $(\mathcal{T}, \star)$, and by replacing $\mathbf{0}$ with t_0, the zero of $(\mathcal{T}, \star)$, in the above games.

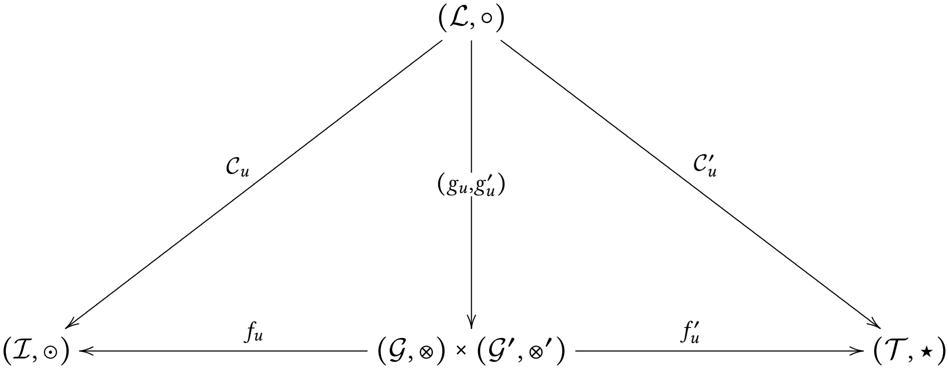


Figure 10.1: The commutative diagram: $\mathcal{C}_u = f_u \circ g_u$ and $\mathcal{C}'_u = f'_u \circ g'_u$

in order to specify a solution as their interdependence through the prior conditional probabilities requires their simultaneous solution. This diagram makes the uniform treatment of all contents by Equilibrium Linguistics even clearer. The following fundamental results establishing the universality of games of partial information for the semantics and syntax of natural language are stated without proof as they are straightforward.[2] They are so basic that from the standpoint of Equilibrium Linguistics they could well be called "The Main Theorems of (Linguistic) Communication."

Theorem 10.1. $g_u : \mathcal{L} \longrightarrow \mathcal{G}$ and $g'_u : \mathcal{L} \longrightarrow \mathcal{G}'$ are isomorphisms.

Theorem 10.2. Given a semantic interpretation function $\mathcal{C}_u : \mathcal{L} \longrightarrow \mathcal{I}$ and given the semantic game function $g_u : \mathcal{L} \longrightarrow \mathcal{G}$, there is a unique function $f_u : \mathcal{G} \longrightarrow \mathcal{I}$ such that $\mathcal{C}_u = f_u \circ g_u$.

Theorem 10.3. Given a syntactic interpretation function $\mathcal{C}'_u : \mathcal{L} \longrightarrow \mathcal{T}$ and given the syntactic game function $g'_u : \mathcal{L} \longrightarrow \mathcal{G}'$, there is a unique function $f'_u : \mathcal{G}' \longrightarrow \mathcal{T}$ such that $\mathcal{C}'_u = f'_u \circ g'_u$.

These theorems imply that there is always a unique solution to the locutionary global game. It may happen that the solution involves a mixed strategy and so in fact involves multiple contents, each with some probability as in a pun, but this mixed strategy is then uniquely given. These results involving the universality of g_u and g'_u and therefore of semantic and syntactic games of partial

[2] They require formal definitions of g_u and g'_u as shown in the Appendix.

information have the following fundamental consequence: if there is *any* other theory (e.g. mainstream rule and convention based semantical theory, syntactic theory, optimality theory, relevance theory, etc.) that succeeds in providing an account of the locutionary content functions $\mathcal{C}_u$ and $\mathcal{C}'_u$, the theorems say that it *has* to be essentially equivalent to the maps f_u and f'_u. Thus, they state that games of partial information are essential to communication or are "universal."[3] If we wish to represent mixed semantic-syntactic products then the two triangles in Figure 10.1 can be folded into one by defining just a single monoid of semantic-syntactic games and a single map to it and a single solution map from it.

Finally, if phonetics is included in the framework, there would be a further set of phonetic games and more or less the same kind of treatment would be available where the words uttered and the equilibrium parse and meaning can be simultaneously derived from the speech wave produced by a speaker.

10.2 Frege's compositionality and context principles

It should be evident that Theorem 8.5 in Section 8.1 is related to Frege's century-old principle of compositionality and context principle because it gives a way to determine the referential meaning of an utterance from first principles just as these principles attempt to do. The compositional principle states that the meaning of a composite expression is a function of the *independently given* meanings of its component parts and the context principle states that the meaning of a word or subsentential expression is what it contributes to the meaning of the whole sentence. Many, notably Dummett (1993: 4–5), have remarked on the tension between these two principles because one requires the meaning of a sentence to depend on the meanings of its constituent words and the other requires the meanings of the constituent words to depend on the meaning of the sentence.[4]

The first thing to note is that the Fundamental Equation of Equilibrium Linguistics is about utterances rather than sentences as contrasted with Frege's two principles. That is, it takes account fully of the context in which sentences are

[3] For more about this, see Parikh (2010: Section 4.11).

[4] There are different ways of interpreting especially the context principle as pointed out by Matilal & Sen (1988). Indeed, they also show that there were quite remarkable debates in classical Indian philosophy of a similar sort from roughly the 5th century CE to the 17th. Their key proponents espoused "sentence holism" (similar to the context principle) and what is called "designation before connection" (similar to compositionality) and "connected designation" (intermediate between the two principles). These discussions went on for centuries. It is a shame that Western philosophers do not generally include such non-Western ideas in their writings.

uttered as sentences by themselves cannot convey contents.[5] I believe there is currently no other competing account that shows how to compute the (locutionary) meaning of an utterance from scratch. Following Grice (1989c), the tendency is simply to say that it is given by convention with the actual processes of lexical and structural disambiguation and pronoun reference fixing left completely mysterious.

The second thing the Fundamental Equation makes clear is that the optimal (referential) meanings of words, subsentential expressions, and sentences are *interdependent*: they codetermine one another via the interdependence of prior probabilities and the concomitant interdependence of the corresponding locutionary games of partial information as expressed through a fixed point process. This shows that Frege's principle of compositionality does not hold for referential meanings because the meanings of the component parts in turn depend on the meanings of the other parts and also on the whole. It is possible to re-express the fundamental equation as follows:

$$\begin{aligned} \mathcal{C}_u(\mathrm{S}) &= h_1(\mathcal{C}_u(\mathrm{NP}), \mathcal{C}_u(\mathrm{VP})) \\ \mathcal{C}_u(\mathrm{NP}) &= h_2(\mathcal{C}_u(\mathrm{S}), \mathcal{C}_u(\mathrm{VP})) \\ \mathcal{C}_u(\mathrm{VP}) &= h_3(\mathcal{C}_u(\mathrm{S}), \mathcal{C}_u(\mathrm{NP})) \end{aligned}$$

where h_1, h_2, h_3 are suitable functions derived from Equation 6 in Section 8.1.5. The function h_1 in particular is obtained by taking the product of the relevant components of $z^\star$ in Equation 6 to get $\mathcal{C}_u(S)$. A similar result can be derived from the more general Equation 7 in Section 8.1.5 as well. The noun phrase NP and verb phrase VP can be further broken down into their constituents in these equations. It is the presence of the second and third equations above that disqualifies the Fregean principle of compositionality since the principle requires the meanings of the component parts to be determined independently and directly. This set of equations can perhaps be more perspicuously expressed as follows:

$$\begin{aligned} \mathcal{C}_u(\mathrm{S}) &= h_1(\mathcal{C}_u(\mathrm{NP}), \mathcal{C}_u(\mathrm{VP})) \\ \mathcal{C}_u(\mathrm{NP}) &= h_4(\mathcal{C}_u(\mathrm{VP})) \\ \mathcal{C}_u(\mathrm{VP}) &= h_5(\mathcal{C}_u(\mathrm{NP})) \end{aligned}$$

[5] See Strawson (1956) and Parikh (2010: 20–21 and Section 2.7) and Section 1.2 of this book.

and as follows:

$$
\begin{array}{rcl}
\mathcal{C}_u(\mathrm{S}) & = & h_1(\mathcal{C}_u(\mathrm{NP}), \mathcal{C}_u(\mathrm{VP})) \\
\mathcal{C}_u(\mathrm{NP}) & = & h_6(\mathcal{C}_u(\mathrm{S})) \\
\mathcal{C}_u(\mathrm{VP}) & = & h_7(\mathcal{C}_u(\mathrm{S}))
\end{array}
$$

the first set corresponding to the first equality in Equation 6 and the second set to the second equality.

Here, too, it is the second and third equations in each set that disqualify Fregean compositionality. Only in the special case where there are no lexical ambiguities will the Fregean principle work. In general, we will have to resort to Theorem 8.5 to determine the meanings of all expressions, whether simple or composite. This is a complex process of simultaneous interaction between the potential meanings of each component part and whole which results in the actual (optimal) meanings of the parts and the whole. Thus, the *principle of compositionality* has to be replaced by the more general *fixed point principle* as indicated by Equation 6 or its more general variant Equation 7.

For greater clarity, I urge the reader to see how the examples we have just discussed, BILL RAN and FED RAISES INTEREST, fit into these equations. For example:

$$
\begin{array}{rcl}
\mathcal{C}_u(\textsc{bill ran}) & = & h_1(\mathcal{C}_u(\textsc{bill}), \mathcal{C}_u(\textsc{ran})) \\
\mathcal{C}_u(\textsc{bill}) & = & h_4(\mathcal{C}_u(\textsc{ran})) \\
\mathcal{C}_u(\textsc{ran}) & = & h_5(\mathcal{C}_u(\textsc{bill}))
\end{array}
$$

and:

$$
\begin{array}{rcl}
\mathcal{C}_u(\textsc{bill ran}) & = & h_1(\mathcal{C}_u(\textsc{bill}), \mathcal{C}_u(\textsc{ran})) \\
\mathcal{C}_u(\textsc{bill}) & = & h_6(\mathcal{C}_u(\textsc{bill ran})) \\
\mathcal{C}_u(\textsc{ran}) & = & h_7(\mathcal{C}_u(\textsc{bill ran}))
\end{array}
$$

Also, since there are no conventional meanings at the level of phrases or sentences, Frege's principle automatically fails to hold at the level of conventional meaning.

Thus, what I am calling the fixed point principle represents a generalization of Frege's principle of compositionality. But these very equations, the third set in particular, also show how the former principle generalizes the context principle because the optimal referential meanings of subsentential expressions depend on the optimal referential meaning of the whole sentence *and* vice versa. Indeed, at this more general level, both principles are reconciled without any tension at all between them. This is a remarkable result.

But this is not all. The *third* observation is that implicit in Frege's compositional principle (and, arguably, in the context principle) is the priority of syntax over semantics and the idea that semantics *mirrors* syntax. The function of the meanings of the component parts of a sentence depends on first optimally parsing the sentence. It is only on the basis of this optimal parse that the component meanings can be properly *identified* and *composed.* The fixed point principle, on the other hand, asserts that semantics and syntax, and phonetics for that matter, are completely interdependent and more loosely *reflect* one another. *Nothing* has priority; everything is simultaneous and interdependent! In particular, each bit of semantics influences each bit of syntax and vice versa. Each component z^*_k of the vector z^* is affected by all the other components z^*_{-k} (not to mention by all the suboptimal content values) irrespective of whether they are semantic or syntactic. All the (possible) contents are intermixed. When an utterance is being processed in real time, the fixed point principle will be modified to take account of the temporal appearance of words as I discuss in Section 12.1. But, in either case, the process of inferring the locutionary *content* of an utterance – its semantic, syntactic, and phonetic values – involves a thoroughgoing interdependence. This, then, is a further generalization of both the compositionality principle and the context principle of Frege.

Not only does Equation 6 provide a way to reduce the computation of locutionary content to a mere comparison of probabilities (or expected values in a more general setting as indicated by Equation 7), that is, to mere arithmetic, this fixed point principle of Equilibrium Linguistics encompasses all the relevant dimensions – generalizing from sentence to utterance, generalizing the nature of the interdependence between word meaning and sentence meaning, and generalizing the relation between syntax and semantics – seamlessly. Frege's principle of compositionality and the context principle are completely transcended.

In fact, if these results are seen in the context of the discussion about Wittgenstein, Austin, and Grice in Section 3.2 and Chapter 5, they show how certain aspects of both Fregean ideas about reference and Wittgensteinian, Austinian, and Gricean ideas about use and communication are not only unified but also superseded. As I said in Parikh (2010) and Chapter 6, the fundamental concepts of semantics are *reference*, *use*, *indeterminacy*, and *equilibrium.* The foregoing shows in a precise way how three of these four notions are unified in Equilibrium Linguistics. The earlier ideas are also superseded because they are rendered more precisely and perhaps more generally.

Interestingly, Equation 6 and its more general variant Equation 7 also apply to other symbol systems, including images and gestures. I discuss in Parikh (2010: Section 7.5) how the meanings of pictures, for example, can be inferred from appropriate part-whole or mereological relationships in a visual "utterance." The syntactic side of the equation has to be interpreted carefully as most symbol systems lack the elaborate structure that language possesses.

10.3 Indeterminacy

Indeterminacy, the fourth notion mentioned above, is dealt with in some detail in Parikh (2010: Chapter 5) so I will be brief here. It is pervasive in both locutionary and especially illocutionary communication for a variety of reasons. One of them is the occasional absence of full common knowledge with the result that each component of the relevant Communication Game can separate into distinct objective and subjective games. Apart from the speaker and addressee inferring slightly different contents, there can also be outright miscommunication. Another factor is the occasional presence of mixed strategies in the global game owing to equal prior probabilities in one or more games of partial information as can happen with puns where the multiple meanings are intended. An interesting aspect of indeterminacy, mentioned at the end of Section 7.4, is the uncertainty surrounding each agent's inferred meaning in the other agent's mind owing to the invisibility of interpretations and, often, also of addressee responses. A key source, perhaps completely overlooked by all writers on meaning, is the indeterminacy of utterance situations. When u is differently carved out by the interlocutors, as happens frequently in literature and art and less often in ordinary communication, there is no unique meaning that can be ascribed. Because the speaker's intentions can be partially *implicit* in examples like "His weight is 150 lbs.," the primacy given to speaker meaning is purely a prejudice issuing from the misplaced Gricean focus on speaker meaning. Addressee meanings are equally important and, in general, there will be as many meanings as there are agents in a conversation. As a result, there is also no need to proclaim *the death of the author* as Barthes (1977) did a half-century ago or Wimsatt & Beardsley (1946) did much earlier, although less dramatically. However, Barthes was more correct than the New Criticism of Wimsatt and Beardsley and others in allowing for an open text with open-ended interpretations, a consequence Equilibrium Linguistics effortlessly enables simply by altering u and related aspects of the discourse in a systematic and methodical manner rather than through vague declarations of the kind many writers influenced by Continental philosophy revel in.

A major source of indeterminacy that I did not consider in my previous book is vagueness. I give a thorough account of this phenomenon in the next chapter.

I will also say a little more in Part IV about indeterminacy as it relates to illocutionary meaning because that is where the bulk of the indeterminacy lies, especially when u is itself indeterminate, but it should be apparent how this fourth concept also ties in smoothly with reference, use, and equilibrium, and flows ineluctably out of the framework. It is something that has largely been neglected in the literature, partly owing to a lack of precise ways to handle it and partly owing to sheer prejudice against the very idea, as it threatens the rigid picture of meaning the ideal language philosophers bequeathed to the field. Uncertainty, generally, can be quite unsettling. But when the tools of probability and game theory are brought to bear on the phenomenon, it is possible to handle it with precision and rigor.

10.4 Meaning and truth

So far, we have managed to derive content from use without any mention of truth. However, there is a minor need for truth in some cases in computing content. Since interpretation involves disambiguation, the truth of a possible proposition, if independently available, can help in this task. If the situation described by an utterance is c and if c is, say, perceptually accessible, then a possible content σ can be evaluated against c to see if it makes the proposition $c \models \sigma$ true. If it does, then this fact can bolster the choice of σ against some other possible content σ'. Naturally, there will be many other factors that also push toward one or other choice of content so truth cannot play an overriding role; it is just one among many pushes and pulls on alternative contents.

Specifically, the way in which truth enters the computation of content is via the prior probabilities of local partial information games that are members of the global game. As we have seen, the priors are affected by multiple objective and subjective elements, and truth, when it is independently available, is one of them. This kind of role does not result in the kind of vicious circularity between meaning and truth referred to in Dummett (1996) and in Section 1.2.

The determination of truth can occur through the fixing of the described situation but the selection of resource situations and even the utterance situation itself can also play a role. Thus, this role of truth in deriving meaning can be reversed and help in fixing or narrowing these three types of situation as well: described, resource, and utterance. When this happens, meaning and truth become more interdependent. For example, a particular choice of described situation might make a particular choice of content more likely via the resulting proposition's being true, and vice versa, the choice of that content might make that corresponding described situation more likely. Such mutual reinforcement and "circularity" are not vicious because the *possible* meanings are not defined in terms of truth as they are based wholly on the Phonetic, Syntactic, and Semantic Constraints which, in turn, depend only on the grammar and on conventional meaning and on the utterance situation. Thus, only the *choice* of optimal meaning may be based partly on truth.

It has never been clear exactly what the place of Lewis's (1981) principle of accommodation is within a theory of meaning. Now that the role of truth in the derivation of meaning from use has been spelled out, it should be possible to extend it to include Lewis's idea but I do not pursue this here.

11 Vagueness

Practically every word in a natural language is vague. And yet, the attention this fact has received in semantics on the whole is surprisingly scant, perhaps because quite new ideas and tools are required to accommodate it. Borderline cases and, indeed, borders that shift with context pose additional difficulties in understanding communication. Without an account of vagueness, any semantic theory is seriously incomplete.

There have been a number of recent attempts to understand vagueness including Keefe & Smith (1996); de Jaegher (2003); Puglisi et al. (2008); Lipman (2009); Égré & Klinedinst (2011); Jäger et al. (2011); Hampton & Jönsson (2012); and Blume & Board (2014). This is a large area and I focus mostly on foundational issues.

I start by characterizing clear cases and borderline cases of a vague concept. Next I tackle the *sorites paradox*. It is perhaps this classical puzzle that has occupied most writers on vagueness, especially philosophers and logicians. I then look briefly at the important subclass of vague concepts called essentially contested concepts. Last, I apply these insights to communication.

I approach these tasks by adapting models from cognitive psychology as it appears that psychologists understand this domain better than philosophers or linguists owing to the psychologists' emphasis on experimental data. I also apply these adapted models in new philosophical and linguistic ways. With the right models, the sorites paradox yields to a natural and intuitive resolution. Essentially contested concepts become easier to understand. Once I describe vagueness, it can be readily incorporated into a larger theory of communication and meaning.

I first clarify some basic terminology. Concepts are taken to be mental representations of collections of things and categories the collections themselves. Concepts correspond to properties or attributes or features, all terms referring to abstract entities. So an agent's vague concept *bald* corresponds to a vague property of baldness. Since each agent will have a slightly different concept *bald*, the corresponding vague property can be thought of in two ways: as a kind of social and abstract average of these individual representations[1] and as a kind of abstract individual counterpart to the concept, one for each agent. Both kinds of property are important, the average kind and the individual kind. Vague words such as BALD have vague concepts conventionally attached to them that serve as their conventional meanings.

It has been taken for granted from classical times until Wittgenstein (1953/1968: Sections 66 and 67, pages 31–32) questioned it that most concepts have clear definitions, that is, noncircular necessary and sufficient conditions. This implies every object is or is not a member of the corresponding category. As discussed by Smith & Medin (1981), this classical view and its variants are untenable primarily because most concepts are vague and have borderline cases.[2] Here is Murphy's (2004: 21) account of *why* vagueness is ubiquitous:

> The Necessity of Category Fuzziness
>
> The existence of unclear examples can be understood in part as arising from the great variation of things in the world combined with the limitations on our concepts. We do not wish to have a concept for every single object – such concepts would be of little use and would require enormous memory space. Instead, we want to have a relatively small number of concepts that are still informative enough to be useful (Rosch 1978). The ideal situation would probably be one in which these concepts did pick out objects in a classical-like way. Unfortunately, the world is not arranged so as to conform to our needs.
>
> ⋮
>
> The gradation of properties in the world means that our smallish number of categories will never map perfectly onto all objects: the distinction between members and nonmembers will always be difficult to draw or will even be arbitrary in some cases. If the world existed as much more distinct clumps of objects, then perhaps our concepts could be formed as the classical view says they are. But if the world consists of shadings and gradations and of a rich mixture of different kinds of properties, then a limited number of concepts would almost have to be fuzzy.

[1] As I said in Section 2.1, properties are individuated from *reality* by agents and so are social but abstract constructs that nevertheless have a certain objectivity. For example, the number 5 can be thought of as being abstracted from collections of five objects just as the property of being blue can be thought of as being discriminated from blue objects.

[2] The other important theoretical reason is that they are unable to account for typicality effects. And much of the experimental evidence disconfirms them.

As described by Murphy (2004: Chapter 3), there are three new views that have emerged: the exemplar approach, the prototype approach, and the knowledge approach. The first uses the information provided by each encounter with an exemplar for a category separately; the second works with a *summary representation* of a category derived from the experience of exemplars; and the third integrates concepts with the broader knowledge schemata in which they must reside via plausible reasoning.

Because concepts are used in very diverse tasks, none of these approaches is able to account for all the empirical data. Indeed, as Murphy (2004: Chapter 13) concludes, some amalgam of the three will probably be required for a "Big Theory of Concepts" as they also appear to be somewhat complementary in their explanatory adequacy. In other words, each approach focuses on a different source of information, and any one or more of these sources may be summoned for a particular task based on its suitability.

I will adapt the first two approaches for my purposes. The third knowledge approach is, in a sense, not really an independent stand-alone approach but one that operates by combining knowledge effects with one or both of the other approaches. I construct the simplest models required for the tasks at hand and do not aim at more comprehensive versions.

11.1 Basic setup

Both the exemplar and prototype approaches rely on the idea that each exemplar of a concept has multiple properties that take on particular values. For example, the concept *bald* may involve features such as the number of hairs on the scalp, the number of completely hairless patches on the scalp, the fraction of the scalp that is hairless, and so on;[3] each exemplar will instantiate these attributes with particular numbers. In other words, each concept is associated with an n-dimensional attribute space and each exemplar can be represented by a point in this space. Some dimensions may be continuous and some may be discrete.

Let b_i with $i = 1, 2, \ldots, N$ be clear exemplars of *bald* that an agent $\mathcal{A}$ has encountered in his experience. Likewise, let $b'_{i'}$ with $i' = 1, 2, \ldots, N'$ be clear exemplars of *not bald*. A negative category such as the latter is a little unusual in that it contains not just persons with a full head of hair but also random items such as clocks

[3]Such a listing of features is an idealization as they are not entirely independent of one another. It is not clear, however, whether agents actually operate with completely independent attributes. Presumably, this depends on what they know and this is one way in which knowledge effects may enter.

and cars. There is no problem with this as all potential exemplars are assessed relative to the relevant attributes which come from the corresponding positive category. Thus, only persons with relatively full heads of hair will qualify as exemplars and items such as clocks and cars will be discarded as junk. A different way to think about negative categories is that in any particular situation where its exemplars are accessed, there will always be a default reference category that will automatically limit the possibilities to the relevant types of individuals. In the case of *not bald*, the possibilities will be limited to persons; in the case of *not tall* to men or women or basketball players, depending on the situation; and in the case of *not chair*, the default category might be items of furniture.

Let x_{ij} be the value of the jth attribute of b_i and, similarly, let $x'_{i'j}$ be the value of the jth attribute of $b'_{i'}$. That is, $b_i = (x_{ij})$ and $b'_{i'} = (x'_{i'j})$, the right-hand side of both equalities being vectors with $j = 1, 2, \ldots, n$.

Now suppose $\mathcal{A}$ has to judge whether the candidate a is bald or not bald or borderline bald in u. Then a will also be a point in the same space with value x_{aj} in the jth dimension. That is, $a = (x_{aj})$.

The basic idea underlying both approaches is to see how "far" a is from all the exemplars taken separately or from an "average" exemplar (i.e. the prototype) and, based on this, to see how similar a is to the other members in the category. This computation allows $\mathcal{A}$ to decide where a stands with respect to *bald*.

11.1.1 The exemplar model

This model has its roots in Medin & Schaffer (1978); Nosofsky (1992); and Nosofsky & Palmeri (1997). I build upon the description in Murphy (2004: 65–71). Schiffer (2010) informally mentions the possibility of using weighted distance in the context of vagueness.

In order to get at the psychological distance between a and b_i, we need to first note the following. For certain attributes such as the number of hairs on an individual's scalp, if a's value x_{a1} is less than b_i's value x_{i1} then the psychological difference between these values along this dimension is not $|x_{a1} - x_{i1}|$ but 0 because b_i is an *exemplar* and a has, so to speak, met the *bar* set by b_i. Likewise, if the attribute is the number of completely hairless patches on the individual's scalp, and if $x_{a2} > x_{i2}$, then again the difference is 0 by the same reasoning. There may, of course, be attributes where only an exact equality $x_{aj} = x_{ij}$ results in a zero difference.[4] So we can define a psychological difference function $\delta(x_{aj}, x_{ij})$

[4] For example, the category *blue* is such because overshooting the relevant color frequency in either direction counts as a nonzero difference. With such attributes, only an exact equality results in a zero difference.

which is either 0 or $|x_{aj} - x_{ij}|$ based on the nature of the concept and attribute being considered.[5]

Now define the weighted psychological distance between a and b_i as follows:

$$d_u(a, b_i) = \sqrt{\sum_{j=1}^{n} w_j(u)\ \delta(x_{aj}, x_{ij})^2}$$

Here, $w_j(u)$ are weights issuing from the situation u. The psychological distance that $\mathcal{A}$ perceives between a candidate and an exemplar thus varies with the situation he is in. This variation implies that certain attributes and therefore certain exemplars will play a more or less important role in $\mathcal{A}$'s judgment.

This distance function is *not* a metric in the technical sense as it is not symmetric: $d_u(a, b_i)$ may not equal $d_u(b_i, a)$ because the underlying psychological difference function δ is not symmetric. Also, many different forms for it can be used; I have restricted myself to the commonest Euclidean variety.

Correspondingly, the weighted psychological distance between a and $b'_{i'}$ will be:

$$d_u(a, b'_{i'}) = \sqrt{\sum_{j=1}^{n} w'_j(u)\ \delta(x_{aj}, x'_{i'j})^2}$$

Shepard (1987) has shown that behavioral similarity between items is an exponentially decreasing function of their psychological distance.

$$s_u(a, b_i) = e^{-c(u)\, d_u(a, b_i)}$$

where $c(u) > 0$ is a situation-based parameter. Again, a larger or smaller $c(u)$ will determine the relative importance of items that are near and items that are far.

Analogously:

$$s_u(a, b'_{i'}) = e^{-c'(u)\, d_u(a, b'_{i'})}$$

Finally, define the psychological probability that a is bald rather than not bald for the agent $\mathcal{A}$ as follows:

$$P(bald|\, a; u) = \frac{\sum_{i=1}^{N} s_u(a, b_i)}{\sum_{i=1}^{N} s_u(a, b_i) + \sum_{i'=1}^{N'} s_u(a, b'_{i'})}$$

[5]There is some empirical warrant for such a result as reported in, for example, Hampton et al. (2005).

Then:

$$P(not\ bald|\ a;u) = \frac{\sum_{i'=1}^{N'} s_u(a, b'_{i'})}{\sum_{i=1}^{N} s_u(a, b_i) + \sum_{i'=1}^{N'} s_u(a, b'_{i'})}$$

Note that $P(bald|\ a;u) + P(not\ bald|\ a;u) = 1$. These psychological probabilities are measured with respect to the agent $\mathcal{A}$ as they are based on his exemplars. So they are agent-relative probabilities. But they are not "subjective" probabilities in the usual sense of being $\mathcal{A}$'s beliefs.

11.1.2 The prototype model

This model has its roots in Rosch & Mervis (1975) but the account below is based on a certain natural construal of a summary representation of a category.

The only difference between the exemplar model and the prototype model is that the latter does not compute the psychological distance between the candidate and each exemplar separately as above but first averages the values of all the exemplars and then computes the distance from this average.

So we first define the average values as follows:

$$\overline{x_j} = \frac{\sum_{i=1}^{N} w_i(u) x_{ij}}{N}$$

$$\overline{x'_j} = \frac{\sum_{i'=1}^{N'} w'_i(u) x'_{i'j}}{N'}$$

This tells us that the prototypes for *bald* and *not bald* are just $\overline{b} = (\overline{x_j})$ and $\overline{b'} = (\overline{x'_j})$. The weights $w_i(u)$, $w'_i(u)$ are different from the earlier weights above described in the exemplar model, and are also indexed with respect to i and not j as before. These weights play an important role because sometimes extreme examples such as a completely hairless person count as prototypes. Such extreme exemplars can be selected as the relevant prototypical average by adjusting the weights suitably. Alternatively, they can be selected as the minimum or maximum of the relevant attribute values.[6]

We can use the same idea for the psychological difference as before except that it is measured with respect to the average values. So $\delta(x_{aj}, \overline{x_j})$ is either 0 or

[6] Generalized means are a family of functions for aggregating sets of numbers and we can draw upon any of these based on the nature of the concept and its attributes. See, for example, http://en.wikipedia.org/wiki/Generalized_mean.

$|x_{aj} - \overline{x_j}|$ based on the nature of the concept and attribute being considered and likewise with $\delta(x_{aj}, \overline{x'_j})$.

Now the weighted psychological distance from the prototype is defined as follows:

$$d_u(a, \overline{b}) = \sqrt{\sum_{j=1}^{n} w_j(u)\ \delta(x_{aj}, \overline{x_j})^2}$$

$$d_u(a, \overline{b'}) = \sqrt{\sum_{j=1}^{n} w'_j(u)\ \delta(x_{aj}, \overline{x'_j})^2}$$

This in turn leads to similarity.

$$s_u(a, \overline{b}) = e^{-c(u) d_u(a, \overline{b})}$$

$$s_u(a, \overline{b'}) = e^{-c'(u) d_u(a, \overline{b'})}$$

And finally to the psychological probabilities for $\mathcal{A}$ as above.

$$P(bald \,|\, a; u) = \frac{s_u(a, \overline{b})}{s_u(a, \overline{b}) + s_u(a, \overline{b'})}$$

$$P(not\ bald \,|\, a; u) = \frac{s_u(a, \overline{b'})}{s_u(a, \overline{b}) + s_u(a, \overline{b'})}$$

As stated above, the key difference is that distances and similarities are measured with respect to a "summary representation," an *average* (or generalized mean) of all the exemplars.

Both models give us somewhat different ways to compute the same psychological probabilities $P(bald \,|\, a; u)$ and $P(not\ bald \,|\, a; u)$. I now put them to use.

11.2 Characterizing vagueness

Intuitively, if a candidate is sufficiently similar to clear exemplars of both *bald* and *not bald*, it is reasonable to think it is a borderline case. This suggests the following definitions:

Definition 11.1. A candidate a is a borderline case of a concept C for an agent $\mathcal{A}$ in situation u if and only if $|P(C \,|\, a; u) - P(not\ C \,|\, a; u)| < \epsilon_u$ where $0 < \epsilon_u < 1$

is $\mathcal{A}$'s threshold in u. If a is not a borderline case, then it is classified as clearly belonging to C if and only if $P(C \mid a; u) > P(not\ C \mid a; u)$ and as clearly belonging to *not C* if and only if $P(C \mid a; u) < P(not\ C \mid a; u)$.

Definition 11.2. A concept is vague for an agent $\mathcal{A}$ in situation u if and only if it has borderline cases. Otherwise, it is precise or "classical."

The decision to count an item as borderline or clear is *derived from* agent-relative psychological probabilities but is not itself probabilistic or belief-based (e.g. a is borderline with "subjective" probability q). So it is doubly immune to the charge against subjective probability views made by Schiffer (2003: Chapter 5): it is not based on beliefs and it is deterministic.

The threshold ϵ_u is a kind of limiting value but it should not be confused with any *direct* cutoff between clear and borderline cases as that would be assuming what has to be established. It operates at the more basic level of psychological probabilities and *enables* us to draw a line between clear and borderline cases. Moreover, ϵ_u can be conceived as a precise number or a fuzzy number.[7] It should also be seen as something the agent does not know for himself in different situations. Lastly, it arises through communicative interactions and so agents in the same community tend to share it to a greater degree than intuition might suggest. Here, knowledge effects of the kind alluded to earlier when I described the knowledge approach to concepts may play an important role as such thresholds tend to partly arise also from the goals and interests of agents.

$\mathcal{A}$'s cognitive system determines when a case is borderline or clear. It is not entirely a conscious decision. This is confirmed by the familiar feeling of being stymied when we are asked to make a conscious judgment about a borderline case. There is simply no way to reason *decisively* about it based on the external facts.

An item a that is a borderline case for one agent need not be so for another agent. Likewise, the borderline cases of a concept shift with u for the same agent because all aspects of the definition depend on u, the probabilities as well as the threshold. The same agent $\mathcal{A}$ may choose to call Bill bald in one situation but

[7] A fuzzy number A is generally expressed as a function $A : \mathbb{R} \longrightarrow [0, 1]$ such that:

$$A(x) = \begin{cases} f(x) & \text{for } x \in [a, b] \\ 1 & \text{for } x \in [b, c] \\ g(x) & \text{for } x \in [c, d] \\ 0 & \text{for } x < a \text{ and } x > d \end{cases}$$

where $a \leq b \leq c \leq d$, f is a continuous function that increases to 1 at point b, and g is a continuous function that decreases from 1 at point c. See Klir et al. (1997: 170).

not in another as the sentence BILL IS BALD may be true for $\mathcal{A}$ in one situation, false for $\mathcal{A}$ in another, and indeterminate for $\mathcal{A}$ in a third. Consider the sentences SHE WON'T DATE BILL – HE'S BALD and BILL ISN'T BALD – HE NEEDS A HAIRCUT. In the first case, the situation u is such that a less stringent membership condition for *bald* is operative, either because relatively less-bald exemplars are weighted more or because the magnitude of the threshold ϵ_u is relatively smaller making the penumbra, the region of borderline cases, correspondingly smaller as well, or possibly because both factors apply simultaneously. In the second case, the situation u is the opposite – a more stringent membership condition is used. This explains what Schiffer (2010) calls *penumbral shift* in a natural way.

In order to obtain "natural" concepts of the kind that would be useful in thinking and communication, we must assume that the positive exemplars (i.e. the set $\{b_i\}$ for *bald*) are so distributed that all members of their convex closure are instances of C. Otherwise, we would get strange and seemingly arbitrary outcomes for what belongs to C, what is borderline C, and what is not C. This requirement translates into a restriction on ϵ_u: it must be sufficiently small. In other words, if $b = (b_i)$, $b' = (b'_{i'})$, the latter in each case being the vectors of positive and negative exemplars of *bald*, then $0 < \epsilon_u < \zeta_u(b, b') < 1$ where $\zeta_u(b, b')$ is a function of the N positive exemplars and N' negative exemplars that derives from the convexity assumption. The same kind of condition is obviously not required for the negative exemplars as they can, in general, lie anywhere outside the convex closure of the positive exemplars in the n-dimensional attribute space. A candidate a that is judged to belong to C *need not* lie within the convex closure. All that is required is that it be sufficiently close to it. Indeed, subsequently, a would become a positive exemplar itself and the convex closure could be correspondingly enlarged. This suggests a dynamic model of concept learning that results in possibly expanded convex closures as more exemplars are encountered. After a while, the category would converge to a convex polytope in the attribute space with somewhat different boundaries for different situations u.

This convexity assumption is very similar to the convexity assumption made by Gärdenfors (2000) and Warglien & Gärdenfors (2013). My approach of using exemplars to derive concepts seems to allow a clearer development of these ideas from a more foundational starting point. Also, their decision to banish the *external* significance of language from their model seems unnecessary and raises too many problems (e.g. Putnam 1975; Barwise & Perry 1983: 28–31). As I show presently, conventional meanings are mental representations but referential meanings are external entities such as the individuals and properties that make up propositions. It is possible to have one's cake and eat it too.

What may be true or false or indeterminate for one agent may not be so for another. There is *no* agent-independent or objective truth value in other words. However, because the agents must belong to the same linguistic community and the exemplars each agent draws upon are often shared through communication, they may agree more often than expected. In fact, for a concept to be socially useful as most are, its exemplars must be *sufficiently* shared among the community. This points to a community model of interacting agents where concepts are constantly being revised to have sufficient overlap.

The definitions above suggest that if the threshold is a precise, nonfuzzy number there is no *higher-order vagueness*, the phenomenon that the borderline between clear and borderline cases is itself unclear, resulting in borderline borderline cases and so on ad infinitum. If we wish to allow for higher-order vagueness, the threshold can be identified with a fuzzy number. In this case, there is no precise cutoff between clear and borderline cases and higher-order vagueness can be admitted.[8] The evidence seems to indicate that higher-order vagueness is real. This implies that the threshold ϵ_u must be a fuzzy number and not a precise number.

The mistake "epistemicicists" such as Williamson (1994) seem to make is to assume the existence of sharp cutoffs between clear cases of a concept and its complement. That is, they not only reject higher-order vagueness but also first-order vagueness, which is completely unrealistic.

Indeed, it is possible to *characterize* higher-order vagueness by treating Definition 11.1 as a base case for an inductive definition. The key idea is to identify exemplars at each level, and therefore, psychological distance, similarity, and psychological probabilities at each level. For example, Definition 11.1 provides a precise or fuzzy account of first-order borderline cases. Then we can identify positive and negative exemplars for what is clearly borderline and what is clearly not borderline which, in turn, gives rise to psychological distance and similarity and psychological probabilities at the next level. The latter can then be used in a manner analogous to Definition 11.1 to define borderline borderline cases or, in other words, second-order vagueness. And so on to higher-order vagueness for all n.

In the definition below, I assume that we have the exemplars for nth-order vagueness and therefore the nth-order psychological probabilities $P_n(C_n \mid a; u)$

[8] I believe my model is richer than Hampton's (2007) because it allows for both graded membership in a category as well as fuzzy judgments of when a case is clear or borderline whereas Hampton (2007: 377) only allows for the former while pointing to the latter as important experimental evidence. Something like Definition 11.1 would have to be added to his model.

and $P_n(\textit{not}\ C_n \mid a; u)$ where C_n is the nth-order concept of being nth-order borderline C and *not* C_n is the corresponding complementary concept. When $n = 0$, this is understood as just standing for the concepts C and *not* C.

Definition 11.3. A candidate a is an $(n+1)$st-order borderline case of a concept C for an agent $\mathcal{A}$ in situation u if and only if $|P_n(C_n \mid a; u) - P_n(\textit{not}\ C_n \mid a; u)| < \epsilon_{n,u}$ where $0 < \epsilon_{n,u} < 1$ is $\mathcal{A}$'s threshold in u. If a is not an $(n+1)$st-order borderline case, then it is classified as being clearly nth-order borderline or clearly not nth-order borderline according to whether $P_n(C_n \mid a; u)$ or $P_n(\textit{not}\ C_n \mid a; u)$ is greater.

Combining this definition with Definition 11.1 yields a characterization of higher-order vagueness for all n. Now, the threshold $\epsilon_{n,u}$ has to be understood as fuzzy and this gives rise to a fuzzy *fractal*[9] set with no crisp boundaries even in the limit. There is some indirect evidence for the fractal nature of higher-order vagueness in Hampton et al. (2012). In practice, of course, an agent will not actually possess or construct the threshold $\epsilon_{n,u}$ for all n, only for first- and possibly second-order borderline cases as there is no practical utility in having such higher-order thresholds.

Vague properties can now be easily characterized.

Definition 11.4. A property is vague if and only if it is based on the community's corresponding vague concepts, either as an average or as an individual counterpart.

Definition 11.4 covers both types of property, the average kind and the individual kind. It is deliberately vague as there are somewhat messy issues relating to what happens if some members of the community have incorrect concepts and also if some members are vague and others are precise about the same concept. As should be obvious, the same sorts of observations, mutatis mutandis, hold for vague properties as for vague concepts. However, propositions involving average vague properties are objectively true or false or indeterminate as the latter are derived from the concepts of all the individuals in a community. But in general there is no way to know with certainty which of these truth values actually obtains as we have only approximate epistemic access to such a property.

Incidentally, the foregoing observations about how properties result from the averaging of individual concepts may provide a new way of defining properties as social constructs in a precise way. All sorts of averaging operations may be utilized for this purpose based on the particular property being defined.

[9] A fractal is an object or quantity that displays "self-similarity" on all scales. See https://en.wikipedia.org/wiki/Fractal.

11.3 The sorites paradox

The sorites paradox can be formulated for any vague property. It consists of the following type of argument:

1. A hairless person is bald.
2. For all k, if a person with k hairs is bald, then a person with $k + 1$ hairs is bald.
3. Therefore, all persons are bald.

Most proposed solutions to the paradox deny the second premise on either semantic or epistemic grounds. Schiffer (2003: Chapter 5) does an able job of dispelling such proposals. My resolution also denies the same premise but on *psychological* grounds. First, consider an agent-relative concept-based (or individual property-based) restatement of this premise:

- For all k, if $\mathcal{A}$ judges a person with k hairs to be bald then he would judge a person with $k + 1$ hairs to be bald.

The key to the resolution is that such judgments are made on the basis of *multiple* exemplars or a category prototype which is also based on multiple exemplars. So it is quite possible for $\mathcal{A}$ to judge a person with k hairs to be bald and then judge a person with $k + 1$ hairs to be only borderline bald for some definite or fuzzy $k^\star$. This would be true even if we restricted our attention to all the exemplars in the sequence, that is, $1, 2, 3, \ldots, k$ that have already been judged by $\mathcal{A}$ to be bald. As the value of k increases, the distance from the early members of the series (or from the dynamically changing prototype) keeps growing, their similarity keeps dropping, and, at $k^\star$, $\mathcal{A}$ finds himself with a borderline case that is not clearly bald. This follows easily from Definition 11.1.

I am *not* saying that there is a definite cutoff between bald and not bald as many attempts at solution do; I am saying there is either a precise or fuzzy cutoff between clearly bald and borderline bald (and also between borderline bald and not bald). Whether $k^\star$ is a precise or fuzzy number depends on whether $\mathcal{A}$'s threshold ϵ_u is a precise or fuzzy number. Moreover, $\mathcal{A}$'s own cutoff will change with the situation u in which he is asked to make the judgments because ϵ_u depends on u. Finally, different agents will have different cutoffs because their thresholds will generally be a little different.

The reason why the sorites argument seems plausible is that its formulation tricks us into consciously focusing on just a single exemplar: the previous case

k in the second premise. Because our judgments about vague concepts are typically subpersonal and nonconscious, we are not aware of the multiple exemplars $1, 2, 3, \ldots, (k-1)$, and k (and we are not aware of the thresholds) that go into our judgments. Indeed, our judgments also lack the kind of firm conviction we have in judging that 2 + 2 = 4 and we may waffle over the exact value of $k^\star$. So the sorites works by forcing us to make intermediate judgments in a conscious and unnatural way and then freeing us to judge the conclusion that all persons are bald in a nonconscious natural way.

Since properties are abstract social constructs built out of community members' individual concepts, it follows that the agent-independent property version of the sorites paradox will also have the same kind of resolution. That is, there will be some function of the individual $k^\star$ values for each agent that yields some social cutoff $K^\star$ although it will not be possible for anyone to know what its precise or fuzzy value is.

11.4 Essentially contested concepts

Because practically every word in a natural language is vague, practically every word is potentially evaluative, that is, it involves standards of judgment. These standards result from the *exemplars* present in the definition of vagueness and the calculation of distance from them and also from the thresholds. This observation applies not only to weighty concepts like *art*,[10] *science*,[11] *politics*, and *good* but also to mundane ones like *bald* and *tall*. Such concepts are inherently contestable as pointed out by Gallie (1956) in a pioneering analysis because, in the context of my model, individual and group agents may have somewhat different exemplars, may consider somewhat different attributes for each exemplar, or may weight these dimensions differently, and, as a consequence, arrive at potentially quite different concepts with substantially less overlap than I have discussed so far. My use of the term is wider than Gallie's as he restricts it to traditional areas of philosophy such as ethics, aesthetics, and political philosophy where contests over the meaning of a word are endless and conceptual differences are greater. But admittedly transitory and less material contests also occur over everyday concepts such as *bald* that can therefore be assimilated to the same idea.

Besides, not all of Gallie's necessary conditions seem essential to the core notion. He offers seven: the concept must be evaluative, internally complex, variously describable, and undergo penumbral shift; the parties to the contest must

[10]The ensuing discussion has some relevance for attempts to define art. See Adajian (2012).

[11]The discussion that follows has some relevance for the demarcation problem in the philosophy of science. See Popper (1935/2002).

be aware of the contest and must share an original exemplar; and the contest itself must enable the original exemplar's "achievement" to be optimally developed or at least sustained. Of these, the first four result from my model of vagueness, and the last three can be dropped. In particular, a vague concept's attributes make it internally complex and also variously describable by altering either some attributes themselves or their values or their relative weights. It is also susceptible to penumbral shift. This suggests the following definition.

Definition 11.5. A concept is essentially contestable if and only if it is vague.

Since most concepts are vague and therefore evaluative, language is an agonistic site of innumerable large or small potential contests against a background of *partially* shared (conventional) meanings. The actual disputes that take place over the meaning and use of terms reflect the large or small differential *interests* of the members of the linguistic community. Two individuals or groups may diverge over whether someone is bald or, more significantly, over whether some policy is democratic.[12] Thus vagueness has an extremely important consequence: it gives language a *normative* and even a *moral* dimension. While the gross structure of meanings classifies and partially constitutes the entities individuated by a society as described in Section 2.1 and Chapter 6, its fine structure as given by the definitions in this chapter corresponds to and partially constitutes the contrasting values that prevail in society. As Wittgenstein (1953/1968: Section 242, 88) said: "If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments." If communication is understood as an identity between content conveyed and content grasped, then Wittgenstein is fully right but, as I argued in Parikh (2001; 2006b; 2010), this identity condition is just an ideal limiting case that is seldom realized in practice. Less than ideal communication is in fact the norm. This entire book also reinforces the same point.

In Section 6.2, I described Taylor's (1980/1999) arguments for an expressive dimension of language. It should now be clear that the constitutive role he assigns to language is not restricted just to subject-referring emotions, social relationships, and moral values but is widespread as it applies equally to ordinary words such as BALD and TALL. The uses of such words, too, partially constitute their contents. This kind of constitution does *not* make these contents *linguistic*: they

[12] Connolly (1993) offers an illuminating discussion of the contests surrounding fundamental concepts in political theory such as *politics*, *interests*, *power*, and *freedom*. He goes on to show how these debates are constitutive of politics itself and therefore how conceptual revision is a necessary condition for political change.

retain their abstract character. Moreover, vagueness in language is essential to human beings because, given the fuzzy nature of the world as described in the quote at this chapter's start, it allows us to agree and disagree in subtler and more efficient ways about such matters that can only be subjective or intersubjective but not objective, and the possibility of such flexible (dis)agreement is necessary to our emotions, concerns, relationships, and moral lives.

Lastly, because properties may be viewed as appropriately derived averages of the individual concepts in a community, it is not just vague language but also vague properties that are normative in the sense discussed. Since most concepts are vague, most properties are vague, and thus we get the somewhat startling view that the individuated informational space or world described in Section 2.1 is largely one based on human norms.[13] As I said in Section 1.2, one of the benefits of approaching meaning via communication is that the normative character of language and the world emerge in an especially clear light.

11.5 Back to communication

Recall from Section 7.7 that the Communication Game $\Gamma_u = (SG_u, CSG_u, GG_u(\sigma), UG_u(\varphi), G_u(\varphi)) = (SG_u, CSG_u, GG_u, UG_u, G_u)$. As I have addressed locutionary communication quite thoroughly already, I will confine myself to a few remarks about where vagueness may alter the foregoing analysis. It turns out that the conceptual difficulties it poses do not overly complicate our semantic frameworks.

One somewhat new phenomenon that arises with vague concepts is how precise an agent needs to make his utterance to balance the conflicting demands of costs and benefits. If $\mathcal{A}$ makes the content more precise, then it may be more costly but may also yield greater benefits depending on the nature of his goals. This is similar but not identical to the issue of how much to explicitly disambiguate the lexical and structural ambiguities in a sentence as discussed in Section 8.4. There, the issue had to do with the particular *sentence* that was optimal in GG_u to capture the same optimal content from CSG_u. With vagueness, the issue is how precise to make the content itself and, therefore, the corresponding sentence. This requires us to look at CSG_u rather than GG_u.

For example, the question before $\mathcal{A}$ might be whether he should convey the content σ = *Bill is very bald* or σ' = *Bill is bald* to $\mathcal{B}$ in some utterance situation

[13] Vagueness is not the only language-related source of normativity. The Romantic conception of language provides another way for norms to arise through language. A third way is based on the speech act theory of Austin (1975) and especially Habermas (1976/1998; 1988/1998).

u based on some Setting Game SG_u. The former content would more tightly circumscribe the range of baldness conveyed and this may be desirable on account of the response $\mathcal{A}$ wishes to elicit from $\mathcal{B}$ in u and the *effect* it might have on $\mathcal{B}$ in u.

Specifically, let the Setting Game SG_u be one where $\mathcal{B}$ faces a decision about whether or not to date Bill and $\mathcal{A}$ has information that can help her. Assume it is common knowledge in u that $\mathcal{A}$ is a well-wisher of $\mathcal{B}$ and that $\mathcal{B}$ prefers not to date Bill if he were very bald but would be okay with it if he were somewhat bald. Owing to this common knowledge, $\mathcal{B}$ generally trusts $\mathcal{A}$'s judgment and, other things being equal, would prefer not to go against his word. Further, assume Bill can be either very bald or just bald as far as $\mathcal{B}$ knows but $\mathcal{A}$ knows it is the former that is true.[14] Note the quite realistic and complicated preferences and knowledge the two agents have in the situation. This somewhat involved scenario together with the decision problem SG_u induces the Content Selection Game CSG_u shown in Figure 11.1.

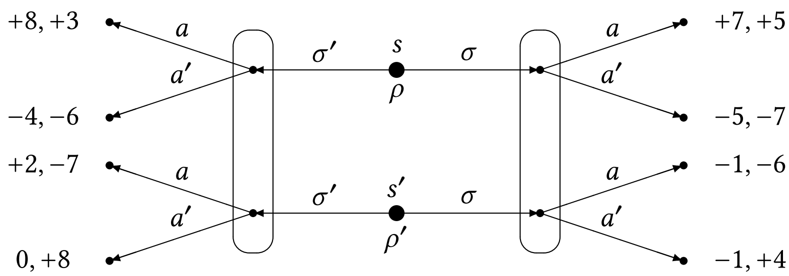


Figure 11.1: A Content Selection Game with vague contents

Now, there are two initial situations s and s' in the center of the diagram instead of just one as we had in Figures 7.1 and 8.9. The first of these is factual and is a part of u and contains the fact that Bill is very bald. The second situation is also a part of u and includes the counterfactual possibility that Bill is just bald. $\mathcal{A}$ knows s is factual but $\mathcal{B}$ does not and so both agents have to consider both initial situations. In the absence of any further information, the prior probabilities ρ, ρ' can both be taken to be 0.5.[15] In both situations, $\mathcal{A}$ can convey either σ or σ'

[14] It is the *property* of being very bald that matters here, not the concept, and, as I said earlier, the property does allow objective truth.

[15] There is a certain subtlety involved here because $\mathcal{A}$ knows that s is factual and therefore that $\rho = 1$ for him. One could say either that the game is played prior to his knowing this (the usual assumption in standard game theory) or, what I prefer, that they have common knowledge of $\mathcal{B}$'s belief of equiprobability. That is, $\mathcal{A}$ adopts $\mathcal{B}$'s ignorance for the sake of the situation.

assuming that when Bill is bald, that is, in s', $\mathcal{A}$ could be said to *exaggerate* Bill's baldness if he were to choose σ. These choices lead to two ovals or information sets (as we had in the earlier partial information games) and then to $\mathcal{B}$'s possible responses a = *don't date Bill* and a' = *date Bill*.

The preferences described informally above together with the slightly greater cost of σ translate into the numerical payoffs shown in Figure 11.1. It is a little tedious to go through how each pair of numbers is arrived at so I will explain just two pairs. Consider the path s, σ, a on the upper right. Here, in the unnamed terminal situation that results, $\mathcal{A}$ has accurately conveyed that Bill is very bald as that is the fact in s and $\mathcal{B}$ has chosen not to date Bill. Both parties are happy with the outcome given their preferences and so both get positive payoffs of +7 and +5, respectively. Now consider the path s, σ', a on the upper left. Here, $\mathcal{A}$ has chosen the slightly vaguer and cheaper content σ' but $\mathcal{B}$'s decision remains the same so $\mathcal{A}$ gets a slightly higher payoff of +8. On the other hand, $\mathcal{B}$ has chosen not to date Bill despite receiving the content that Bill is just bald (rather than very bald) so, while she is happy with the outcome since s contains the fact that Bill is very bald, she is also a little conflicted because she has gone against $\mathcal{A}$'s judgment that Bill is just bald (and that, in effect, it is okay to date Bill). So she gets a positive payoff of +3 which is a little lower than she got in the first case. All the other payoff numbers can be analyzed in a similar way. Perhaps the key thing to note is that there is a combination of cooperation and conflict in this Content Selection Game.[16]

I have gone into some detail to show how the preferences and payoffs are determined by psychological and social factors in the utterance situation of an intricate and nuanced sort. I repeat what I said in Section 8.6: a great diversity of human and social phenomena get resolved into relatively few numerical slots, the prior probabilities and the various payoffs. In this instance, however, the game tree has grown more complex with many more branches and so there are more numbers than before to be set.

This completes the description of the relevant CSG_u in Γ_u. Its more complex form has nothing to do with vagueness per se. It all depends on u and I happened to include a choice between a relatively more precise statement and a less precise one as this kind of occurrence is not uncommon. The (Nash) solution to this CSG_u is that $\mathcal{A}$ should choose to convey σ = *Bill is very bald* in s, which is the situation that matters since it is factual. As yet, $\mathcal{B}$ does not even know CSG_u exists as nothing has been uttered. But when $\mathcal{A}$ does utter something like φ =

[16]It is possible to complicate the model further by noting that the property of being bald or very bald is an average of the individual properties of the members of the community and so is epistemically inaccessible. This would make the payoffs depend on the *beliefs* of the two agents and would require the psychological games of Geanakoplos et al. (1989).

BILL IS VERY BALD as part of $GG_u(\sigma)$ and $\mathcal{B}$ interprets the utterance via $UG_u(\varphi)$ as conveying σ, she can then construct and play CSG_u. When she does this, her part of the (Nash) solution will be to respond with *a = don't date Bill*. $\mathcal{A}$ will thus get a payoff of +7 and $\mathcal{B}$ of +5 by playing CSG_u.

I have mentioned both full rationality and partial rationality. In the partial information games that constitute the (locutionary and illocutionary) global game, full rationality is likely to operate as those games essentially involve comparing probabilities based on Equation 6 from Section 8.1.5. Even here, there are further simplifications that I will describe in the next chapter that make it easier to be fully rational. But when more complex calculations of the kind shown in Figure 11.1 are involved and when they have to be done in real time, resource bounds are likely to kick in, making full rationality harder to employ. In such circumstances, there will be a variety of measures the agent may adopt: using a fragmentary or approximate form of CSG_u; using a different model of payoffs and solutions as dictated, for example, by a theory such as Kahneman & Tversky's (1979) prospect theory; using heuristics to solve the game; or a combination of these. As the theory of partial rationality is still in its infancy, it is best to simply display the full model and outline the solution without trying to anticipate how it would be implemented. On the addressee's side, often she may see only the fragment of the game issuing from the actual content conveyed but there will be times when she might want to think about what alternative contents the speaker could have conveyed but chose not to.

How the issue of common knowledge of CSG_u should be dealt with is unclear. One could baldly assume it despite all the partial information the agents are likely to have about the game. As handling the alternative possibilities is messy, I will just leave the matter unresolved with the warning that realism about the details is going to be quite thorny. This is where games with unaware players of the kind studied by Halpern & Rego (2006; 2007) are likely to be of use.

There is no need to go through $GG_u(\sigma)$ and $UG_u(\varphi)$ in detail. Both involve four constraints: Phonetic, Syntactic, Semantic, and Flow. I will confine myself to the Semantic Constraint that says that every word in an utterance is transformed by a conventional map into its conventional meaning(s) and then further transformed by a referential map into its referential meaning(s).

So far, the conventional meanings of a word have been its conventionally associated and more or less *shared* properties. But, as mentioned in Chapter 3, sharing a language partly involves sharing its conventional meanings but now, in light of our discussion of vagueness, it becomes clear that there is at best an overlap among, rather than common knowledge of, the concepts (or the corresponding individual, not average, properties) associated with a word.

It is convenient to say that each conventional meaning for an agent is the word's conventionally associated concept (rather than individual property), and each referential meaning is the corresponding individual or property into which the former is mapped relative to the utterance situation u.

$$\text{word} \longrightarrow \text{conventional meaning(s)} \xrightarrow{u} \text{referential meaning(s)}$$

Symbolically expressed for each agent separately, this becomes:

$$\omega \longrightarrow C^{\omega} \xrightarrow{u} P^{\omega}$$

assuming the concept C^{ω} is converted into the corresponding property P^{ω} in u. The u on top of the second arrow implies that it is an argument of the referential map together with the conventional meaning. This schema combines the internal and external significance of language as claimed earlier. It allows agents to compute referential meanings based on the conventional meanings in their heads and to convey abstract propositions that are not in their heads.

Consider now an utterance by $\mathcal{A}$ to $\mathcal{B}$ in u of the sentence φ = BILL IS VERY BALD. Then the two maps above will apply to each of the four words in the utterance. As our interest is in BALD, we get the picture:

$$\text{BALD} \longrightarrow C^{\text{BALD}} \xrightarrow{u} P^{\text{BALD}}$$

Each concept and property can be understood as marked by the relevant agent $\mathcal{A}$ or $\mathcal{B}$ implying that there are actually *two* such arrow diagrams for the word, one for each agent. The vague concept C^{BALD} can be more or less any of the many situated concepts that $\mathcal{A}$ (or $\mathcal{B}$) has used in the past or it can be some average of these. Further, in the current situation u, it gets transformed via a new u-relative concept into its corresponding vague property. That is, the concept C^{BALD} that is the conventional meaning *shifts* to a related concept C'^{BALD} relative to u and thence to the corresponding property. This shift is required to accommodate the different kinds of uses of BALD that may occur – for example, in the sentences SHE WON'T DATE BILL – HE'S BALD and BILL ISN'T BALD – HE NEEDS A HAIRCUT. Here, the penumbra of the same agent's concept shifts in accord with the use by accessing different exemplars or attributes or weights. The transformation of C^{BALD} to the shifted concept C'^{BALD} also depends on truth as discussed in Section 10.4. Working out the details is likely to be a bit involved but the broad contours of penumbral shift, the change of truth value of the same sentence in different situations, do not seem to raise any special problems once the general context-sensitivity of language is accounted for. This can be displayed as an extended Semantic Constraint as follows:

$$\text{BALD} \longrightarrow C^{\text{BALD}} \xrightarrow{u} C'^{\text{BALD}} \xrightarrow{u} P^{\text{BALD}}$$

The property that is the referential meaning can be taken to be either the *intersubjectively derived* average property or the subjective property and, in the former case, it can be assumed that each agent has just a partial and "vague" understanding of the content of the utterance. Also, with an actual utterance of φ, there would be more than two diagrams as the word BALD is *ambiguous* besides being vague. For example, it can also mean *plain* or *blunt* as in A BALD STATEMENT. So there will be two or more conventional meanings, each of which will be mapped into their respective referential meanings for each agent. But we can ignore this complexity here. I will also omit consideration of the use of VERY with BALD but it can be said in passing that it makes the threshold ϵ_u more stringent.

A third question to ask is what made it possible for $\mathcal{A}$ to choose σ and utter φ in the first place. In order to do so, he would have had to determine that Bill *is* (very) bald and to make that determination he would have had to resort to the earlier calculations. That is, he would have ascertained that $|P(C'^{\text{BALD}} \mid \text{Bill}; u) - P(\textit{not}\, C'^{\text{BALD}} \mid \text{Bill}; u)| \geq \epsilon_u$ and that $P(C'^{\text{BALD}} \mid \text{Bill}; u) > P(\textit{not}\, C'^{\text{BALD}} \mid \text{Bill}; u)$, applying Definition 11.1 to the shifted concept. Because the threshold is more likely to be fuzzy, and fuzzy numbers involve membership functions, the calculation will involve some situated rule based on interval arithmetic for deciding what degree of membership is sufficient for counting someone very bald. In other words, such calculations are an integral part of content selection and natural language generation. Since most words in language are vague, this shows that speakers have quite a bit to do and it is something of a psycholinguistic mystery how so much is accomplished so quickly. Perhaps the human brain's parallel processing just is very fast with such probabilistic comparisons. On $\mathcal{B}$'s side, she simply has to access her concept C^{BALD} and then her shifted concept C'^{BALD} and corresponding property P^{BALD}, although in circumstances where the truth of the proposition conveyed matters in determining it as described in Section 10.4, she would have to go through the same arithmetic with her own threshold.

The rest of the analysis for such an example is identical to what we have seen with nonvague language. Thus, once one has the right approach to vagueness, its apparent hurdles seem to melt away.

I have construed the exemplar and prototype approaches of cognitive psychology in certain ways to characterize vagueness, approach the sorites paradox in a new way, analyze essentially contested concepts, and describe certain relevant aspects of vague communication. The models appear to have some empirical support as well though I want to emphasize the *kind* of reasoning that is involved in addressing these problems. More realistic models of vagueness will doubtless become available but the underlying structure of explanation I have offered is not likely to change materially. For example, the explanation for the sorites based on the presence of multiple exemplars as opposed to a single exemplar is likely to survive further refinements of the underlying models of vagueness.

11.6 Communication and categorization

I want to now briefly point out that I have so far applied ideas of categorization to communication but it is possible to go in the reverse direction as well. There could be a two-way interaction between the study of category formation and language use. One could apply my model of communication to categorization and, in particular, compare the exemplar and prototype approaches in a domain where these competing ideas have not been compared before, namely their potential to explain the use of vague concepts in communication. This could ground psychological categorization in a broader context of linguistic interaction, which could further our understanding of its social nature. This integration of mathematical models of category formation and a mathematical theory of communication is a contribution to both fields separately as well as to the growing interdisciplinary effort to bridge pure psychological and language-related research.

Incidentally, this model also provides a basic scaffolding for formulating computational models of vague utterance generation and interpretation that can be applied to building dialogue and conversational agents, question-answering systems, and related areas in artificial intelligence.

I cannot address these matters here but do briefly discuss the generation of conventional word meanings in Part V.

12 Psycholinguistics and natural language processing

12.1 The connection with psycholinguistics

The foregoing analysis presents the broad framework for (locutionary) communication. Psycholinguists study the actual processes of utterance production and comprehension. So the Communication Game Γ_u has to be specialized and refined in various ways to fit with their experimental data. Fortunately, the basic approach of Equilibrium Semantics is largely compatible with the findings reported in, for example, the textbook by Fernández & Cairns (2011: Chapters 5–8) and elsewhere.

As confirmed by Swinney (1979); Onifer & Swinney (1981); and Kawamoto (1993), multiple senses of ambiguous lexical items are activated and disambiguated exactly as dictated by games of partial information. Rodd et al. (2002) found that homonyms, that is, words like BANK with unrelated conventional meanings take more time to process than polysemes, words like EYE or SCHOOL with related senses. Pylkkänen et al. (2006) and Frisson (2009) suggest that the conventional meanings of polysemous words are likely to be relatively abstract and impoverished underspecified cores relative to the full meanings they are given during processing. This implies some further alterations to our conception of conventional meaning that I discuss in Chapter 17. These findings are based on ingenious experiments rather than mathematical models and so this provides an opportunity to connect the latter from Equilibrium Semantics with the former. This kind of work cannot be pursued here but I say a little about one refinement to my model to illustrate the point.

As I said in Section 8.1, the meaning and parse vectors x and y are parts of the *content vector* $z = (x, y)$ with $2n$ components where n is the number of words in the sentence. This vector ranges over all the possible lexical meanings and parses of φ uttered in u. I repeat Equation 6 from Section 8.1.5 below for the reader's convenience.

$$z^\star = \operatorname*{argmax}_z P(z_k \mid z_{-k}; u) = \operatorname*{argmax}_z P(z; u), \quad k = 1, \dots, 2n$$

This can be broken down into two symmetric equations, one for meanings and one for parses, as follows:

(18) $$x^{\star} = \operatorname*{argmax}_{x,y} P(x_i \mid x_{-i}, y_i, y_{-i}; u) \quad i = 1, \dots, n$$

(19) $$y^{\star} = \operatorname*{argmax}_{x,y} P(y_i \mid x_i, x_{-i}, y_{-i}; u) \quad i = 1, \dots, n$$

In real-time computations, the meanings and parses of earlier words can affect those of later words but not vice versa unless backtracking is required owing to a garden-path sentence or similar problem. So the symmetric equations above can be modified to reflect this temporality.

(20) $$x^{\star} = \operatorname*{argmax}_{x,y} P(x_i \mid x_1, x_2, \dots, x_{i-1}, y_1, y_2, \dots, y_{i-1}, y_i; u) \quad i = 1, \dots, n$$

(21) $$y^{\star} = \operatorname*{argmax}_{x,y} P(y_i \mid x_1, x_2, \dots, x_{i-1}, x_i, y_1, y_2, \dots, y_{i-1}; u) \quad i = 1, \dots, n$$

Equations 20 and 21 capture some of the constraints on real-time sequential processing. These can be further refined to reflect different experimental results (e.g. the fact that the comprehension of the underspecified core sense of a polysemous word occurs instantaneously but the full meaning is realized only at the end of the sentence, if at all – Frisson 2009).

At a more general level, it is far from clear from the data whether the various games themselves are played or whether just equations like the ones above involving probability comparisons are solved directly. Secondly, I have developed the broadest construal of Generation Games as including a model of the addressee and the various partial information games. This is likely to require some modifications. The Content Selection Game appears to be completely absent in psycholinguistic studies because the experiments start with an already selected content to be conveyed. It is also unclear whether and how meanings and parses interact in actual processing especially when we take account of the different sites in the brain where such activity might occur. For example, it may be that the utterance situation plays a greater role in the determination of optimal meanings than in the determination of optimal parses. The same observation also applies to phonetic processing. In other words, while the abstract framework of Equilibrium Linguistics makes possible the interdependence of phonetics, syntax, and semantics in relation to an utterance situation – which may be useful in building artificial agents – the psycholinguistic evidence may suggest that these three types of content are not so closely coupled. If this turns out to be so, then for psycholinguistic purposes a slightly less integrated framework can be devised keeping its basic principles intact.

Such observations and others suggest the possibility that an entire suite of psycholinguistic and neurolinguistic experiments may be conducted in light of Equilibrium Linguistics and there is much scope for fruitful interactions between the two. This could lead to one kind of unification of the separate fields – psycholinguistics and semantics – of the science of language mentioned in Section 1.4 and at the end of Chapter 6. Current work in semantics is so remote from psycholinguistic practices that there is no easy way to bring them together. It is to the credit of Equilibrium Linguistics that its results are both more general and more precise than what exists in semantics and are therefore more directly testable.

12.2 The connection with natural language processing

Not only does my framework connect with psycholinguistics and neurolinguistics, it also offers multiple bridges to computational linguistics and natural language processing. Artificial intelligence involves different kinds of constraints as compared with real-time human processing and is generally divided into logical, statistical, and hybrid approaches, although most contemporary methods involve some use of probabilities.

Hobbs (2004) provides an overview of logical techniques involving abduction in natural language understanding. Recall from footnote 4 in Section 5.1 that in abduction we conclude from an observable Q and a general principle $P \implies Q$ that P must be the underlying reason that Q is true. We assume P because it explains Q. In practice, there may be many such possible P's, some contradictory with others, and therefore any method of abduction must include a method for evaluating and choosing among alternative explanations. This provides an inference to the *best* explanation. Indeed, as I show now, the choice-based approach of Equilibrium Semantics is just a species of abduction in which the choice structure is far more sophisticated and detailed than that provided by Hobbs or in the literature.

Consider the game shown earlier in Figures 7.2 and 8.1 and repeated for convenience in Figure 12.1. Recall that φ_1 = BILL, σ_1 = *Bill Smith*, and σ_1' = *Bill Jones*. The speaker is conveying σ_1 in s_1 and σ_1' in $s_{1'}$. Both facts imply that $\mathcal{A}$ utters φ_1. So if we identify P with $\mathcal{A}$ *is conveying* σ_1 and P' with $\mathcal{A}$ *is conveying* σ_1' and, further, if we identify Q with $\mathcal{A}$ *utters* φ_1, we can write $P \implies Q$ and $P' \implies Q$. The

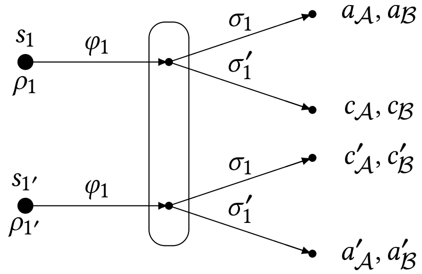


Figure 12.1: Semantic lexical game g_1

pure abductive task is then to infer which of P and P' best explains the observable Q. But the game g_1 in Figure 12.1 resolves precisely this question *in conjunction with* the other partial information games associated with the whole utterance. As we have seen, this complex strategic inference results in the Fundamental Equation. Thus, strategic inference is a more refined form of abduction.

Those pursuing purely logical approaches do not seem to realize that their setups miss a great deal of structure that is present in the problem. This is part of the reason they tend to be rather brittle and do not scale well in real natural language understanding applications. This lack cannot be rectified simply by adding in a few costs here and there as Hobbs attempts to do. A strategic inference, on the other hand, is a very tightly knit structure containing all kinds of relations among the various bits of information in the problem. That is why we get something as elegant and powerful as Equation 6.

Incidentally, these observations are orthogonal to the issue of how so-called real world knowledge is best represented as I also said in Section 8.1. Both pure abduction and more refined strategic inference require it as any full-fledged approach would. Some of it may be present in the encyclopedic entries for lexical items and some of it in some kind of knowledge base. It would enter the computation of meaning both through conventional (lexical) meanings and through the prior probabilities in the locutionary global game. But, importantly, the need for explicit representation of some of it would be circumvented by using corpus-based probabilities for different alternatives. But it cannot be avoided altogether. Consider the following example:

(22) The weight went through the wall because it was made of iron.

Here, some world knowledge has to be used to set the conditional probabilities for the two possible referents of the pronoun. The probability for *weight* would presumably be higher than for *wall*. We also used such world knowledge in the earlier examples, including the simple sentence BILL RAN (e.g. knowing that politicians run in elections). Humans have ready access to it so it can be assumed to be available. For the sentence above, it is enough to say that most human interlocutors would have the background knowledge that iron would go through many ordinary materials and so the possible referent *weight* has a higher probability. This is really all that is required for a semantic theory for humans in my opinion. It is not a gap in the theory to say we cannot assign higher or lower probabilities without such background assumptions. Artificial agents, on the other hand, do not have such knowledge so it has to be made explicit. This is best done in the context of actually building such agents. In any case, such orthogonal matters cannot be addressed here but suffice it to say that Equilibrium Linguistics offers a hybrid approach to the problem.

Statistical natural language processing has exploded in the last two decades and the textbooks by Manning & Schütze (1999) and Jurafsky & Martin (2000; 2009) have been quickly outpaced by ongoing work.

Perhaps the key connection between Equilibrium Linguistics and contemporary approaches in statistical NLP is that they both see the resolution of ambiguity broadly construed as the common feature of many problems that arise in communication and other tasks. Jurafsky & Martin (2009: Section 1.2) make this observation up front and it is something I have emphasized in all my work. Ambiguity is not just lexical or structural or phonetic but occurs wherever there are multiple possible contents that can be phonetic, syntactic, or semantic, the last category of which includes all of pragmatics and discourse.

In my view, and as Jurafsky & Martin (2009: 592) also say, this key problem remains unsolved despite many advances in computational lexical semantics and in parsing. Many models such as Naive Bayes, Hidden Markov Models, Maximum Entropy Models, Conditional Random Fields, and others have been developed to address a variety of tasks.[1]

The results of Section 8.1 show that Equilibrium Linguistics can potentially offer a more satisfying theoretical foundation for such statistical models by grounding them in rational human agency in a way that is philosophically sound and empirically adequate and can also offer more general computational techniques. All the models listed in the previous paragraph make the basic assumption that communication is a stochastic process with no further constraints. The Communication Games developed in this book capture the further constraint that

[1]See Rabiner (1989); Sutton & McCallum (2006); Manning & Schütze (1999: Chapters 7, 9); and Jurafsky & Martin (2009: Chapters 5, 6, 20).

communication is more or less *rational*. This brings in more cutting power and enables sharper conclusions to be drawn as Theorem 8.5 and the more general Theorem 8.6 show.

For example, it is easy to see that Equation 6 is more general than the Naive Bayes equation for word sense disambiguation:

$$\begin{aligned} s' &= \operatorname*{argmax}_{s_k} P(s_k \mid c) \\ &= \operatorname*{argmax}_{s_k} [\log P(s_k) + \sum_{v_j \in c} \log P(v_j \mid s_k)] \end{aligned}$$

where the s_k are the word senses, the v_j are the words in the context c, and the P represents probability.[2] The key generalization is that the conditioning (and conditioned) variables in my equation are the "senses" and parses themselves, as opposed to words, because of the circular and fixed point nature of my conception.

While Hidden Markov Models are also not sufficiently general as they rely essentially on Bayesian inference, Maximum Entropy Markov Models and Conditional Random Fields *are* sufficiently general and do allow in principle for the kind of fixed point inference that occurs in Equation 6. All these models involve manipulations of conditional probabilities and the latter two make it possible to condition on a wide range of features including the ones required by Equation 6. But because these models are not based on rationality, they cannot derive Equation 6 itself and cannot conclude that the probability distributions involved in communication are *compatible* distributions. In practice, moreover, semantic and syntactic (and phonetic) contents are seldom thrown in together as they are in Equilibrium Linguistics so even though the latter two models *could* include the relevant conditioning variables, they almost never do.

To illustrate how bringing in rationality can generalize such purely probabilistic considerations, consider the problem of classifying a piece of text as spam or not spam. This example also shows, incidentally, how the tools of Equilibrium Linguistics can be extended to related tasks such as text classification as mentioned earlier in Section 7.3. A rudimentary version of the spam identification task can be modeled by the simple game shown in Figure 12.2. Being spam or not spam is just a more abstract kind of content of an utterance.

Clearly, the Nash solution involves comparing expected values of payoffs rather than pure probabilities. By adjusting the $\mathcal{B}$ payoffs we can get different, more fine-grained results. For example, ordinarily we would want false positives

[2] See Manning & Schütze (1999: Section 7.2.1) for more details.

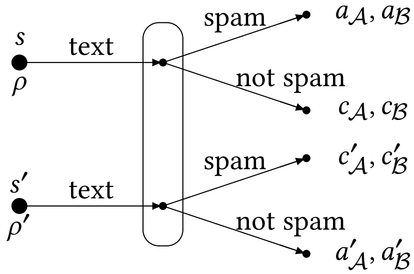


Figure 12.2: Text classification game g_{tc}

(i.e. classifying some text as spam when it is not spam) to be penalized more heavily. If it is assumed that the sender in s is spamming and in s' is not spamming then such a heavier penalty can be realized by making $c'_{\mathcal{B}}$ sufficiently low, that is, by making the cost of interpreting a text that is not spam as spam high. If, instead, we were doing sentiment analysis, a false negative, that is, a negative sentiment in s interpreted as positive may be penalized more heavily by making $c_{\mathcal{B}}$ sufficiently low. And so on. Needless to say, this only scratches the surface of the text classification problem but it shows how Equilibrium Linguistics offers additional degrees of freedom. This flexibility carries over to the much harder task of inferring the content of communication as well, as made clear by Figure 8.6 and Theorem 8.6, because it is also just a classification problem. It can also help in natural language generation and in building conversational agents and in other tasks such as machine translation (see Chapter 21) and information extraction.

I hope I have convinced the reader that Equilibrium Linguistics has something to offer to statistical NLP. Its main contribution is its comprehensive definition of the problem and its better grounded and more general solution. I believe it is the closest any theory in semantics has come to computational accounts.

As one would expect, the illumination goes in the other direction from statistical NLP to Equilibrium Linguistics as well. Recall from Section 7.3 where the Flow Constraint and partial information games were first discussed that the probabilities ρ were taken to be the prior conditional probabilities that $\mathcal{A}$ was conveying a certain content *given* that he was conveying the other meanings and parse trees of the rest of the sentence in the utterance situation u. This assignment was based on the empirical observation that the meanings and parses of words are interdependent in a rational utterance. As statistical NLP shows, such conditioning variables can be thought of very generally as a wide variety of *features*. For example, if the word BROWN is capitalized in a written sentence, it helps the reader to see that it is likely to be a name rather than an adjective. So features like capitalization can further aid disambiguation. In human understanding, too, such features can supplement the primary features of semantic, syntactic, and phonetic contents. All that is required is to augment the conditioning variables in the prior probabilities by such extended features and then carry through the game-theoretic analysis as before.

This shows that if the approaches of Equilibrium Linguistics and statistical NLP are combined then both can be enriched, the former by adding more features to the list of conditioning variables, and the latter by bringing in both rationality as well as a more satisfying and foundational basis for the results.

I believe Part III gives a more or less complete solution to the problem of locutionary communication that is, as advertised in Section 1.4, philosophically sound, mathematically solid, computationally tractable, and empirically adequate. The Setting Game, Content Selection Game, Generation Game, and Interpretation Game that constitute a (locutionary) Communication Game are developed in detail. So are the four Constraints: Phonetic, Syntactic, Semantic, and Flow. The solution I offer shows, moreover, that partial information games are universal in the algebraic sense discussed in Section 10.1. Frege's two historic principles of compositionality and context are both reconciled and transcended. I develop a theory of vagueness and address the sorites, relate vagueness to essentially contested concepts, and model vague communication. Finally, as promised in Section 1.4, I connect arguments in philosophy, linguistics, psychology, and computer science, unifying them through the framework of Equilibrium Linguistics.

In the next Part, I consider illocutionary communication and explore the wider varieties of meaning.

Part IV

Illocutionary meaning and beyond

13 Relevance

Much of this Part falls under the topic of the global game $G_u(\varphi) = LG_u(\varphi) \cup IG_u(\varphi)$. I deferred considering this aspect of meaning as it requires some new insights that are best studied separately. Illocutionary meaning includes free enrichment, implicature, modulation, illocutionary force, and perhaps other things, all very interesting phenomena that allow speakers to convey meanings beyond what is conveyed *directly* by the words in an utterance, that is, beyond locutionary meaning. The same sorts of Constraints operate in this realm. There is no need for a new Phonetic or Syntactic Constraint; they are the same as with locutionary meaning as it is the same sentence φ that gives rise to illocutionary meaning. The Semantic Constraint, which generates the possible illocutionary meanings of an utterance, is no longer based on the conventional meaning of the words in the sentence and so does not split into two sub-Constraints, Conventional and Referential, as happened with locutionary meaning. It is based instead on two different sub-Constraints of Relevance and Distance that are needed only for illocutionary meaning, not for locutionary meaning. They perform the same function as the earlier sub-Constraints of generating the *possible* (illocutionary) meanings. The Flow Constraint then disambiguates these possibilities via partial information games as before. Thus, there are many similarities between locutionary and illocutionary meaning, the fundamental one being that the possibilities are first generated and then disambiguated. This is one reason to take all of meaning as belonging to the single realm of *semantics* as argued in Section 5.4.

After describing how the illocutionary meaning of an utterance is computed, I will look at further kinds of meaning that lie beyond illocutionary meaning. At the end of the Part, we will classify all these meanings and discuss the issue of what literal meaning is. But, as was pointed out in Section 5.4, all these distinctions are somewhat indeterminate, even the so-called primary one between literal meaning and implicature.

I will start by discussing the two sub-Constraints of Relevance and Distance in this chapter and the next.

I had defined relevance as the *value* of information, an idea from choice theory, in Parikh (1992; 2006b; 2010). In the later two references, I had extended the standard decision-theoretic notion to game theory, though this extension remained incomplete there for technical reasons. In this chapter, I complete it.

I first examine the general notion more closely, starting with some of the limitations of the way the Relevance Theorists have defined it, and then broadening the discussion to my own idea.

13.1 The Relevance-Theoretic concept of relevance

Relevance Theory[1] has been around now for over thirty years and has contributed many rich insights to the fields of semantics, pragmatics, and the psychology of language comprehension. Yet, the formulation of its foundational concept remains weak. This is partly because the idea of relevance is extraordinarily difficult to pin down. But it is also partly because the Relevance-Theoretic approach is flawed in certain key respects. This weakness threatens the integrity and operation of the theory because both of its primary pillars – how gaps between sentence meaning[2] and speaker meaning are to be filled and how, more broadly, utterance comprehension occurs – rest upon computations based on relevance. For this reason, a close examination of its eponymous concept is warranted. Here is their own description.

> When is an input relevant? Intuitively, an input (a sight, a sound, an utterance, a memory) is relevant to an individual when it connects with background information he has available to yield conclusions that matter to him: say, by answering a question he had in mind, improving his knowledge on a certain topic, settling a doubt, confirming a suspicion, or correcting a mistaken impression. According to relevance theory, an input is relevant to an individual when its processing in a context of available assumptions yields a POSITIVE COGNITIVE EFFECT. A positive cognitive effect is a worthwhile difference to the individual's representation of the world: a true conclusion,

[1] There are many accounts of the theory in the literature. I have referred primarily to Sperber & Wilson (1986; 2002) and Wilson & Sperber (1986; 2004).

[2] Relevance theorists take sentence meaning to be a kind of logical form that is an incomplete proposition as opposed to the mainstream Gricean view that it more or less coincides with literal meaning modulo certain linguistically mandated items that need to be filled in contextually. My own view, described in Parikh (2010) and in this book, is more radical in that I believe there is nothing like sentence meaning involved in interpretation, just conventionally given word meanings and utterance meaning based on a generalized compositionality, as discussed in Part III. That is, the difference between the Relevance-Theoretic view and the Gricean view is one of degree, not one of kind – as mine is.

for example. False conclusions are not worth having; they are cognitive effects, but not positive ones.

The most important type of cognitive effect is a CONTEXTUAL IMPLICATION, a conclusion deducible from the input and the context together, but from neither input nor context alone. For example, on seeing my train arriving, I might look at my watch, access my knowledge of the train timetable, and derive the contextual implication that my train is late (which may itself achieve relevance by combining with further contextual assumptions to yield further implications). Other types of cognitive effect include the strengthening, revision or abandonment of available assumptions. For example, the sight of my train arriving late might confirm my impression that the service is deteriorating, or make me alter my plans to do some shopping on the way to work. According to relevance theory, an input is RELEVANT to an individual when, and only when, its processing yields such positive cognitive effects.

Relevance is not just an all-or-none matter but a matter of degree. There are potentially relevant inputs all around us, but we cannot attend to them all. What makes an input worth picking out from the mass of competing stimuli is not just that it is relevant, but that it is MORE relevant than any alternative input available to us at that time. Intuitively, other things being equal, the more worthwhile conclusions achieved by processing an input, the more relevant it will be. According to relevance theory, other things being equal, the greater the positive cognitive effects achieved by processing an input, the greater its relevance will be. Thus, the sight of my train arriving one minute late may make little worthwhile difference to my representation of the world, while the sight of it arriving half an hour late may lead to a radical reorganization of my day, and the relevance of the two inputs will vary accordingly.

What makes an input worth attending to is not just the cognitive effects it achieves. In different circumstances, the same stimulus may be more or less salient, the same contextual assumptions more or less *accessible* [my emphasis], and the same cognitive effects easier or harder to derive. Intuitively, the greater the effort of perception, memory and inference required, the less rewarding the input will be to process, and hence the less deserving of attention. According to relevance theory, other things being equal, the greater the PROCESSING EFFORT required, the less relevant the input will be. Thus, RELEVANCE may be assessed in terms of cognitive effects and processing effort:

(1) **Relevance of an input to an individual**:

a. Other things being equal, the greater the positive cognitive effects achieved by processing an input, the greater the relevance of the input to the individual at that time.

b. Other things being equal, the greater the processing effort expended, the lower the relevance of the input to the individual at that time.

(Wilson & Sperber 2004: 608–609)

The key word in the foregoing is *worthwhile* difference to the individual's representation of the world. Sperber and Wilson develop this idea through the "maximization" of positive cognitive effects relative to processing effort. Cognitive effects are taken to be *contextual implications* or the strengthening, revision, or abandonment of assumptions. This last equation – of cognitive effects with contextual implications and other kinds of inference – is primarily where the Relevance-Theoretic idea falls short in a number of ways.

13.1.1 The first difficulty

To start with, it seems that there is no practical way for the individual to assess which of many cognitive effects is greater. In most of their examples, the magnitude of these effects is taken to be the *number* of contextual implications or inferences each input generates. This raises the obvious difficulty that if p and q are contextual implications of an input, are these to be taken as two propositions $\{p, q\}$ or as just the single conjunction $\{p \wedge q\}$ or as the triple $\{p,\ q,\ p \wedge q\}$, and so on? Is p itself just one proposition or more than one?[3] There is no unique way to individuate implications and each way is equally arbitrary. While the authors appeared to take this exclusively numerical route in Sperber & Wilson (1986), they try to avoid it in their more recent Wilson & Sperber (2004: 609–610) by expanding the scope of cognitive effects to other less quantitative aspects: "In the first place, only some aspects of effect and effort (e.g. processing time, number of contextual implications) are likely to be be measurable in absolute numerical terms, while others (e.g. strength of implications, level of attention) are not."

Unfortunately, their comparative approach is applicable only to the very simplest of comparisons involving entailment of one input by another. If two inputs were picked randomly from a large pool of inputs, the likelihood of their comparability would be zero, making the determination of optimal relevance a rarity.

[3]For example, for some input an implication may be that the ball is round. Now, is that one implication or two or more? It could be said that the implications are that the ball has a shape and that this shape is round, and so on.

13.1.2 The second difficulty

The problem of comparing the relevance of different inputs is in practice sidestepped by bringing in the idea of accessibility mentioned in the first quote. Based on the "definition" of relevance in (1) from the quote above, the following comprehension procedure is described in Wilson & Sperber (2004: 613):

> Relevance-theoretic comprehension procedure:
>
> a. Follow a path of least effort in computing cognitive effects: Test interpretive hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.
>
> b. Stop when your expectations of relevance are satisfied (or abandoned).

But it is not clear what is accessible and there is a grave danger of circularity because what is regarded as optimally relevant is usually what is most accessible *and* vice versa. There is no explanation for why things always seem to turn out this way. Why isn't a less accessible interpretation ever optimally relevant?

Besides, nothing of consequence is said about an interlocutor's expectations of relevance. How much relevance is enough? This question is never addressed though it plays a crucial role in stopping the process before a comparison of the relevance of different interpretations can occur.

Their examples (e.g. Wilson & Sperber 2004, Sperber & Wilson 2002) are framed in a way that encourages the reader to conflate their own *intuitive* concept of relevance with the Relevance-Theoretic one. They are lulled into thinking that there is a definite procedure being followed when in fact a highly incomplete and vague series of steps is described with crucial gaps filled in with one's own intuitions.

The problem is that no comparison between the relevance of competing inputs is ever required because the comprehension procedure stops operating after the first and most accessible input's relevance invariably meets the addressee's expectations of relevance.

Basically, we are simply *told* that certain interpretations are more accessible than others based on commonsense reasoning that *we* employ, not the protagonists in their examples. By pointing out that the correct interpretation is the one that is the most accessible, no real work is done by the theory because there is no *theory* of accessibility. What starts off as an account of a relevance-theoretic procedure for comprehension ends up as an accessibility-theoretic procedure but there is, in fact, no such procedure made available by the theory because there is no such theory to begin with.

Unfortunately, even if an empirical theory of accessibility became available, more serious difficulties with following the Relevance-Theoretic comprehension procedure and evaluating relevance would still remain.

13.1.3 The third difficulty

The problem with the Relevance-Theoretic notion is graver still. Even its qualitative notions appear, on closer inspection, to be on the wrong track *conceptually*.

As I said earlier, the key word in the first quote above was *worthwhile* differences to an individual's representation of the world. In those paragraphs, relevance is related to positive cognitive effects and positive cognitive effects are related to *worthwhile* differences in an individual's representation of the world. These differences are cashed out initially in epistemic terms: "a true conclusion, for example." It is asserted that false conclusions are not worth having; they are not positive cognitive effects. Later, such effects are elaborated in terms of the number and strength of contextual implications, as we have seen.

Both of these dimensions of positive cognitive effects are misplaced. It is just not true that false conclusions do not have positive cognitive effects. Indeed, most if not all human beings entertain their share of false beliefs; it is precisely these false beliefs that enable them to get by. The truth may be too unbearable and so people invent all sorts of polite behaviors and euphemisms and other circumlocutory modes of being that border on or are falsehoods. Not only that, even entire systems of belief – like various religions and ideologies – may help alleviate human suffering to a considerable degree while being completely false. People typically choose hope over despair even when such a choice is not warranted.[4] Often, believing something false will get a person into trouble but, depending on how marginal the false belief is to the practical realities of life, the person may in many cases be better off holding onto the false belief.[5] This psychological phenomenon is well known enough to not require further argument.

More importantly, though, it is the identification of cognitive effects with the number and strength of contextual implications, the heart of the Relevance-Theoretic view, that is the real problem. Implicit in this identification is the equation of effects with the *amount* of information extractable from an input rather than with the *value* of that information to the individual. That is, a *worthwhile* difference or effect is one that is *valuable*, not one that involves large amounts

[4] As Woody Allen has said: "More than any time in history mankind faces a crossroads. One path leads to despair and utter hopelessness, the other to total extinction. Let us pray that we have the wisdom to choose correctly."

[5] "She loves me, she loves me not … ."

of information. Once one has this clear insight – first expressed in Parikh (1992) and later in Parikh (2006b; 2010) – then the *conceptual* flaw in the Relevance-Theoretic notion becomes obvious. Simply put, more information is often but not always better for the individual than less information because that additional information may simply not *matter*. Would you prefer a gift of the Encyclopedia Brittanica or the winning lottery number?

It is to address such issues that Grice's (1975; 1989c) maxim of quantity referred to the *purposes* of the exchange. Suppose two bank robbers had robbed a bank some years ago and one of them had hidden the loot without the knowledge of the other at a river bank they frequented. Now, the first robber is dying and the second asks him where the loot is, and he says:

(23) It is at the bank.

It appears prima facie that the financial bank should be more accessible and should yield sufficient relevance to meet the second robber's expectations because, as Willie Sutton reminded us, "that is where the money is." But it is equally arguable that the river bank is the right choice because both the robbers know the location and it is possible to stash loot in such places. Indeed, with some more reasoning it is clear that the dying robber meant the river bank because he would otherwise have provided some more information about how to retrieve the money from the financial bank (e.g. what the account number was, etc.). On the other hand, the financial bank reading may have many more contextual implications (e.g. how long the loot had been accumulating interest and how much the balance might now be, whether it was in a savings account or certificate of deposit, and so on). But, intuitively, all these implications are *irrelevant* to the matter at hand.

It might be said that the context could be construed differently but such maneuvers begin to beg the question. It is as if the context is being tailored to fit our pre-theoretic notion of relevance which is then shown to yield an outcome based on the Relevance-Theoretic notion resulting in an identity of the two.

It is the purposes of the exchange that are missing from the Relevance-Theoretic account. By confining worthwhile differences to the individual's representation of the world, Relevance Theory treats relevance in a purely *epistemic* way. There is no room for purposes or goals or preferences in this view. A better intuition is that an input is relevant *if it warrants a change in the individual's intended actions* because this idea is intimately bound up in a concrete way with the purposes and goals of the individual. Here, the term "action" includes both external behavior as well as internal, that is, "mental" actions like assessing a statement as true and forming a corresponding belief. As I point out in Parikh (1992; 2006b; 2010), this is just the choice-theoretic notion of the *value of information*. It readily admits of degrees of relevance and allows one to compare the relevance of two inputs. And it more accurately captures the idea of *worthwhile* differences because worth and value are one and the same thing.

13.1.4 The fourth difficulty

Lastly, the point made in the last paragraph above is worth making more broadly. Relevance Theory, influenced by Grice, Chomsky, and Fodor as it is, faces an insurmountable meta-difficulty with the very *way* in which it approaches semantics – what it calls pragmatics – involving a narrow focus on the *cognitive* aspects of agency and ignoring its practical aspects. This leads to prematurely conflating the related domains of semantics and psycholinguistics by directly importing psychological notions like *accessibility* without any theoretical underpinning via speculations about specialized pragmatic modules and the like.[6] Semantics and psycholinguistics ought to be studied at somewhat different levels of abstraction, as the discussion in Section 12.1 makes clear.

For example, memory certainly plays a key role in utterance comprehension but a semantic theory of communication would not typically attempt to incorporate its structure and would abstract from this constraint. In cases of disambiguation, it is true that we are not aware of all the senses of all the words we encounter in a sentence; our memory plays a key role in enabling the mechanism of *priming*, for instance, which leads to the suppression of some alternative senses of words in a sentence. But the semantic task is to explain how disambiguation occurs. If such a theory is right, its main ideas will be transferable to the psychological level. A communicative theory of priming will abstract from the details of memory and use just its broad contours.[7] By seeing linguistics as a branch of cognitive psychology – as Chomsky first proposed – Relevance Theory falters methodologically because the proper domain of semantics is the domain of actions and interactions, not just the cognitive processes of the mind. And this requires a more abstract level of inquiry.

[6] See Sperber & Wilson (2002).

[7] While Equilibrium Semantics as described in Parikh (2010) and in this book does not consider the phenomenon of priming in any detail, its framework is set up in a way that readily allows explicit disambiguation to occur when required and allows priming to eliminate alternative possibilities directly when required. It is then up to a psychological theory to specify when different mechanisms operate based on how memory functions. The precise details of how priming works will have to be described at the level of memory. This is analogous to the way in which, say, the levels of physics and chemistry are separated.

Such a level is, indeed, provided by decision and game theory, as I have argued in my publications. This abstract level does *not* mean that the framework is not intended to be an empirical one – it is. But it will still be a more general framework compatible with multiple psychological frameworks. It is interesting to observe this empiricism in its own domain. Earlier, rational choice theory was regarded by many as the right theory to employ when modeling action and interaction because it was sufficiently abstracted from the details of psychology and sociology and it appeared to give the right predictions. But, as Allais (1953), Simon (1955; 1956), and Kahneman & Tversky (1979); Kahneman et al. (1982) among many others in a large literature have argued, the utility-theoretic axioms of rationality do not in fact always give the right results and this has led to a major revolution in the theory of choice. So far, no *single* theoretical alternative has emerged as the clear successor although there are many competing approaches. The number of experiments has also mushroomed and the field of behavioral decision theory has begun to partially merge with the even more nascent and experimental field of neuroeconomics.[8] If a clear theoretical winner does emerge eventually, it will also be relatively abstract and devoid of the details of neuroscience and psychology and sociology, and just like utility theory in its broad form, but will include new features abstracted from such details.

Incidentally, Grice himself is also always clear that he is pursuing *philosophical* psychology, not psychology per se, even though he, too, remains confined to an epistemic view of agency as I have argued in Section 5.1. This collapsing of the distinction between semantics and psychology creates insuperable methodological problems for Relevance Theory. Overall, Relevance Theory is an advance on traditional theories because it does try to countenance the pervasiveness of context and offer a method to derive content. It unfortunately errs by remaining trapped within the same *epistemic* view of agency that also ensnared the tradition and by conflating semantics with psychology. And its key underlying concept of relevance is deeply flawed.

13.2 The value of information

I will not develop the idea of the value of information in any mathematical detail here; I have done this in the references cited at the start of this chapter, although, as I said, the extension of the concept to games as opposed to single-person decisions remained incomplete, something I rectify here.

[8] See Glimcher (2004).

The basic idea is that an agent may be faced with some choices and, based on his goals and preferences as well as on the uncertain consequences that may ensue if he were to pursue each particular choice, he identifies an optimal course of action. Now, he may also be in a position to get some information that could reduce his uncertainty about the events following his actions. If he acquires this information by asking an informed person a question, he may find that, with this additional information, he may be able to pursue either the same action or a different action with better payoffs. In either case, he can assess the value of the information he has received by computing the difference in payoffs. If his optimal choice does not change, we may say that the information was not relevant or that it was not relevant enough. But if his optimal action changes, then we can not only say that the information was relevant but we can also say *how* relevant it was depending on how much better off he is after receiving the information. In other words, information is relevant to an agent when it makes him better off. The virtue of decision theory is that it allows us to express this commonsense reasoning in a precise way. An excellent informal introduction can be found in the classic by Raiffa (1970).

This is very different from counting the number or strength of implications; the latter notion is logical and epistemic and is concerned solely with relations among propositions. Practical agency does not enter into this calculus at all and so it is no wonder that it fails to get a grip on how information might *matter* to an agent. Relevance Theory falls short because it, like much of linguistics and the philosophy of language, has only a conception of a reasoning agent but none of practical agency as discussed in Section 5.1.

Now, to make the ideas above a little more concrete, recall that $\mathcal{A}$ and $\mathcal{B}$ are involved at the outset in a Setting Game. This Setting Game may in fact be just a decision problem for either agent which I am calling a game to make the terminology more convenient. In this event, the value of information as dictated by decision theory would apply directly. A possible implicature, for example, would be tested for its value in the context of this decision problem. If its value is sufficiently positive relative to some threshold, it might become a candidate for further processing by the Flow Constraint. In other words, this criterion of Relevance or the value of information would be one sub-Constraint of the Semantic Constraint as mentioned at the start of this chapter.

The difficulty arises when there is no apparent isolated decision problem that either agent faces in the Setting Game. The choice problem is really a two-person game (or even a multi-person game). How then do we evaluate the value of information? The idea, which eluded me in Parikh (2006b; 2010), is straightforward. One can always extract a single-person decision problem from a game because a game is just an interactive decision problem between two or more persons. All that has to be done is to set up some choices of action for a player based on his available strategies in the game and uncertainty about what the other player will do based on her strategies and then use the first player's payoffs in the game to determine his payoffs in the decision problem so constructed. And this allows one to compute the value of information of a possible implicature in the usual manner of decision theory. So all that has to be done is to *reduce* a game to a component decision problem to compute the relevance of possible illocutionary meanings.

However, there are two formidable difficulties with the value-theoretic concept as well.

13.2.1 The first difficulty

Suppose John has tentatively decided to opt for a surgical procedure but decides to get some more information from his doctor. His doctor may say either of the following:

(24) a. 90% of the patients survive the procedure.
b. 10% of the patients do not survive the procedure.

It is well known that an agent's response to each of these two statements is often not the same.[9] With (24a) John is likely to want to go ahead with the procedure but with (24b) he might, like a number of other imperfectly rational agents, balk at the choice. This is so even though the information conveyed by both statements is equivalent. That is, if σ represents the information conveyed by (24a) and σ' the information conveyed by (24b), we have $\sigma \iff \sigma'$. This is clearly a situation where the form in which information is couched matters and so this phenomenon is called *framing*, a result of the intensionality of choice.

In other words, relevance cannot simply be taken as the value of information without assuming that agents are perfectly rational in the sense of utility theory. If they are not, the concept has to be defined in terms of *behavioral* choice theory rather than rational choice theory and this kind of behavioral relevance

[9]See Arrow (1982). The experiment Arrow discusses concerns clinicians' recommendations for surgery based on information about the survival probabilities. When told that the chance of survival was 90%, some 80% of surgeons who were asked recommended surgery. However, of those told that the same course of action was associated with a 10% chance of death, only 50% recommended surgery. Since the contents of both statements were equivalent, this implies that the preferences of some surgeons must have been intensional. I have changed the example a little.

may not yield the right results. If behavioral relevance can be defined in a general way that accounts for intensionality, that is, if it encompasses the value of *framed* information, then it may predict an agent's behavior correctly but give an incorrect view of relevance from a normative standpoint. This would create problems in explaining how communication occurs between imperfectly rational agents because each agent would have to consider the other agent's possible imperfections. As I argued in Section 5.2, all utterances involve framing, whether by design or not, and this poses problems for a theory of relevance.

One way to address the intensionality of choice is to use the prospect theory of Kahneman & Tversky (1979), though I will not pursue it here. Another way to begin considering the value of framed information is to first notice that situation theory as opposed to competing frameworks like possible worlds theory that game theorists and philosophers tend to favor *does* differentiate between equivalent infons σ and σ' because it is more fine-grained and can potentially avoid the problem of logical omniscience.[10] So it is possible to represent an agent's knowing σ without knowing σ'. Then an agent who is partially rational in a certain situation can be modeled as knowing some (framed) information such as σ but not knowing other equivalent information such as σ'. With these assumptions, the value of this partial information can be computed. This procedure would, in principle, give the behavioral relevance of some information. But it is not entirely clear how such a computation might be carried out in practice and when such partial rationality would have to be assumed.

The intensionality of choice rests on framing or the intensionality of language but, interestingly, the intensionality of language itself rests on what might be called the intensionality of metaphysics, the phenomenon that I briefly alluded to in Section 2.1 as that of the glass being half full or half empty. If the infon $\tau = \langle\!\langle \text{half full; glass} \rangle\!\rangle$ and $\tau' = \langle\!\langle \text{half empty; glass} \rangle\!\rangle$ then we have the further fact that $\tau \iff \tau'$. These are two equivalent *ways* of individuating the world. This kind of situational framing is basic to how we interact with the world and with each other and has profound effects on us even though, often, the different facts or infons we individuate are equivalent. Whether we see τ or τ' may result in the difference between hope and despair. Even surgeons, experts in their fields, behaved differently when the survival rates of surgery were differently but equivalently described.

[10] There is a large literature on this problem in epistemic logic. It occurs because with certain standard closure axioms, whenever an agent knows all of the formulas in a set and a formula follows logically from this set, the agent also knows this formula. See Hendricks & Symons (2015), Fagin et al. (2003: Chapter 9), and Stalnaker (1991), for example. If Dretske's (1981: Chapter 9) argument is right, then his account does not have this problem.

Why is it that human beings respond differently to τ and τ' even though $\tau \iff \tau'$? It is because they do not always realize that $\tau \iff \tau'$ and, in my view, this lies at the heart of the problem of rationality. This is why an intensional ontological framework like situation theory that distinguishes between τ and τ' is required. Only the underlying *reality* is the same, almost everything else – situations, infons (and facts), properties and relations, and perhaps other objects – are intensional. It is not just that human beings *perceive* the same fact in different ways; the facts themselves – of the glass being half full and half empty – are distinct but equivalent. (I believe this situation is different from Wittgenstein's (1953/1968) duck-rabbit, which *is*, as he says, a case of *seeing* different aspects.) What happens is that people interact with reality to individuate it and produce the ontologies we then see as real. This is very similar to the Buddhist idea of dependent origination as well as Quine's ontological relativity mentioned in Section 2.1. Where both fall short is in missing the further idea that only a subclass of ontologies is *optimal* (as argued in Parikh (2010: Section 7.6), where I call the codeterminative process *equilibrium metaphysics*) and so cannot be dismissed either as merely conventional as the Buddhist does or as merely relative as Quine does. This optimal subset has a certain *immovability* and this makes the "transcendental" aims of Buddhism all the more difficult to realize and the radical indeterminacy claims of Quine suspect.

I am not suggesting some kind of unbridled relativism (with optimality). Reality does offer resistance and does constrain the class of possible ontologies. The glass can only be half full or half empty or both; it cannot be one-third full in the situation under consideration. Frameworks like possible worlds theory that do not make such distinctions are ineffectual because they cannot hope to explain how human beings act. It is also revealing that possible worlds theory describes this indistinguishability as the problem of *logical omniscience* which indicates its logicist origins as it sees the problem as concerning just the *epistemic* dimension of agency rather than being much more widespread.

In any case, as I just said, such finer ontologies lie at the heart of the problem of rationality. Even though a half-full glass is the "same" as a half-empty glass, we may not realize this and so our responses to them may be different. Sometimes, even when their equivalence is pointed out, we may persist in giving them different weights and thereby respond differently even when we know they are the same. (Also, a glass being one or the other is usually just a metaphor for more complex situations, and then the posited equivalence can be somewhat more difficult to see.)

Of course, the intensionality of metaphysics is not confined to different ways of individuating *equivalent* infons. The problem is far more pervasive. In Section 8.6, $\mathcal{A}$'s goal is to have $\mathcal{B}$ accept his statement that Bill Smith ran in the local election and he has two fundamentally different and also nonequivalent ways of realizing this goal: either to issue a bald statement or to give $\mathcal{B}$ a reason to accept the information by adding that he read the information in a newspaper. While these two ways may not be informationally equivalent, they do possess what we may call *act-equivalence*, that is, they are meant to induce the same act by $\mathcal{B}$. This ties in how we individuate our choices in a situation very closely to human psychology, relationships, and perception, and the way we individuate our choices depends on how we can individuate the world. We will discuss an example of this act-equivalence in some detail later in Section 18.2.

While the behavioral revolution in decision-making of the last three decades covers a wider ground than the intensionality of metaphysics and language and choice, I believe the latter lies at its core. Because of this, a certain level of *creativity* or *idiosyncrasy* or *indeterminacy* may be ineradicable from human action and choice and, even though we may be able to explain certain actions partially after the fact, a truly predictive science that is also explanatory and not just data-driven, especially for large-scale phenomena like economy and society, may lie beyond our means. Intensionality also makes the problem of natural language generation much harder than the reciprocal problem of natural language understanding.

In Section 2.1, I mentioned the presence and general indeterminacy of context as a barrier to developing a complete science of communication; now the intensionality of metaphysics and language and choice and action can be seen as a further impediment.

13.2.2 The second difficulty

A different sort of difficulty arises when we recognize that communication does not always involve statements. What if a speaker says something that changes an addressee's goals or preferences? For example, the doctor in the example above may recommend the procedure to the patient by saying something like:

(25) I recommend the procedure.

How, now, should the addressee evaluate the relevance of the exhortation? Or, a mother may command her child:

(26) Drink your milk.

In such cases, it could be argued that the addressee changes his current goals and preferences in light of his larger goals and preferences; the patient could reason that the doctor's recommendation would be good for him in the long term and the child could reason in the same sort of way. But this complicates the simpler theory of relevance as the value of information where a goal and decision problem are simply *given*. For example, the patient may earlier have decided against the procedure on the basis of his initial preferences; now, he changes these preferences by taking account of his longer-term preferences and then decides to opt for the procedure. In the course of this evaluation, the immediate preferences have changed, so how exactly should the value of information be computed? This raises thorny questions of intertemporal choice among others. Portner's (2004) proposal that imperatives result in a change in the addressee's To-Do List does not help because he seems to lack the crucial idea of evaluating actions (i.e. the members of To-Do Lists) in terms of preferences. The value of information involves calculating the change in payoff and the problem is which payoffs to use, the prior ones or the posterior ones after the utterance.

Both these and perhaps other such issues pose hurdles to developing a satisfactory theory of relevance, even one that is able to incorporate ideas of the value and worth of information. This behooves us to look at what role relevance is required to play in a theory of communication.

13.3 The role of relevance

In Relevance Theory, relevance is meant, roughly speaking, to replace the four Gricean maxims (Grice 1975; 1989a) with a single principle that can serve to explain human interpretive behavior. But the natural question that arises is whether *any* maxim or principle is required to explain human behavior once one has the apparatus of rationality and rational action (and of possible deviations from rationality in the sense of utility theory). If this rhetorical observation is correct, then of the three choices – four maxims (and their variants[11]), one principle, or just rational action with no further layer – it is clear that the last option is to be preferred. This is not only because of an obvious scientific principle of economy but also because it would bring the field of communication in line with the rest of the social sciences by recognizing a full-fledged notion of *agency* that is conspicuously absent from linguistics and the philosophy of language.

As I have argued in several places and as we have seen in detail in Part III, rationality by itself suffices to explain the transmission of locutionary meaning.

[11]See the neo-Gricean work of Horn (1984) and Levinson (2000).

It is only when *illocutionary* meaning comes into play that we need to rely on relevance viewed as the value of information. The key reason for having to introduce some external Constraints like Relevance (and Distance) is that for illocutionary contents like implicature we do not have any counterpart of linguistically sourced entities like conventional meanings (e.g. the conventional meanings of BANK enable us to identify its range of possible referential meanings) to generate the *possibilities* which the Flow Constraint would disambiguate. Illocutionary meanings are more or less purely contextual and emerge through a kind of bootstrapping process which is facilitated by having access to the notion of the value of information. In other words, the sub-Constraints of Relevance and Distance together identify the illocutionary possibilities from the entire universe of contents; then the games of partial information disambiguate among these and determine one or more contents as those communicated. This is very different from the role of relevance in Relevance Theory where it determines all contents directly, both locutionary and illocutionary.

To the extent relevance is required, perhaps the main obstacle to using it is intensionality. As described in footnote 9, *all* the surgeons did not deviate from rationality, only some did. This makes things somewhat unpredictable since different agents in identical situations behave quite differently. I will accept these limitations of the decision-theoretic notion of the value of information to serve as our criterion of relevance and use it in what follows.

Because illocutionary meanings are not constrained by conventional meanings as locutionary meanings are, they may bear no very direct relation to locutionary meanings as is familiar from many examples of implicature. Almost anything can be illocutionarily implied given the appropriate circumstances. This means the potential candidates from which the Flow Constraint enables a choice of one or a few optimal meanings can come from anywhere in the infon space $\mathcal{I}$ mentioned in Section 2.1. In the main example $\varphi = \varphi_1\varphi_2$ = BILL RAN of Part III, an enrichment or completion such as *Bill Smith ran for President of the US* might be very relevant by almost any criterion of relevance, whether the *amount* of information notion of Relevance Theory or my less problematic *value* of information notion. By both these criteria, this content would be far more relevant than the more plausible and intuitively acceptable content *Bill Smith ran in the local election* because it would either generate many more contextual implications (i.e. positive cognitive effects) as required by Relevance Theory or would deliver a greater change of payoff as required by my account. This observation likely applies even to any intuitive criterion of relevance such as Grice's.

This ought to make it clear that relevance by itself is not sufficient for identifying the potential candidates or possibilities that enter into the Flow Constraint.

Some other constraint is also required and this turns out to be the new idea of *distance*. Incidentally, this idea has nothing to do with the measure of weighted psychological distance introduced in the discussion of vagueness in Chapter 11. As we will see, it may even be possible to manage just with this new idea and do without the idea of relevance altogether.

14 Distance

It is reasonable to suppose that just as people form intuitive judgments about physical distance which may be more or less accurate, so people routinely form intuitive and subpersonal judgments about the distance between two infons. This judgment about informational distance is always relative to the situation in which it is made and also relative to the goals of the agents in that situation. It is a subjective assessment that sometimes becomes conventionalized.

Just as one thinks of physical distance in terms of the path that takes one from here to there, one can think of informational distance in terms of the steps that take one from one infon to another, so perhaps the best way to fix this notion is through human reasoning. Agents attempt to derive information that enables them to fulfill their goals. Such inferences involve a variety of modes of reasoning including especially abduction, but also deduction, induction, probabilistic (e.g. Bayesian) and plausible reasoning, and so on.[1]

To start with, there is a situation u and a corresponding Communication Game Γ_u. Γ_u contains a Setting Game SG_u which has information about the setting with respect to which some communication takes place. Either because of prior utterances or because of the interaction between the agents, both of which are housed within SG_u, some information about the agents' explicit or implicit goals may become public. If the goals are not already common knowledge, the agents invoke the Cooperative Principle to assume that their goals are shared. For example, $\mathcal{A}$ and $\mathcal{B}$ may be discussing where to eat, in which case their (implicit) goal would be to eat out together. This goal is common knowledge so the Cooperative Principle is already satisfied.

Next, an agent may utter a sentence φ based on a Generation Game GG_u induced by a Content Selection Game CSG_u which in turn is induced by SG_u. This utterance is evaluated by $\mathcal{B}$ in the Interpretation Game UG_u and will have some locutionary content σ^ℓ determined via the locutionary game $LG_u(\varphi)$ as discussed in Part III. Recall that the Setting Game or conversational goals influenced the formation of preferences and prior probabilities in $LG_u(\varphi)$ which helped determine locutionary meaning. Now we will see how they help in identifying illocutionary meaning.

[1] Adler & Rips (2008) is a large collection of papers that discuss a variety of modes of reasoning.

Once $\mathcal{B}$ derives σ^{ℓ} he checks if it fulfills their shared goal, shared either by virtue of it being common knowledge or by being assumed via the Cooperative Principle. If it fulfills the goal, there is nothing more to be done. In this event the whole intended meaning of the utterance is just σ^{ℓ}. If it falls short of the goal, then a search for one or more illocutionary meanings is *triggered*.

It is in the course of this search that distance and relevance play a role in generating the possibilities for illocutionary meaning that are later disambiguated by the illocutionary Flow Constraint which is made up of the illocutionary global game $IG_u^{\mathcal{B}}(\varphi)$. There are no conventional meanings to generate the possibilities as there were for locutionary meaning. So the illocutionary Semantic Constraint can no longer rely on the Conventional and Referential sub-Constraints. Instead, it must use the Relevance and Distance sub-Constraints to generate the possibilities. The search for illocutionary possibilities takes place within the infon space $\mathcal{I}$ and involves a *derivation* of one or more contents from the locutionary content relative to u. This derivation involves a sequence of steps from initial premises to conclusions and has a length.

In general, the situated distance $d_u^{\mathcal{B}}(\sigma, \tau)$ between two infons σ and τ for $\mathcal{B}$ with respect to u can be defined as the length of the derivation from σ to τ that $\mathcal{B}$ actually undertakes in u relative to $\mathcal{B}$'s goals in SG_u (that are assumed to be shared with $\mathcal{A}$) and relative to other infons in u. These other infons in u may be part of the agent's explicit beliefs, they may come from what she is perceiving, they may come from the common ground of the dialogue taking place, and so on. In any event, they constitute a *finite* set.[2]

To clarify what is meant in the above definition of distance by "relative to $\mathcal{B}$'s (shared) goals in SG_u," there is always a step in the reasoning involving an implication such as $\tau' \longrightarrow$ Goal where Goal is an infon representing a conversational goal. Then, either because Goal is common knowledge or because it can be assumed via the Cooperative Principle, $\mathcal{B}$ can infer τ' by abduction. Usually, $\tau' = \tau$, the end state of the derivation, but occasionally there may be one or more steps beyond τ' to τ. Also, there may be other such implications like, for example, $\tau'' \longrightarrow$ Goal, in which case τ'' could be another conclusion reached by the search. This will become clearer when I discuss concrete examples below. As I will also

[2]It is not clear how the underlying space of infons through which the agent searches is structured but it may be thought of as a graph of some kind. A graph is a set of vertices representing the infons in the space together with directed or undirected edges linking them that represent various kinds of inferential steps. This notion of distance is not a metric as the symmetry condition $d_u^{\mathcal{A}}(\sigma, \tau) = d_u^{\mathcal{A}}(\tau, \sigma)$ for all σ and τ may not be satisfied. In any case, this condition is not required for the idea to work. See Barwise & Etchemendy (1991: Section 3) for a partial implementation of such a graph.

show, because human reasoning is extremely broad and versatile, employing all kinds of inferential steps, it is probably not possible to capture it in any logical calculus or pin down the definition of distance more precisely.

There will be a corresponding distance $d_u^{\mathcal{A}}(\sigma, \tau)$ for $\mathcal{A}$ as both agents have to reason from σ to τ, one in the Generation Game and the other in the Interpretation Game. This is what would make the final result τ common knowledge if it survives all the other requirements for being an illocutionary meaning. But the precise derivations for the two agents may differ and so $d_u^{\mathcal{A}}(\sigma, \tau)$ may not equal $d_u^{\mathcal{B}}(\sigma, \tau)$.

This idea of distance constitutes the second sub-Constraint of the Semantic Constraint for illocutionary meaning. The way the two sub-Constraints of Relevance and Distance are meant to work together once a search is triggered is as follows. Assume for the moment that the only indirect meaning being considered is an implicature.[3] Essentially, the infon σ and other infons present through u and through $\mathcal{B}$'s goals provide a space through which $\mathcal{B}$ searches either consciously or nonconsciously for derivations to a sufficiently *relevant* set of conclusions. When the distance, that is, the length of these derivations, is also sufficiently small, they become possible candidates for the Flow Constraint. In other words, distance and relevance have to be compared with situated thresholds. Crucially, one does not have to check each infon in $\mathcal{I}$ to find the candidates, which is an infinite task. One starts from the locutionary content or its proper parts and then proceeds *locally* until one reaches a relevant conclusion that fulfills one's goal within a specified radius.

Suppose the distance and relevance thresholds are $\epsilon_{d,u}$ and $\epsilon_{R,u}$ respectively. I have omitted the agent superscripts on these thresholds to avoid clutter but different agents will, in general, have different thresholds. One can now identify an "open ball" $\text{Ball}_u^{\mathcal{B}}(\sigma^\ell, \epsilon_{d,u}) = \{x \in \mathcal{I} \mid d_u^{\mathcal{B}}(\sigma^\ell, x) < \epsilon_{d,u}\}$ inside which all distances from σ^ℓ to x must lie.[4] Further, only those x that also have $\text{Relevance}_u(x) > \epsilon_{R,u}$ qualify as possible implicatures. Thus, the set of possible implicatures is given by $\{x \in \text{Ball}_u^{\mathcal{B}}(\sigma^\ell, \epsilon_{d,u}) \mid \text{Relevance}_u(x) > \epsilon_{R,u}\}$. Because agents are finite, the space searched is finite. So it becomes possible to search locally in a computationally tractable way within the ball for sufficiently relevant candidate implicatures. Jurafsky & Martin (2009: Section 1.3) discuss a variety of algorithms (e.g. graph

[3]In earlier chapters, I have used the terms *locutionary* and *illocutionary* for direct and indirect contents. Here, I will use the two pairs of terms interchangeably to make the writing and reading smoother. The first terms of each pair correspond to the content that results primarily from the conventional meanings of the uttered words whereas the second terms of each pair correspond to the content that arises primarily from the context. This is not meant to be a black-and-white distinction.

[4]I have borrowed the term "open ball" from the context of metric spaces.

algorithms such as depth-first search) for such tasks when they have to be performed by artificial agents. These candidate implicatures can then be supplied to the illocutionary Flow Constraint which disambiguates among them and chooses the best ones.

The reason why the two thresholds above are subscripted with u is that they may be different for different kinds of communication. When interpreting a novel or a poem, for example, it is reasonable to form extended derivational chains consciously. In such cases, the open ball will be much larger than it would be for ordinary face-to-face communication. In the case of Relevance, too, different kinds of conclusions might be found relevant depending on what is being interpreted, whether some immediate communication or some remote text. In the latter case, an addressee's tolerance may be greater, making the corresponding threshold lower.

I repeat that the local search for implicatures and other indirect meanings is triggered by the locutionary content or its proper parts falling short of the agent's goals in one way or another. If such a content is adequate, there will be no need to search for further meanings. This means there are three steps in eliciting indirect meanings: first check if the locutionary content or its proper part fulfills the conversational goal appropriately and, if not, then search locally via the illocutionary Semantic Constraint consisting of Relevance and Distance to produce candidate implicatures or other indirect meanings, and, last, submit these candidates to the illocutionary Flow Constraint for disambiguation.

The procedure sketched above provides a more or less complete description of the task of computing the implicatures of an utterance, when to look for them and how to find the possibilities and how to identify the best ones. In implementing such an idea computationally it would be necessary to set up an actual search space of infons and various inferential modes of traversing this space. Needless to say, this is nontrivial to do. To make things more concrete, reconsider Grice's example of the stranded motorist I discussed in Section 5.2. Let the sentence uttered by the passerby in response to the motorist saying "I am out of gas" be:

(27) There is a garage around the corner. (ψ)

As I noted earlier, an utterance of ψ in the context described has been generally taken to have the implicature that the speaker thinks the garage might be open. How would such an implicature be derived in a rigorous and complete way given just the locutionary content of the utterance and u?

First, note that the prior utterance, "I am out of gas," makes the goal of the motorist public in the Setting Game for the utterance of ψ in u. Since it is not common knowledge this goal is *shared* by the passerby, this is assumed by invoking the Cooperative Principle. The locutionary content σ^ℓ of ψ in u does not, by itself, fully meet this goal of getting gas. So a local search for one or more implicatures is triggered.

If the implicature referred to above is denoted by τ, then it is plausible to say that in the utterance situation u the two infons σ^ℓ and τ are relatively *near* each other, that is, $d_u(\sigma^\ell, \tau)$ is within the threshold $\epsilon_{d,u}$ or, equivalently, τ lies inside the corresponding open ball. This is because the chain of reasoning from σ^ℓ to τ is quite short and the distance may even be just 3 or 4 units depending on how one counts the steps, which depends on how the search space is set up in the agent's head. For example, the addressee may start with common knowledge of σ^ℓ that the garage is around the corner and common knowledge that σ^ℓ falls short of his goal. Then he may observe that the speaker has uttered the sentence ψ in response to his goal. If the speaker's cooperation is assumed, it follows that the speaker would not have said this if he knew the garage was closed. Indeed, it is plausible to conclude that the speaker thinks it is open because this fulfills the addressee's goal. This is about 3 or 4 steps to τ and all of them are common knowledge. So the distance from σ^ℓ to τ is about 3 or 4 units as undertaken in u and would certainly be contained within the open ball. So the first requirement of Distance has been met.

Since clearly $\text{Relevance}_u(\tau) > \epsilon_{R,u}$ as well because the motorist is out of gas and needs to fill up (i.e. this value of information is also related to the interlocutor's goal), one strand of the local search stops with τ and we can conclude that τ qualifies as a *candidate* implicature. Crucially, it is not necessary to pull τ out of thin air and then evaluate its distance and relevance. On the contrary, τ is *discovered* by the agent as he searches locally through the associated space. This candidate implicature then becomes one *possible* interpretation in the relevant illocutionary partial information game. As will be shown in the next chapter, such an implicature would be derived with some probability as a solution to this game so that its being an implicature is *seldom* certain. This is because it is possible that the passerby does not give any thought to whether the garage is open or not.

Both the Distance and Relevance sub-Constraints are evaluated relative to the agent's goals. But the nature of the computation in each case is quite different. In the former, a derivation is constructed, and in the latter, the *value* of the information is computed. It may happen that there is no derivation that fulfills the goals and then a suitable conclusion has to be drawn (e.g. that the speaker does not know the required information or is cooperating only partially or not at all). I will look at such an example in Chapter 16.

How does the search algorithm know where to stop as it reasons along the search space starting from σ^ℓ? With each step it takes to some $x \in \mathcal{I}$, it tests whether $\text{Relevance}_u(x) > \epsilon_{R,u}$ and whether the goal is fulfilled. If not, it keeps going. For example, the intermediate steps in the three- or four-step reasoning above were not sufficiently relevant. Finally, it either reaches an infon like τ which *is* sufficiently relevant and which fulfills the goal or it goes outside the ball. If the former, τ is accepted as a candidate; if the latter, it returns empty-handed to σ^ℓ. In either case, a new local search begins.

If we assume both interlocutors are aware of a few gas stations nearby then such a new search might lead to another *possible* implicature τ' that the gas available at the garage around the corner is inexpensive as the passerby may be offering the best option among the available ones. This is because τ' is also near σ^ℓ as τ was and is discovered in the same way. It is possible that $d_u(\sigma^\ell, \tau) < d_u(\sigma^\ell, \tau')$ even though τ' may also lie within the open ball. Its relevance is also sufficiently high in u though, again, it is likely that $\text{Relevance}_u(\tau) > \text{Relevance}_u(\tau')$. So τ' could also become a candidate implicature. It is hard to be sure whether the speaker is possibly conveying τ' or not because goals such as a general preference for inexpensive gas can be implicitly shared. Most likely, it can be taken to be an implicature by the Flow Constraint with an even lower probability than τ. Such considerations show how much indeterminacy can be present at the level of both the Semantic and Flow Constraints.

A possibility I have never seen addressed is why the passerby might not be implying that he would fetch the gas from the garage for the motorist. Call this τ''. It is assumed because Grice assumed it that the implicature is τ and not τ''. The latter certainly fulfills the motorist's goal and is certainly relevant: indeed, its value may be potentially higher even if its probability may be lower. But, in most circumstances, τ'' would be ruled out by everyone. It would *not* be near σ^ℓ as even though $\tau'' \longrightarrow \text{Goal}$, the abductive inference to τ'' would be blocked because it is not the *best* explanation. So even though it is relevant, its distance is infinite or at least beyond the open ball. Practically every method of deriving implicatures (e.g. Gricean or Relevance-Theoretic or Benz 2010) fails to disqualify such possibilities. As such examples are omnipresent, this shows their account of implicature is seriously inadequate.

Finally, as mentioned in Section 5.4, there is a certain indeterminacy between whether the illocutionary meaning in question is the implicature τ or is the modulated meaning τ''', the content that there is a *possibly open* garage around the corner. That is, the distance $d_u(\sigma^\ell, \tau''')$ is also small.

The search algorithm terminates its search when no new steps can be taken from σ^ℓ. Note that I have informally described one version of depth-first search and it is not the only way that this search might be organized. The goal of any algorithm is to locate all the infons in $\{x \in \mathrm{Ball}_u^{\mathcal{B}}(\sigma^\ell, \epsilon_{d,u}) \mid \mathrm{Relevance}_u(x) > \epsilon_{R,u}\}$, the set of possible indirect meanings starting from σ^ℓ.

In principle, there could be other sets of possible indirect meanings issuing from the proper *parts* of σ^ℓ as well. The distance to τ''' that I took to be $d_u(\sigma^\ell, \tau''')$ would actually be $d_u(\langle\!\langle \text{garage} \rangle\!\rangle, \tau''')$ where $\langle\!\langle \text{garage} \rangle\!\rangle$ is the proper part of σ^ℓ that corresponds to the word GARAGE in ψ. Such examples arise not just with modulated meanings but also with implicatures and free enrichments and other indirect meanings triggered by subsentential expressions.

In the discussion of this example I have throughout used distances without any agent superscript. In fact, similar computations have to be undertaken by both agents, the speaker in the Generation Game and the addressee in the Interpretation Game. Even though the details of the calculations are private, each agent has common knowledge that such calculations exist in the other's mind. This is what results in common knowledge of the implicature once it passes the further requirement of the illocutionary Flow Constraint.

I have given a fairly detailed analysis of this example but it is not, in fact, detailed enough. I have focused mainly on the addressee's goal which is assumed to be shared via the Cooperative Principle. But the speaker's goals also play a role as they are also implicitly shared. For instance, the inference to τ'' is blocked in most circumstances because it is implicitly shared that the passerby has other goals of his own and would not want to fetch the gas for the motorist. Thus, there is generally a mutual sharing of both agents' explicit and implicit goals that permit some candidates to be licensed while others are eliminated. This implies that cooperation is almost always *partial*, a nuance that is missing from Grice's Cooperative Principle.

I have implicitly assumed that each step of a derivation is equally easy for the agent. But this may not be true. For example, the steps involved in reaching τ' may be harder than those involved in reaching τ as τ' is less related to the agent's immediate need. So one can assume that each step comes with a weight or cost and the distance then is not the length of a derivation but its cost. If the search space is a graph, then its edges would have weights. This idea of a weighted graph could model the more imaginative kinds of reasoning involved in interpreting the indirect meanings of a poem or novel. There is a certain creativity involved in such interpretations and the task of the literary critic is a costly one.

The conception of distance I have offered does not yet tell us how the implicature of an utterance of HE IS A FINE FRIEND might be derived because the implicature is precisely the negation of the locutionary content as discussed in Section 5.2. How can the agent start from a content and arrive at its negation?

For this, a slight adjustment in the procedure is required. Recall that the distance between two infons σ and τ is to be calculated relative to other infons present in u. Such other infons include background knowledge that the agents share or are perceiving as well as the common ground of the dialogue taking place. So the procedure has to start with a whole *set* of premises, not just a single infon σ. This set could be internally inconsistent – the speaker could not be conveying the locutionary meaning because it contradicts the common knowledge the agents have that the person is question has cheated the speaker – and so the reasoning process must include cases where such a set of infons may need some revision, just as happens in situations where our beliefs have to be revised sometimes in the light of new knowledge or because they are internally inconsistent. This kind of revision may result in rejecting σ, the starting point, or one of the other premises, depending on which of these are potentially alterable. In our example, only the locutionary content can be negated as the other situational infons are taken as true. This broader conception of reasoning that includes such revision is therefore required to account for all cases of illocutionary meaning.

In general, then, the locutionary meaning σ^ℓ or its proper parts serve as the baseline relative to which distances for various types of potential illocutionary meanings are measured. I will assume henceforth that people are able to form such intuitive judgments (either consciously or nonconsciously) about the magnitude of distances like $d_u(\sigma^\ell, x)$ for various x in $\mathcal{I}$. If one wishes, one can call the Distance sub-Constraint a *theory* of "accessibility," recalling that this was precisely one of the questions Relevance Theory had simply begged.

In Relevance Theory, relevance and accessibility are more or less equated because the most accessible meaning always turns out to be the most relevant. In Equilibrium Semantics, Relevance is the *value* of information which is completely distinct from Distance or the derivational accessibility of information, and both of these sub-Constraints are quite different from the Flow Constraint. It is the three constraints together that lead one to a complete theory of illocutionary meaning. And, moreover, in Equilibrium Linguistics, the goals and preferences of the interlocutors present in the Setting Game always enter all three constraints, the calculation of relevance and distance, and the calculation of equilibria of partial illocutionary games.

Why are illocutionary partial information games required? Isn't it enough to identify the maximally relevant meanings inside the open ball? Unfortunately, *all* such nearby relevant meanings may not be compatible with the speaker's and addressee's viewpoints. They also have to satisfy the requirement of "optimality" based on the agents' preferences. Otherwise, for instance, an addressee could infer a very valuable proposition that is easy to access derivationally but that was not intended by the speaker. This kind of miscommunication occurs quite frequently, in fact. For instance, in the garage example, as I just noted, the passerby may not have given any thought to whether the garage is open or not. In such a case, the motorist would be wrong to infer τ as an implicature even though it is both relevant and near. But this ties in with the general fact of the indeterminacy of indirect meanings.

To the best of my knowledge, neither Grice nor any of his followers have ever shown a way to *find* candidate implicatures or other candidate illocutionary meanings such as free enrichments and modulations. They have, at best, provided more or less convincing ways – usually less convincing, as I have argued in Section 5.2 and in this chapter – to determine if a candidate is actually an implicature, say, *given* that it is a candidate. But the candidate itself is always pulled out of thin air. It is to the credit of my first book, *The Use of Language* (2001: Chapter 7), that this problem was first considered partially, and to the credit of Equilibrium Semantics that it now offers a *complete* and *computationally tractable* solution to the problem, at least in principle. If suitable assumptions are made about the search space, it may become possible to put the theory of implicature and other indirect meanings on a solid scientific footing and explain how it applies to humans at a psycholinguistic level. It may also make it feasible for robots to communicate effectively with indirect meanings, maybe for limited purposes rather than for open-ended conversations, as they can be endowed with appropriate search spaces. Texts such as Manning & Schütze (1999) and Jurafsky & Martin (2000; 2009) are largely silent about how indirect meanings may be derived. Jurafsky (2004) does compare and contrast the logical Plan Inference (or Belief-Desire-Intention) Model developed mainly by Allen, Cohen, and Perrault (see Allen 1995) with his probabilistic Cue-based Model in the context of speech act interpretation and similar tasks like reference resolution and discourse structure interpretation. As will be seen below, the approach of Equilibrium Semantics integrates the two methods in some sense, as it provides the depth of the first and the breadth of the second.

I now consider an example in greater detail to show how the sub-Constraints of Relevance and Distance work together in enabling the identification of candidate illocutionary meanings, that is, the illocutionary possibilities, that would then be disambiguated by the Flow Constraint, that is, by $IG_u(\varphi)$, the global set of local illocutionary games. I discuss this in the context of free enrichment so it also shows how the same method applies to other types of indirect meaning.

15 Free enrichment

Everyone acknowledges the phenomenon of incomplete utterances requiring free enrichments, contents that Grice never quite foresaw. But there are significant differences in how they are perceived and, to the best of my knowledge, no one has seriously tried to present a method to derive them from first principles. Before we do so, it is helpful to set the background and context for the discussion.

15.1 Representationalism and Contextualism

Representationalism and Contextualism are two influential approaches to enrichment. The Representationalists, descendants of the logicists, admit that illocutionary contents intrude into what is said[1] but argue that these meanings are *conventionally* mandated by covert parts of the utterance in line with Grice's conventionalist conception of what is said. They allow context and inference to play a role in determining enrichments but insist that they are nevertheless triggered by "syntactic" elements *represented* in the utterance. In other words, they see all such meanings arising from *saturation*, the filling of an explicit or *implicit* slot (i.e. a constrained variable) in a sentence. The Contextualists, descendants primarily of ordinary language philosophers, regard these contents as *not* being conventionally enjoined but resulting entirely from the context.

Representationalism is a frankly rearguard action to defend a losing position. If illocutionary encroachments upon what is said are granted and the role of context and inference in their determination is allowed, then much of the battle has been lost. The conventionalist view that what is said is largely conventional *has* to give ground to the contextualist view that what is said is largely contextual. Indeed, if an actual derivation from first principles is attempted, it would become obvious just how much of the former position is untenable. So bringing in conventional triggers is a feeble bulwark from this larger perspective, especially when it is grasped how rapidly these hidden variables multiply. A simple

[1]By "what is said," I mean the literal meaning or content or what is generally called the *truth*-conditional content, though I prefer to call it the *information*-conditional content. The reason for preferring the latter is that truth conditions are not fine-grained enough to distinguish between an utterance of "He is playing" and "He is playing and 2 + 2 = 4."

sentence with only five words like "Every man wants to leave" could have as many as *four* "aphonic" or covert slots!

Sometimes the best defense is an offensive action and it is the Representationalists' criticisms of Contextualism that are important. One advantage Representationalists tout is that the slots they posit *constrain* enrichments. They see context as being too unruly to be able to control what can become an indirect content. Can an utterance of "Every Frenchman is seated" ever convey that every Frenchman or Dutchman is seated, and can an utterance of "Everyone loves Sally" ever convey that everyone loves Sally and her mother?[2] Representationalists and Contextualists both share the intuition that they cannot, though the Contextualists seem unable to rule out such additions to the direct content as they appear not to have any access to the kinds of constraints associated with the slots of the Representationalists. This is the *overgeneration problem* for the Contextualists: they have no mechanism to outlaw certain types of seemingly unwarranted indirect or illocutionary contents. Recanati (2010), one of the foremost proponents of Contextualism, has even conceded that modulation is *unsystematic*, that is, that no principled derivation of modulations is possible.

This has led to a kind of impasse because Representationalism is just too baroque to be taken at face value and Contextualism appears unable to subdue problems like overgeneration. In my view, many of these difficulties stem from the fact that neither side in this debate actually knows how to derive any content, whether illocutionary or even locutionary (e.g. "straightforward" disambiguation of a word like "bank" or "ran" in an utterance as presented in Part III).[3] If they did, many aspects of this situation would no longer require an appeal to questionable intuitions.

I will show how to solve, even *dissolve*, the overgeneration problem. Both sides take indirect contents to be an *either-or* or *zero-one* matter. Either some content is possible or it is not. I will argue not only that almost any indirect content can attach to any direct content in some context or other – *contrary* to the common intuition of Representationalists and Contextualists – but also that content as such is often more nuanced than being a simple *all or nothing* issue. Grice (1975; 1989a: 39–40) himself more or less *required* such indeterminacies in the calculation of implicatures but this subtle point has been largely forgotten. Indeed, the procedure I outline applies equally to *all* indirect contents *including* implicature, and once this is recognized, it becomes clear that overgeneration is a potential

[2] See Stanley (2007: Chapter 7) and Recanati (2010: Introduction) for a discussion of these examples.

[3] I have already argued in Section 13.1 that the mechanism offered by Relevance Theory is seriously flawed.

problem for implicature as well.[4] But the real lesson from this observation of a uniform approach to all illocutionary contents is this: since no slots are possible for implicatures as they would be too limiting and even perhaps a little absurd, so no slots are necessary for other illocutionary contents like enrichment and modulation either! If a method can be shown to work for implicatures without slots, then the same method is likely to work for other indirect contents as well.

15.2 How to think about content

Early on, Frege (1892/1980) and Russell (1919) and others who followed attended to just the putative single literal content of a *sentence*. The pervasive ambiguities of utterances were seldom acknowledged, let alone dealt with.[5] The later Wittgenstein (1953/1968) and Austin (1961) did much to broaden this picture by bringing in *utterances* as the bearers of meaning, but even they succumbed to the lure of single determinate contents. Grice did allow indeterminacies in implicature, in fact making them one of its necessary attributes. But, influenced by the logicist view of what is said, he left literal content determinate and singular. This view still persists and even the indeterminacy of implicatures has been largely ignored. So while new kinds of meaning have been identified – such as enrichments and modulations – the idea that the overall content is singular and determinate still holds sway.

In Parikh (2006b; 2010), I argued for a very different way to think about the content of an utterance. I tried to show that meaning is almost never determinate or singular. As I say in Parikh (2010: 217–218):

> First, the content of an utterance is not just a single infon (or proposition) as is usually assumed, but rather a collection of infons, possibly infinite in number, relative to a described situation. Second, each of these infons is just *partially* present as a member of the content through a probability attached to it. Third, not every component of the content may be explicitly intended by the speaker or explicitly inferred and represented by the addressee. And lastly, the content may be different for the speaker and the addressee – there is no "objective" content. This lack of objectivity is reinforced by … various sources of indeterminacy … . Thus, in general, the flow of information is also

[4]As I said in Section 5.2, Grice's own mechanism is also seriously flawed and, in any case, it provides little protection against overgeneration because the mechanism is so vague.

[5]To repeat, by "ambiguity" I mean something broader than mere lexical and structural ambiguity, though these are the plainest kinds. I intend to include all the multiple interpretations an utterance might possibly convey.

> a partial affair, determined by the degree of overlap between the speaker's meaning and the addressee's interpretation.
>
> A visual image one might draw of content is as follows. Given a space of points representing the entire class of infons (i.e. a lattice), the content for each agent would be a subspace with each point assigned a number representing its probability. Then each agent involved would have his or her subspace and the overlap between them would determine the degree and nature of the flow of information that took place between them.

This is a very different conception of content, whether literal or not. If we reflect in a naïve way how we experience utterances in ordinary exchanges and in literature, it should be apparent that there is some plausibility to the picture I am proposing. In this view, meaning is not a *singular* or *binary* affair as the Representationalists and Contextualists both assume. *Multiple* contents are all *partially* present in the overall meaning of an utterance. This partial presence is captured via the probabilities that the respective infons or propositions are being *conveyed* by the speaker; they are *not* the probabilities of the contents themselves. The points in the subspaces above can be shaded darker or lighter depending on the strength of their presence.

Regarding the third point in the quote above about the necessity of intentions, here is a quote from Thomason (1990: 349) that is indirectly related to this view:

> All this shows, I think, is that 'Did you mean to say' is used to query people about the consequences of what they have said, without distinguishing these clearly from what they said. More generally, 'Did you mean to do' does not distinguish between intentions and foreseeable side effects, which is exactly the distinction we need to make here. To achieve some progress in these negative cases, it seems that we will need to say more about the intensionality of intentions in general, and of intentions to assert in particular, and also to find tests that extract more reliable conclusions from the evidence. The philosophical background, of course, suggests that it won't be easy to make progress on these matters.

This passage reveals that Thomason was grappling with this very issue of indeterminacy in how to attribute meaning to an utterance in the special case pertaining to what is said and its consequences. But because he was locked into the intentional and Gricean way of seeing content, he felt that progress would be difficult. As observed in Section 5.3, intentions are not required to be explicitly present for all that is meant. Looking to the example of φ below, the fact that

Smith's weight refers to his weight *on Earth* in most ordinary situations could be part of what was meant without it being intended by the speaker, simply because we are usually on Earth and this is just taken for granted. (And it could be understood as such by the addressee without being explicitly inferred and represented in her beliefs.) It would be quite reasonable for the speaker to assent if he was asked later if "he meant to say it." In addition, it is equally plausible that such contents be only partially present in the overall content of the utterance with greater or lesser probability.

Consider the sentence φ discussed by Cappelen & Lepore (2005b):

(28) Smith weighs 150 lbs. (φ)

An utterance of φ by itself does not tell us whether Smith weighs 150 lbs.

1. when naked in the morning
2. when dressed normally after lunch
3. after lunch in heavy outer clothing
4. after being force-fed four liters of water
5. four hours after having ingested a powerful diuretic
6. on Earth
7. on the moon
8. and his height is five and a half feet

In the appropriate circumstances, one or more of the above completions could be legitimate probabilistic meanings of φ, even #8. It all depends on how much is shared between speaker and addressee, what is available to both in the utterance situation, and what their goals are. Indeed, there is even an indeterminate *number* of possible completions for such an utterance, not all of which might have been explicitly intended by the speaker but which neither agent can entirely eliminate because the circumstances might be sufficiently ambiguous and unclear. We can often ask after an utterance has elapsed: could the speaker also have meant this? And equally often the answer will be: possibly, though maybe with rather small probabilities. It is this more generous conception of meaning I am advocating as it appears to fit the evidence better. Whether an indirect meaning is triggered or not depends on the interlocutors' goals in part and these may not always be explicit and explicitly shared. Of course, when a context is specified more or less determinately, some of these enrichments may well become intuitively implausible and their probabilities will then plummet to zero.

The solution to the overgeneration problem then is to recognize two things. First, it is almost impossible to eliminate a possible enrichment or modulation as a *potential* candidate relative to *some context or other*. There could always be a context that allows a certain indirect content to become plausible. Secondly, depending on the utterance situation, a potential illocutionary candidate could become a *partial* presence in the overall meaning of an utterance. A probability of zero would be attached to the majority of infons simply because there are so many infons, but zero is also not so different from a vanishingly small positive number.

These two things, possibility and probability, correspond to two steps in the method of Equilibrium Linguistics. First, the possible enrichments have to be generated via the Semantic Constraint consisting of the sub-Constraints of Relevance and Distance *after* the need for them has been triggered because the direct content or some proper part of it falls short of the shared conversational goals. Second, some of these have to be retained and others discarded in a probabilistic way via the Flow Constraint. I now proceed to show how these two steps work.

15.3 The Semantic Constraint: Generating possible indirect contents

The direct or locutionary content of φ, *Smith weighs 150 lbs.*, is given by the infon $\sigma^{\ell} \equiv \tau_0$. This content, which I will refer to as τ_0, *together* with each of the completions above will give rise, respectively, to the infons τ_i with $i = 1, 2, \ldots, 8$. Thus, τ_1, for example, will be the content *Smith weighs 150 lbs. when naked in the morning*, τ_2 will be the content *Smith weighs 150 lbs. when dressed normally after lunch*, and so on.

Each of these contents will be plausible candidates in some context or other. In many ordinary situations, τ_0 itself will satisfy the goals in the Setting Game and so there will not be any need for a completion. But, in other situations, τ_0 might fall short of these goals and then a search for completions will be triggered. I leave it to the reader to imagine such circumstances for each completion and will simply assume that they are available to trigger a search. Thus, as we have seen in Chapter 14, the set of possible completions for $\mathcal{B}$ for each u is given by $\{x \in \mathrm{Ball}_u^{\mathcal{B}}(\tau_0, \epsilon_{d,u}) \mid \mathrm{Relevance}_u(x) > \epsilon_{R,u}\}$.

If, for example, Smith's doctor is making a statement about his weight, then completions τ_1 and τ_2 may have high relevance and so would perhaps τ_5 and τ_6 although it is possible that, say, τ_3 could also be included for either the speaker or the addressee. The locutionary content τ_0 would also potentially be sufficiently relevant by itself. In such a scenario, the distance to these infons from τ_0 might also be sufficiently small. One can estimate the distances as was done in the example of the stranded motorist.

Given this specification of *u*, a new facet of triggering becomes visible. As I have just said, a local search for completions would have to be triggered by some inadequacy in τ_0 with respect to the conversational goals inferred from the Setting Game in *u*. But this inadequacy may not always be clear-cut because τ_0 may only *partially* or hazily fulfill it. In the case of the garage, this inadequacy seemed relatively straightforward to ascertain but even there it is not completely obvious. This is why the locutionary content (or a proper part of it) often figures as a probabilistic presence in the illocutionary content. This will become clearer presently.

Returning to the Semantic Constraint, if we consider only relevance and ignore distance we may end up with a content the addressee would like to infer but for which there is little warrant because it is too remote – as argued at the end of Section 13.3. This is also the kind of case that was mentioned earlier when I raised the question whether "Everyone loves Sally" could ever imply the content that everyone loves Sally and her mother. This may be very relevant for the addressee to know but it may be too distant a content given what has been uttered. It all depends on the utterance situation and it may be possible to describe a context where the distance is not so great. In any case, relevance *by itself* is not sufficient.

Likewise, if we include distance but drop relevance, then we may get possible completions that do not matter to the addressee and that the addressee would not have any reason to entertain. This could occur if φ were uttered in a casual conversation where the detailed nuances of Smith's weight are irrelevant to the addressee and would never be inferred *even* postdictively if he were asked for his assent or dissent. So distance by itself is also not sufficient but it should be noted that it is often less onerous to admit an irrelevant meaning than it is to admit a remote meaning because the latter may lead to more serious errors about what the speaker was trying to convey.

Nevertheless, it may be possible to develop a complete theory based only on distance and the Flow Constraint, that is, without considering relevance. This would naturally greatly simplify the picture as we would be rid of the problems (e.g. the intensionality of choice) the notion of relevance faces. The argument for this is that the notion of distance accounts sufficiently for the goal of the interlocutor. That is, if the distance requirement is satisfied then the relevance requirement will automatically be satisfied. In yet other words, sufficient informational proximity *implies* sufficient relevance. This is an interesting conjecture. For now, I prefer to retain both sub-Constraints since it is easier to drop something later than add it in. I will return to this alternative way of developing the account in Section 18.1.

The set of candidate completions $\{x \in \text{Ball}_u^{\mathcal{B}}(\tau_0, \epsilon_{d,u}) \mid \text{Relevance}_u(x) > \epsilon_{R,u}\}$ does not require τ_0 to be a part of x. This allows one to include possibilities where the direct content may be rejected altogether, as happens in cases of irony – for example, an utterance of "He is a *fine* friend" where exactly the opposite is being conveyed. I have also simplified the notation by using τ_0 in the distance calculation rather than any *part* of τ_0 because it can and often does happen that enrichments attach to parts of the locutionary content rather than to the whole locutionary content.[6] Not only that, it may often *not* be necessary to compute the whole locutionary content before deriving a local enrichment.

Because agents are limited, they may fail to compute the set of candidate completions exhaustively. Later, someone may inquire of a speaker if some content was part of what he meant or of an addressee if that content was part of what she inferred. If it had not been evaluated earlier, these agents would need to then test that content postdictively. For example, in most ordinary situations, the completion #6 (i.e. the content τ_6) would be felt to be both of high relevance and close proximity relative to the agents' goals but neither agent may consciously pause to compute its relevance and distance because it may not appear "salient": it may be taken for granted implicitly and so no triggering inadequacy would be detected. It would remain latent unless it is brought up later in the exchange or an ethnographer asks the interlocutors about it.

Not only are the assessments of relevance and distance rough and subjective, the two thresholds are also rough and subjective, and all these things vary from situation to situation too. So it is not unlikely that the speaker and different addressees come up with somewhat different possible enrichments. This is part of the indeterminacy of any utterance. What about completion #8 or τ_8? Can it ever, so to speak, cross these thresholds? I don't see why not. Certainly, its relevance may be high in u as the addressee may be interested in Smith's height. Its distance from τ_0 would generally also be high rather than low as required because it is unlikely to bear an inferential relation to τ_0 in most circumstances where the goals do not permit it. So it would generally be eliminated as a possibility. But weight is roughly correlated with height within a certain range and if two

[6]See Parikh (2010: 156) for more details.

doctors were discussing Smith's weight it is conceivable that his height could figure as an indirect content even if only his weight was made explicit because the distance between the two would be relatively low in such circumstances relative to their goals.

While the inferential relations between weight and height may be partly physiological, the possible inferential relations between Sally and her mother may be social and the possible inferential relations between Frenchmen and Dutchmen may be social or historical. Even merely hypothetical inferential relations can contribute to the candidacy of an indirect content by making the relevant distances low enough. So almost any indirect content can be added to any direct content given the right circumstances and goals if the inferential relations between the infons make the distance between them small assuming the infon also has high relevance. The reason why such enrichments appear impossible in ordinary circumstances is that they are generally *inferentially unrelated* to the direct content. That is, the distance lies outside the open ball in ordinary circumstances as no appropriate goal has been specified.

It now becomes clear that not only are slots *not* required for triggering the search for possible enrichments (or modulations), they may in fact be too constraining if they prevent certain types of possibilities and they may be too lax in licensing other possibilities. This is an interesting twist to the overgeneration problem because in a certain sense this observation *dissolves* the problem altogether. The trouble with positing slots is that they are *always* present in the sentence itself quite independent of the circumstances and, in particular, goals. This means they have no connection to what is actually going on in the communication and so will sometimes undergenerate (e.g. as in the examples of weight and height or Sally and her mother) and sometimes overgenerate (e.g. when no completion is warranted). Goals and the circumstances are appropriately connected to the communication and they, therefore, generate just the right completions for every utterance.

Here is a quote from Recanati (2010: 11):

> I agree with Stanley that certain things don't happen, that would happen if modulation were totally unconstrained. But who claimed that modulation was totally unconstrained? Work in this area precisely needs to address the issue of what is possible and what is not (step one), in order to arrive at suitable generalizations (step two), which it will then be incumbent upon pragmatic theory to derive (step three) (Elbourne 2008).

I believe the criterion involving Relevance and Distance does carry out all three steps. In a related footnote, Recanati says, "One obvious constraint, implicit in my writings on the topic, is that modulation should preserve semantic types, just as adjunction preserves syntactic types." If such a constraint on semantic types really does turn out to be required, it would automatically be handled via the Distance sub-Constraint because two infons involving different types might naturally be too remote in *most* situations. This situated way of bringing in possible constraints on semantic types is better because it allows a more contextual formulation of the generalization. Language is usually too rich for unsituated "semantic" diktats to work.

More conscious and deliberate processes of determining candidate completions and other illocutionary candidates are required when interpreting literature. Consider this poem by Emily Dickinson (1993: 18):

> Tell all the Truth but tell it slant –
> Success in Circuit lies
> Too bright for our infirm Delight
> The Truth's superb surprise
>
> As Lightning to the Children eased
> With explanation kind
> The Truth must dazzle gradually
> Or every man be blind –

The broad meaning of the poem appears to be fairly clear, but what about its nuances? What, for example, is to be included in "all the Truth?" Interpreting such utterances requires some *creativity* and not only because the context for the text is incompletely specified and, indeed, specifiable. There is a range of possibilities and even coming up with them involves some imagination. This can now be explained by the cost of certain inferential paths being higher and requiring greater effort. One may have to consciously but informally judge distances as part of an extended deliberation about the poem. This is part of the reason why literary criticism is a professional activity. And it is part of the pervasive indeterminacy of language and meaning.

Deciding which candidates are in fact part of the poem's meaning is another matter altogether. It is to this latter problem, the matter of deciding what in fact *is* the content of an utterance, that I now turn. This brings in a further filtering action, this time via game theory.

15.4 The Flow Constraint: Eliminating possible indirect contents

For the sake of concreteness, I will assume that the Relevance and Distance criterion is triggered and generates just τ_1, τ_2, and τ_6 as possible indirect contents for our example. The rest can be deemed to be either irrelevant or too distant or both.

The first thing to notice is that some items in the list have to be mutually exclusive whereas others can coexist. Both τ_1 and τ_2 cannot hold simultaneously though they can both coexist with τ_6. This introduces a slight complication in how we evaluate the possibilities. Sets of mutually exclusive candidates have to be considered together in the same game whereas those that do not conflict must be treated separately.

The second thing to notice is that τ_6 may never be consciously entertained by either speaker or addressee. It is, so to speak, *fully* situated though it may come up later in the conversation in an explicit way.

Let us start with the games that arise from contemplating τ_1 and τ_6. Both games will involve comparing the relative merits of just τ_0, the direct or locutionary content, with the possible indirect or illocutionary contents τ_1 and τ_6. Take a look at the illocutionary game $g_1^\iota \equiv g_u^\iota(\varphi)$ in Figure 15.1.[7]

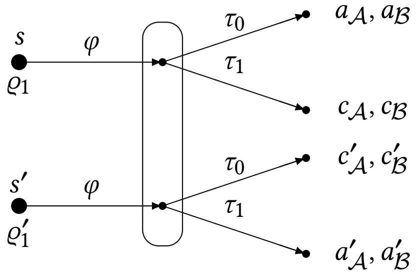


Figure 15.1: The game g_1^ι for illocutionary content τ_1

This is a partial information game as before but it is an *illocutionary* partial information game. $\mathcal{A}$ utters φ = SMITH WEIGHS 150 LBS. in u, its optimal locutionary meaning $\sigma^\ell \equiv \tau_0$ = *Smith weighs 150 lbs.* (and optimal parse t) are determined by $\mathcal{B}$ via the locutionary game $LG_u(\varphi)$, and then they trigger various illocutionary enrichments that need to be derived via corresponding illocutionary games.

[7]The superscript of g_1^ι is the Greek letter "iota," a mnemonic for "illocutionary."

I have avoided subscripting the initial situations s and s' which should be s_1 and s'_1. The reader should read the symbols as carrying the appropriate intentions to convey τ_0 and τ_1 in this game and appropriate contents in later games that will soon appear. The payoffs are as they were in the locutionary games and obey the same inequalities. The symbols for the prior probabilities are ϱ_1 and ϱ'_1 and are distinct from the earlier symbols of Part III which were ρ_1 and ρ'_1. They are different forms of the Greek letter "rho."

These prior probabilities are also conditional probabilities like those of the locutionary games of Part III. In particular, $\varrho_1 = P(\tau_0 \mid \tau_0; u)$. Recall the meaning of the probability notation: $P(\tau_0 \mid \tau_0; u)$ indicates the probability $P(\mathcal{A}$ is conveying τ_0 in $g_1^I \mid \mathcal{A}$ is conveying τ_0 in $LG_u(\varphi); u)$ or, more simply, $P(\mathcal{A}$ is conveying τ_0 indirectly $\mid \mathcal{A}$ is conveying τ_0 directly; $u)$. Here, the first occurrence of τ_0 is part of the illocutionary game g_1^I whereas the second occurrence of τ_0 is part of the locutionary game $LG_u(\varphi)$. If the two games had been the same, then obviously $P(\tau_0 \mid \tau_0; u)$ is just 1. But we are trying to determine whether τ_0 is also an indirect content *given* that it is a direct content. As I said a short while ago, a locutionary content may be hazily inadequate and so it may occur probabilistically as an indirect content as well.

In the same way, $\varrho'_1 = P(\tau_1 \mid \tau_0; u)$. Here, there is no clash between the symbols of different games as the indirect content is τ_1 and the direct content is τ_0.

We can naturally allow t, the optimal parse of φ = SMITH WEIGHS 150 LBS., to also influence these indirect contents by simply adding it in as follows: $\varrho_1 = P(\tau_0 \mid \tau_0, t; u)$ and $\varrho'_1 = P(\tau_1 \mid \tau_0, t; u)$. Whether t actually influences the indirect content is an empirical question; my purpose here is just to show that both situations can be easily accommodated in the framework.

This also raises the much more interesting question whether indirect meanings influence direct meanings and parses. If this is the case, then again it is very easy to capture this simply by adding in the relevant indirect meaning τ_0 or τ_1 as a *conditioning* variable that would play a role in determining locutionary meanings and parses. And simultaneous two-way influence can be readily entertained by having variables such as x'_1 for the indirect meanings τ_0 and τ_1 and variables such as x_i and y_j for direct meanings and parses. Then, everything would be one grand system of equations encompassing both locutionary and illocutionary contents. However, there may be no empirical warrant for this in ordinary face-to-face communication and one should be cautious about advocating such a comprehensive system. When reading is involved, on the other hand, especially with literature where interpretive efforts are more likely to be extended, such a comprehensive system is more likely to be present than not, as a reader may consciously reject a direct meaning and parse because their indirect meanings prove unsatisfactory relative to the indirect meanings of alternative direct meanings and parses. In such cases, the direct meaning and parse are selected on the basis of their more satisfying indirect meanings and so depend on them. In general, the matter should be decided purely on empirical grounds and my purpose here is just to show that both kinds of possibilities can be effortlessly represented in the framework of Equilibrium Linguistics. An entirely new effect that can be conceived is the influence of indirect meanings on optimal parses and vice versa.

One such example of a seemingly common type that can occur even in face-to-face communication is as follows. Suppose $\mathcal{A}$ and $\mathcal{B}$ are talking about Russia and $\mathcal{B}$ asks: "Is Russian literature any good?" Then $\mathcal{A}$ may respond: "Well, Dostoyevsky was Russian." Then "Dostoyevsky" could refer to either the novelist or other Russians with the same last name and it seems it is the implicature that Russian literature is great that enables the addressee to choose the novelist. In other words, the indirect meaning helps to determine the direct meaning and vice versa.

One difficulty such a comprehensive system faces is in explaining why suboptimal direct meanings and parses do not seem to generate their own illocutionary possibilities which would then feed back into bolstering them, however feebly. For example, φ = SMITH WEIGHS 150 LBS. has the alternative suboptimal (in u) locutionary meaning τ_0' that Smith determines the weight of £150 (in coins). This suboptimal meaning would require its own enrichments (e.g. *in coins*) but we are not generally aware of them. We are not generally aware of the locutionary meaning τ_0' either but, as mentioned in Section 12.1, we seem to unconsciously activate alternative *lexical* meanings. Most likely, the processing does not go beyond this to a full-scale sentential referential meaning such as τ_0' and so the warrant for alternative indirect meanings triggered by τ_0' may be weak in such face-to-face interactions *unless* the indirect meanings result from a proper part of τ_0'. In any case, in more literary situations, such interdependent effects are likely to be more common.

From a computational standpoint, too, it may be interesting to consider such comprehensive influences as they provide some more data for eliminating unwanted possibilities though they also substantially complicate the calculations. To keep things simple in this book, I will assume just a one-way influence (without the intervention of the optimal parse t) and take $\varrho_1 = P(\tau_0 \mid \tau_0; u)$ and $\varrho_1' = P(\tau_1 \mid \tau_0; u)$.[8]

Even to the newcomer to game theory, g_1^ι should now be easy to solve. All that is required is a simple Nash equilibrium that can be computed by evaluating the inequalities between corresponding expected payoffs:

[8]Such two-way influences were explicitly considered in my previous book in Chapter 4.

either

$$\varrho_1 a_{\mathcal{A}} + \varrho_1' c_{\mathcal{A}}' < \varrho_1 c_{\mathcal{A}} + \varrho_1' a_{\mathcal{A}}'$$

and

$$\varrho_1 a_{\mathcal{B}} + \varrho_1' c_{\mathcal{B}}' < \varrho_1 c_{\mathcal{B}} + \varrho_1' a_{\mathcal{B}}'$$

or

$$\varrho_1 c_{\mathcal{A}} + \varrho_1' a_{\mathcal{A}}' > \varrho_1 a_{\mathcal{A}} + \varrho_1' c_{\mathcal{A}}'$$

and

$$\varrho_1 c_{\mathcal{B}} + \varrho_1' a_{\mathcal{B}}' > \varrho_1 a_{\mathcal{B}} + \varrho_1' c_{\mathcal{B}}'$$

which simplifies to:

$$\frac{\varrho_1}{\varrho_1'} \lessgtr \frac{a_{\mathcal{A}}' - c_{\mathcal{A}}'}{a_{\mathcal{A}} - c_{\mathcal{A}}}$$

and

$$\frac{\varrho_1}{\varrho_1'} \lessgtr \frac{a_{\mathcal{B}}' - c_{\mathcal{B}}'}{a_{\mathcal{B}} - c_{\mathcal{B}}}$$

As assumed in Section 8.1, if the payoffs are "symmetric" then each of the two numerators on the right-hand side equals the corresponding denominator on the right-hand side, and this leaves us simply with:

$$\varrho_1 < \varrho_1' \text{ or } \varrho_1 > \varrho_1'$$

Often, there will be no situational evidence to warrant either strict inequality and we will simply have:

$$\varrho_1 = \varrho_1'$$

as both possibilities will be equally likely. When this happens, an infinite number of mixed strategy equilibria with some probability weight π_0 on τ_0 and therefore $\pi_1 = 1 - \pi_0$ on τ_1 become available. It is possible for π_0 to be 0 or 1 in which event the corresponding pure strategy, either τ_1 or τ_0, would be selected as the equilibrium.[9] In actual practice, as I pointed out in Section 8.1, all $\mathcal{A}$ and $\mathcal{B}$ have to do is compare certain probabilities, which is what the computation of Nash equilibria under the assumptions about symmetric payoffs amounts to, and this seems within the grasp of partially rational agents. Moreover, if we do not assume a two-way influence between locutionary and illocutionary meanings then the probability comparisons are much simpler. How do the agents pick the weights π_0 and π_1? Again, in practice, these are generally rough estimates like *high*, *me-*

[9]I have omitted φ from these equilibria as it is always present.

dium, and *low*, which can be thought of as fuzzy intervals (but see the discussion below as well).

So if $\varrho_1 < \varrho_1'$ then τ_1 is the solution with probability $\pi_1 = 1$ and if $\varrho_1 > \varrho_1'$ then τ_0 is the solution with probability $\pi_0 = 1$. In cases where $\varrho_1 = \varrho_1'$ and some mixed strategy solution is expected, the solution to g_1^ι might be (τ_1, π_1) with the understanding that this implies the complementary solution (τ_0, π_0). For all situations then, we can write the general form of the content roughly as collections of sets like $\{(\tau_i, \pi_i) \mid i \in I\}$ where τ_i are the range of indirect contents, π_i their probabilistic weights as derived from the relevant games g_i^ι, and I an index set. For example, the game in Figure 15.2 might have the solution (τ_6, π_6) where τ_6 = *Smith weighs 150 lbs. on Earth.* Both of these indirect solutions would be gathered together and the indirect content expressed as $\{\{(\tau_1, \pi_1)\}, \{(\tau_6, \pi_6)\}\}$.

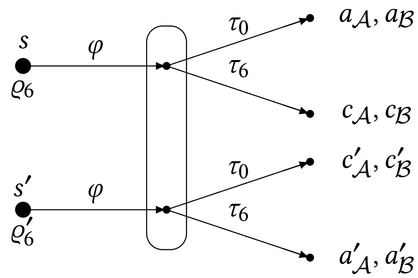


Figure 15.2: The game g_6^ι for illocutionary content τ_6

Intuitively, this tells us that τ_1 (i.e. *Smith weighs 150 lbs. when naked in the morning*) is a meaning of the utterance of φ in some situation u with a probability of π_1 (which would be some number like 0 or 0.3 or 0.7 or some interval) and that τ_6 (i.e. *Smith weighs 150 lbs. on Earth*) is also a meaning of the same utterance with a probability of π_6 (also a number or interval, presumably higher, like 0.9 or even 1 depending on u). In other words, there will be a range of indirect contents, each with some probability weight. Obviously, as I said earlier, these weights are not the probabilities of the infons themselves but the probabilities that the respective contents are being conveyed by the speaker.

When mutually exclusive indirect contents are considered, they have to be evaluated simultaneously in the same game. So the game for τ_1 and τ_2 would not be g_1^ι and a corresponding g_2^ι similar to it, but would be g_{12}^ι as shown in Figure 15.3 where I have deliberately left out the payoffs. In this case, the mixed strategy solution would be written as $\{(\tau_0, \pi_0), (\tau_1, \pi_1), (\tau_2, \pi_2)\}$ with the natural proviso that $\pi_0 + \pi_1 + \pi_2 = 1$.

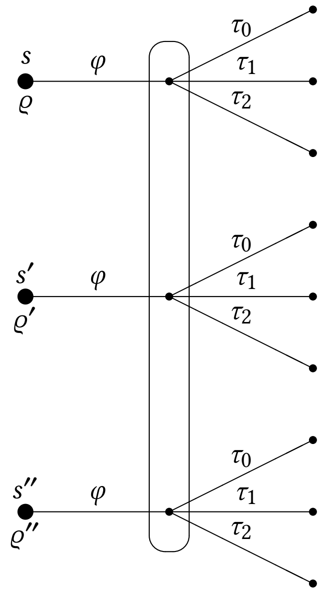


Figure 15.3: The game g^{ι}_{12} with two illocutionary contents τ_1 and τ_2

If we now consider the overall content of the utterance, assuming that only items #1, 2, and 6 or τ_1, τ_2, and τ_6 get through the Relevance plus Distance filter, it would have to include τ_0 with probability 1 as the direct content,[10] it would then include τ_0, τ_1, and τ_2 as one set of mutually exclusive indirect contents with probabilities π_0, π_1, and π_2, respectively, and, last, it would include τ_0 and τ_6 as indirect contents with probabilities π_0' and π_6. This content would be expressed as follows:

$$\{\ \{(\tau_0, 1)\},\ \{(\tau_0, \pi_0), (\tau_1, \pi_1), (\tau_2, \pi_2) \mid \pi_0 + \pi_1 + \pi_2 = 1\},\ \{(\tau_0, \pi_0'), (\tau_6, \pi_6) \mid \pi_0' + \pi_6 = 1\}\ \}$$

If we make the constraints on the probabilities implicit, this simplifies to:

$$\{\ \{(\tau_0, 1)\},\ \{(\tau_0, \pi_0), (\tau_1, \pi_1), (\tau_2, \pi_2)\},\ \{(\tau_0, \pi_0'), (\tau_6, \pi_6)\}\ \}$$

This is *exactly* the form of the content I referred to earlier in Section 15.2 where

[10] The probability that some direct content is being conveyed is not always 1. With a pun, it might be 0.5.

I said it was a collection of infons each with a probability. This can be visualized on the infon lattice as suggested earlier. The entire proposition expressed would be:

$$c \models \{\ \{(\tau_0, 1)\},\ \{(\tau_0, \pi_0), (\tau_1, \pi_1), (\tau_2, \pi_2)\},\ \{(\tau_0, \pi_0'), (\tau_6, \pi_6)\}\ \}$$

where c is some appropriate described situation. For this example, this would be the *literal* content of the utterance. It includes both the locutionary content $\sigma^\ell \equiv \tau_0$ with probability 1 as well as various indirect contents that have crossed the Relevance plus Distance barrier. In other examples, there may be modulations and other indirect contents to consider as well as we will see in the next two chapters.

Note that π_0 is in general different from π_0' and both are in general different from 1, and yet these three numbers all attach to the same content τ_0. This requires some interpretation. The direct content τ_0 (i.e. *Smith weighs 150 lbs.*) is fully part of the overall content with probability one in this example. Its presence in the other indirect contents is *relative* to the other indirect components (i.e. either τ_1 and τ_2, or τ_6). It provides a constraint on the other indirect infons by limiting their participation in each indirect content.

Also, the speaker and addressee are unlikely to share the same content not only because they may have generated different possibilities but also because the game-theoretic computations may yield different solutions. In other words, the single content expressed above would split into two separate contents, one for the speaker and the other for the addressee, with possibly different infons and possibly different probabilities. This again is just a related aspect of indeterminacy. These literal meanings would look as follows:

$$c^{\mathcal{A}} \models \{\ \{(\tau_0, 1)\},\ \{(\tau_0, \pi_0^{\mathcal{A}}), (\tau_1, \pi_1^{\mathcal{A}}), (\tau_2, \pi_2^{\mathcal{A}})\},\ \{(\tau_0, \pi_0^{\mathcal{A}'}), (\tau_6, \pi_6^{\mathcal{A}})\}\ \}$$

and

$$c^{\mathcal{B}} \models \{\ \{(\tau_0, 1)\},\ \{(\tau_0, \pi_0^{\mathcal{B}}), (\tau_1, \pi_1^{\mathcal{B}}), (\tau_2, \pi_2^{\mathcal{B}})\},\ \{(\tau_0, \pi_0^{\mathcal{B}'}), (\tau_6, \pi_6^{\mathcal{B}})\}\ \}$$

We should remind ourselves that the illocutionary games g_i^ι I have been discussing belong to the *objective* illocutionary global game $IG_u(\varphi)$ which is a part of the objective global game $G_u(\varphi) = LG_u(\varphi) \cup IG_u(\varphi)$, as I have said earlier. There are naturally the corresponding *subjective* global illocutionary games $IG_u^{\mathcal{A}}(\varphi)$, $IG_u^{\mathcal{B}}(\varphi)$ which are in turn parts of the corresponding subjective global games $G_u^{\mathcal{A}}(\varphi) = LG_u^{\mathcal{A}}(\varphi) \cup IG_u^{\mathcal{A}}(\varphi)$ and $G_u^{\mathcal{B}}(\varphi) = LG_u^{\mathcal{B}}(\varphi) \cup IG_u^{\mathcal{B}}(\varphi)$. To stay within the bounds of *communication* as delineated by Definition 5.3 in Section 5.3, I have simply assumed that u is such that $G_u(\varphi) = G_u^{\mathcal{A}}(\varphi) = G_u^{\mathcal{B}}(\varphi)$ and $IG_u(\varphi) = IG_u^{\mathcal{A}}(\varphi) = IG_u^{\mathcal{B}}(\varphi)$ and, further, that they all become (nonconscious) common knowledge between $\mathcal{A}$ and $\mathcal{B}$. But, especially with illocutionary games, it is all too common that there is miscommunication or a weaker flow of information than strict communication and this is then often explained by a divergence among these three games or by the lack of common knowledge.

Indeed, the solutions even to identical subjective games that are common knowledge can diverge because different probabilistic weights $\pi^{\mathcal{A}}$, $\pi^{\mathcal{B}}$ may attach to these contents. Since some of the prior probabilities in such games are often equal, an infinite number of mixed strategy equilibria result, and in these circumstances $\mathcal{A}$ and $\mathcal{B}$ may well arrive at different weights for various indirect contents.

There appear to be two possible sources of such weights. Recall that $\{x \in \text{Ball}_u(\tau_0, \epsilon_{d,u}) \mid \text{Relevance}_u(x) > \epsilon_{R,u}\}$ is the set of candidates that contains infons that are both sufficiently close and sufficiently relevant. However, the distance and relevance of each candidate will, in general, be different. Those candidates with greater proximity and greater relevance ought to be more probable than those candidates with lower proximity and lower relevance as they are the contents more likely to be conveyed by the speaker. Thus, the probabilistic weights $\pi^{\mathcal{A}}$, $\pi^{\mathcal{B}}$ come from the *relative* proximity and relevance of different indirect contents in the solution. If there had been just one criterion, either just Distance or just Relevance, the proportion of the probabilistic weights would just be the same as the proportion of either distance or relevance. Since there are two criteria, some suitable function of the two is required to work out the ratio of the probabilistic weights.

There will also be situations when one prior probability outweighs the others and then the corresponding illocutionary content could be the pure strategy equilibrium for both agents with probability 1. This outcome is not unusual but since there will always be a range of indirect contents, it is likely that at least some of them will enter the literal content probabilistically. For example, it may be that, in the above two contents for $\mathcal{A}$ and $\mathcal{B}$, $\pi_6^{\mathcal{A}} = \pi_6^{\mathcal{B}} = 1$ but that the other probability weights lie strictly between 0 and 1.

There may be some interest in actually *measuring* just how much $\mathcal{B}$'s inferred meaning *diverges* from $\mathcal{A}$'s intended meaning.[11] For this, one way is to enlist the

[11] As noted earlier in Part III, $\mathcal{A}$ knows what he wants to convey with certainty but he nevertheless has to derive it based on the Generation Game which contains his subjective representation of the relevant games of partial information. That is, he has to figure out what $\mathcal{B}$ would infer and it is this meaning that is given by $\{\ \{(\tau_0, 1)\}, \{(\tau_0, \pi_0^{\mathcal{A}}), (\tau_1, \pi_1^{\mathcal{A}}), (\tau_2, \pi_2^{\mathcal{A}})\}, \{(\tau_0, \pi_0^{\mathcal{A}'}), (\tau_6, \pi_6^{\mathcal{A}})\}\ \}$.

idea of *relative entropy* or the so-called Kullback-Leibler divergence from information theory.[12] Since each content is a collection of probability distributions $\pi^{\mathcal{A}_j}$ and $\pi^{\mathcal{B}_j}$, $j = 1, 2, \dots$, over the same corresponding spaces of sets of infons,[13] we can compute the relative entropy for corresponding pairs of distributions which is defined as:

$$D(p \parallel q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

where p, q are the two probability mass functions being compared, the logarithm is with respect to base 2, and X is the relevant set of infons. Extremal cases involving zeros are defined thus: $0 \log 0/q = 0$ for all q and otherwise $p \log p/0 = \infty$.

$D(p \parallel q)$ measures the expected number of extra bits required to code samples from p when using a code based on q, rather than using a code based on p. Generally, p represents the "true" distribution of data, observations, or a precisely calculated theoretical distribution. The measure q typically represents a theory, model, description, or an approximation of p.[14]

For each pair of corresponding distributions $\pi^{\mathcal{A}_j}$, $\pi^{\mathcal{B}_j}$ with $\mathcal{A}$'s distribution substituted for p and $\mathcal{B}$'s for q, this will yield a number. Since a content involves *multiple* pairs, we can take the *sum* of these numbers to find the *total relative entropy* of the two contents. To ensure that we have the same corresponding pairs of distributions in $\mathcal{A}$'s and $\mathcal{B}$'s contents, we may occasionally need to introduce a distribution in either content where certain infons get a zero probability. This may pose a small problem of interpretation because when $p(x) \neq 0$ and $q(x) = 0$ the divergence would become infinite. To avoid this, we could apply some kind of smoothing operation (e.g. introducing a small positive number in place of the zero).

As there is no "true" or "objective" meaning, this formula yields simply the total divergence (i.e. the sum of the expected number of extra bits required with respect to each pair of corresponding distributions $\pi^{\mathcal{A}_j}, \pi^{\mathcal{B}_j}$) of $\mathcal{B}$'s derived meaning from $\mathcal{A}$'s intended meaning.[15] The obvious use of this measure is to see how far apart the two are, which would give an idea of the extent to which the communication has been a success. An identity of speaker meaning and addressee

[12] See Manning & Schütze (1999: 72).

[13] I am taking $\pi^{\mathcal{A}_j}$ and $\pi^{\mathcal{B}_j}$ to be just one pair of distributions within the content. As we have just seen, there will in general be *multiple* such pairs of distributions within the overall literal meaning. That is, the superscript j is different from the subscripts on the probabilities π, the latter representing actual probabilities, not whole distributions.

[14] This measure of proximity between two distributions is *not* a metric in the usual sense. There are other ways of measuring distance (e.g. Bhattacharyya distance which is given by $D_B(p, q) = -\ln(\sum_{x \in X} \sqrt{p(x)q(x)})$) which involve real metrics.

[15] I have deliberately ignored the possible differences between $c^{\mathcal{A}}$ and $c^{\mathcal{B}}$.

interpretation is seldom achieved and represents just an ideal case.[16] Yet, this assumption has been the norm for much of semantics.

If there are two or more addressees and one speaker in an utterance situation, we can now compare whose meanings are closer to each other and whose further apart. Another application is to extend this formula to a further sum for entire discourses or dialogues. A third is to situations like Chinese whispers where a relay of utterances that attempts to preserve content is involved.

To recap what we have achieved: I have described the process of deriving indirect contents such as free enrichments from first principles as involving three steps: first, the need for indirect contents is triggered by the agents' goals not being met by the locutionary content, then finding candidate possibilities through local search via relevance and distance, and then solving certain games of partial information to either eliminate or include these candidates in a probabilistic way. The same kind of reasoning would enable us to derive the literal meaning of the sentence φ = BILL RAN from Part III, enriching its locutionary content *Bill Smith stood (for election)* to *Bill Smith ran in the local election.* The steps comprise four illocutionary Constraints: the same Phonetic and Syntactic Constraint as before obviously, a new Semantic Constraint involving the sub-Constraints of Relevance and Distance which generates the possible indirect contents, and a new Flow Constraint involving illocutionary games of partial information that eliminates the suboptimal ones.

Illocutionary meanings such as τ_6 (i.e. *Smith weighs 150 lbs. on Earth*) may never be consciously entertained by either speaker or addressee. In this case, the game g_6^ι would not be played by either agent and so its solution would not enter into either agent's conscious content. It is then up to the theorist or ethnographer to impute this fully situated meaning to them. There is no mystery about this: if they were asked if they counted it as part of their understanding, they would invariably give their assent in most situations.

I have considered whether illocutionary meanings could affect locutionary meanings and parses, a possibility Equilibrium Linguistics effortlessly allows. I have chosen to err on the side of caution in this book for oral utterances by assuming that locutionary meanings influence illocutionary ones but not vice versa.

The picture of the meaning of an utterance as spread out over the entire space of infons with each point in the infon lattice assigned a probabilistic weight based on the likelihood of its being conveyed by the speaker appears very attractive as

[16] See Parikh (2001; 2006b; 2010).

it seems to fit our empirical uncertainty about whether some (direct or indirect) content is in fact part of the meaning or not.

The Nash and Pareto-Nash equilibrium concepts more readily lend themselves to such probabilistic infons than does the Risk Dominant Nash equilibrium concept. If this probabilistic picture of content that I have informally visualized is found to be empirically accurate, then it suggests that the Nash and Pareto-Nash ideas are perhaps the more appropriate ones to consider.

In the example of implicature about the garage in Chapter 14, the reader should have no trouble seeing how both the implicatures τ and τ' might be derived in corresponding illocutionary games, one with σ^{ℓ} and τ as $\mathcal{B}$'s choices and the other with σ^{ℓ} and τ' as $\mathcal{B}$'s choices, the solution to the first assigning a relatively high probability to τ and the solution to the second assigning a lower probability to τ'.

Processing Relevance and Distance as well as the Flow Constraint may seem like a lot psycholinguistically even though the brain is very fast. This is why human agents generally derive and become aware of just a few indirect meanings, especially in oral communication. But it is clear that implicatures such as τ' can be derived if we spend more time reasoning. This is especially true when more complex forms of communication are involved as happens with literature. The inferences take more time and we are more conscious of them. This is an added reason why semantics and psycholinguistics should not be conflated.

I have also tackled the problem of overgeneration in two stages both of which involve indeterminacies. Based on Relevance and Distance and the Flow Constraint, I argued that almost any indirect content can attach to a direct content in some (possibly outlandish) context or other.

This single argument ought to kill two birds, the overgeneration problem raised by the Representationalists and their charge of unsystematicity that Recanati has acknowledged. I have tried to show by actually deriving the enriched content of a sample utterance from first principles that both can be defeated and the broad position of Contextualism defended if one adopts the ideas of Equilibrium Linguistics. This requires giving up many cherished Gricean tenets but these are suspect anyway. And my framework does a lot more: it offers more precise and robust ways of computing not just direct contents but also indirect contents, including not just enrichments but also modulations and implicatures, all using just a few uniform and more or less self-evident principles of ontology and partial rationality.

This completes my discussion of free enrichment. I now briefly consider implicature before moving on to modulation.

16 Implicature

This is the classical type of illocutionary or indirect meaning. It involves a range of examples that are often differently classified. As I have already analyzed a couple of examples of implicature informally in the previous two chapters, I will not develop a more full-blown account here. The considerations are very similar to those in Chapter 15 in the discussion of free enrichment.

Once the need for one or more implicatures is felt because the locutionary content does not fully realize the goals of the conversation that are more or less public in the Setting Game, the addressee undertakes a local search using the Distance and Relevance sub-Constraints and then submits the candidates thus found to the Flow Constraint. This is in fact the procedure more or less for all indirect meanings and especially for free enrichment and implicature.

Just to amplify my discussion, I will briefly analyze two more examples. The first is a slightly modified example from Grice (1975). One agent asks another where Pierre lives, saying he wants to send him something, and the reply he gets is:

(29) Somewhere in the south of France. (η)

First, η needs to be enriched to *Pierre lives somewhere in the south of France* because without this the first agent's goal would not be met. The enrichment is not only derivable at a close distance and high relevance, it also sails through the Flow Constraint. But even with this enrichment, $\mathcal{B}$'s goal is not fulfilled. There is no implicature available given the information in u that allows an actual address to be inferred. So the only possible conclusions that can be drawn are that $\mathcal{A}$ does not know or is unwilling to cooperate (i.e. the Cooperative Principle does not hold). This is within derivational reach, that is, they are within the open ball as they can be inferred from an ambient fact like *if $\mathcal{A}$ had known and if he were cooperating, he would have given the details* and they are also relevant in the sense that they provide a negative answer to the query. And the two possibilities can be shown to go through the Flow Constraint with appropriate probabilities.

This is not all. $\mathcal{A}$ has not simply said he will not reveal Pierre's address, he has provided partial information regarding his whereabouts. This indicates he is not being completely uncooperative which might be taken as rude given some prior relationship between the two agents. As I said in Chapter 14, cooperation is often partial. This further fact – that the Cooperative Principle is nevertheless partially observed – can also be inferred as an implicature in the same way. Such contents based on maintaining relationships and being polite play a role in the corresponding Content Selection Game as well.

The next example is from Hugly & Sayward (1979: 22) who argue convincingly that Grice's own way of handling it was circular. In a conversation about whether Eisenhower was a great US president, one participant offers evidence of Eisenhower's generalship during the war and his great popularity, upon which $\mathcal{A}$ says:

(30) And he had a wonderful grin too. (η')

The locutionary content of this contribution does not achieve the conversational goal. So it triggers the search for an implicature. It can then be inferred in a few short steps that it would be common knowledge that it *obviously* does not meet the goal because having a wonderful grin has nothing to do with being a president, let alone a great one. Because of this obviousness, it draws a parallel with the other statements about Eisenhower's generalship and popularity as also being irrelevant, although less obviously so. This parallel is drawn because otherwise there would have been no good reason to make such an obviously irrelevant statement. In other words, its real meaning is to draw attention to the earlier statements as being similar with regard to its salient irrelevance. Thus, a locutionary content that had no value in the locutionary information it provided with respect to the conversational goal leads to a conclusion within the derivational ball that is also relevant to the goal. Further, it also goes through the Flow Constraint as the ambient facts are all common knowledge and so is the corresponding illocutionary game.

One can see from such examples that the kind of reasoning the agents have to undertake to derive implicatures can be quite sophisticated. It employs the full breadth of inferential modes and strategies available to us and so cannot be easily formalized. It includes even analogical reasoning as happened in the example above where the blatant irrelevance of a statement implies a relation of similarity to the less clear-cut irrelevance of the preceding contributions. Grice was right to emphasize such complex inferences and the Relevance theorists' approach becomes manifestly inadequate because they have no place for goal- and purpose-driven derivations, only for the completely blind process of deriving a large *number* of implications.

How do these analyses square with the maxim of Communication I introduced in Section 5.2? Pretty well, except that the more precise procedure I have offered in this chapter dispenses with the need for such an informally stated rule as implicatures are obtained from basic principles of rational agents trying to fulfill their goals, that is all.

Before I turn to modulation, I should mention that it is not only the addressee who infers implicatures and enrichments. Modulo various indeterminacies, the speaker, too, has to go through the calculations in his Generation Game by imagining how the addressee would carry out the required calculations. Only then can he choose his sentence optimally so that it achieves its ends in an optimal way by making certain things explicit and leaving others implicit.

Lastly, scalar implicatures of the kind discussed by Horn (1972; 1984), Levinson (1983; 2000), and many others can also be handled similarly. The key thing to note here is that not all the implicatures that a scale may license will actually be generated because the relevant goals may not be present. It will all depend on the situation. In this sense, there are possibly no generalized implicatures of the kind Grice had originally envisaged. Incidentally, Ross (2006) was probably the first to employ game-theoretic methods to analyze *complex* scalar implicatures. He showed that games of partial information (and Pareto-Nash equilibria) scale effortlessly toward this end.

17 Modulation

Modulation is possibly the most complex aspect of meaning. It straddles both locutionary and illocutionary meaning, *unlike* both free enrichment and implicature, and involves a more complex Semantic Constraint consisting not only of the Conventional and Referential sub-Constraints of Part III but also of the Relevance and Distance sub-Constraints of Part IV. Moreover, it requires for the first time a back-and-forth interaction between the Semantic and Flow Constraints.

Modulation was perhaps first identified by Cohen (1985; 1986), who cites Ross (1981) as an influence.[1] It was later picked up by several writers including Recanati (2004; 2010) and Wilson & Carston (2007) as an essential component of literal meaning. Cohen's own example is the use of a phrase like THE STONE LION where the content of LION has to be modulated to be made compatible with its adjective STONE. This fascinating aspect of meaning, possibly responsible for much in *meaning change* as one dimension of language change, can only be tackled by applying a more complex combination of the locutionary and illocutionary Semantic Constraints and a Flow Constraint that interacts with them.

It is best to understand the phenomenon through Cohen's (1986) own words. He contrasts "insulationism" and "interactionism" as two different ways in which the meaning of the utterance of a sentence depends on the meanings of its component words.

> According to the insulationist account the meaning of any one word that occurs in a particular sentence is insulated against interference from the meaning of any other word in the same sentence. On this view the composition of a sentence resembles the construction of a wall from bricks of different shapes. The result depends on the properties of the parts and the pattern of their combination. But just as each brick has exactly the same shape in every wall or part of a wall to which it is moved, so too each standard sense of a word or phrase is exactly the same in every sentence or part of a sentence in which it occurs. We may sometimes need to look at

[1]Barsalou (1983; 1987) investigated the related idea of ad hoc categories in a different psychological context.

neighboring words in order to discover the sense in which a word is functioning in the sentence in question, as we might infer the concavity of one brick from the convexity of its neighbor. But even then the meanings that we discover are not made what they are by one another, any more than the presence of a convex brick alters the shape of its neighbor. Rather, the words in the sentence have been given these meanings by diachronic facts of etymology. (page 223)

⋮

Interactionism makes the contradictory assertion: in some sentences in some languages the meaning of a word in a sentence may be determined by the word's verbal context in that sentence, though the extent and nature of this determination shows a wide range of variation. On this view the composition of a sentence is more like the construction of a wall from sand-bags of different kinds. Though the size, structure, texture and contents of a sand-bag restrict the range of shapes it can take on, the actual shape it adopts in a particular situation depends to a greater or lesser extent on the shapes adopted by other sand-bags in the wall, and the same sand-bag might take on a somewhat different shape in another wall or in a different position in the same wall. By exploiting local context in this way a language can be much more prolific of semantic variety than insulationism can give it credit for being. Moreover these sense-refinements or sense-modifications are generated by verbal interaction with a particular synchronic state of a language. Many of them, especially metaphors, are not common enough to be recorded in dictionaries or assigned dates and places of origin. (page 223)

⋮

Once the difference between insulationism and interactionism has been recognized it becomes clear that we cannot construct a semantics for any natural language along the same lines as a semantics for a formal system of any currently familiar kind. Projects like Davidson's or Montague's cannot succeed. And one can see the temptation for philosophers of language to relapse into a later-Wittgensteinian emphasis on anomalism. (page 230)

⋮

The interactionist must certainly grant that the borderline between the two types of meaning-determination is a fluid and flexible one. … Moreover, just as it is possible to shift the borderline between the two kinds of meaning-

> determination so as to make the insulationist story more extensively applicable, so too one could shift the borderline in the other direction. … In short, while homonyms like 'pen' and 'case' are definitely on the side of the borderline where insulationism holds sway, and indefinitely pliable verbs like 'drop' and 'make' are definitely on the interactionist side, there is a lot of polysemy in natural language that can apparently be treated in either way because it is just not known whether all the word-meanings concerned have actually occurred in human utterances. In such ambivalent cases considerations of theoretical simplicity are the only factors that can determine the issue. (232–233)
>
> ⋮
>
> It has to be emphasized, of course, that the details of verbal interaction in natural language are as yet little understood and that very many problems remain as yet unresolved. For example, when one word dominates another, as "stone" dominates "lion" in (1), or "geography" dominates "drop" in (2), what ensures that the domination proceeds in one direction rather than the other?[2] Does the less ambiguous dominate over the more ambiguous, or topic over comment, or the relatively abstract over the relatively concrete, or the inanimate over the animate? But whatever be the correct solution of these problems, it seems highly unlikely that we shall obtain any guidance towards finding it from the ideas of Tarski, Montague, etc., about the semantics of artificial languages. Recognition of the difference between insulationist and interactionist conceptions forces us to treat the semantics of natural language as a largely autonomous discipline, rather than as a topic for Davidsonian or Montagueian theory. (page 234)

This book, and my earlier work, has tried to tread precisely the fine line between the Scylla of "formalism" or, more precisely, "logicism" (Russell, Tarski, Montague, Davidson) and the Charybdis of "anomalism" (the later Wittgenstein's lack of system as described in the quote from Dummett in Section 1.2 as well as the similar outlooks of Austin, Strawson, and Searle). But the modulation of meanings shows us how formidable the task is. To date, there are no adequate theories of this phenomenon except for descriptive attempts by Recanati and the Relevance theorists and a few others.

The kind of disambiguation Part III looked at dealt with homonyms (e.g. words like BANK which can actually be thought of as two or more distinct words with

[2](1) Four stone lions occupy the corners of Trafalgar Square. (2) Most students here drop geography in their final year (where "drop" means "drop studying").

the same spelling). It does not show specifically how such meanings can be *modulated* in the context of a sentence. Cohen also seems to focus exclusively on the verbal context and leaves out modulations that occur on account of the situational context.

There is also sometimes real indeterminacy about whether a particular indirect meaning is a modulation or free enrichment or implicature. Both of the latter can often be viewed simply as modulations as I mentioned with Grice's example of the *possibly open* garage and as could also be said of the example of Smith weighing 150 lbs. *on Earth*. It is nevertheless right to treat the three phenomena as distinct and allow for indeterminacies from time to time. In such events, one agent may treat the matter one way and the other agent may treat it the other way.

The classic – and perhaps obvious – first approach to such a problem would be to mimic Grice's approach to implicature as might be suggested, for example, by Grice (1975) and Searle (1979). This involves computing the whole locutionary meaning of the utterance first, finding it inadequate in one or another way, and then recomputing a related meaning such as an implicature or a modulated meaning to make up for the inadequacy. Unfortunately, the experimental evidence (e.g. Frisson 2009) makes this kind of computation highly implausible as the time delays implied by the model are incompatible with those found empirically.

The alternative is to undertake "local" computations as opposed to "global" ones but use roughly the same kind of reasoning. In the case of THE STONE LION, both the determiner THE and the adjective STONE play a role in constraining the possibilities for the content of LION, the first by guiding the addressee to some resource situation or anaphoric anchor and the second by eliminating the possibility of a real animate lion in most ordinary contexts as inadequate. Modulations are generally activated by other words in the sentence (e.g. STONE) whose referential meanings seem incompatible with those of the word in question (e.g. LION).

Any theory that is developed must be compatible with the psycholinguistic evidence. But, at this stage, such experiments themselves are still in their infancy and so it is not clear what the final findings will be. So it is essential to build a framework that can itself be modulated in different directions and can offer different theories based on the data. I will offer one such theory in what follows. But the framework of Equilibrium Linguistics is flexible enough to house alternative accounts of the phenomena.

As I said in Section 12.1, Pylkkänen et al. (2006) and Frisson (2009) suggest that the conventional meanings of polysemous words are likely to be relatively abstract and impoverished underspecified cores relative to the full meanings they are given during processing. As Frisson says, the comprehension of the underspecified core sense of a polysemous word occurs instantaneously but the full meaning is realized only at the end of the sentence if at all. Such delays are likely to be present even when conventional meanings are modulated. This empirical result gives us a way into the problem. Essentially, it is the Semantic and Flow Constraints that will need to be modified.

To make things concrete, consider the following situation u. Batman and Robin are perched above a tall building in Gotham City in the dead of night and they have the following exchange:

> Robin: People were busy today.
> Batman: The city is asleep. ($\varphi = \varphi_1 \varphi_2 \varphi_3 \varphi_4$)[3]

Here, the Setting Game is just one involving the making of commonplace observations. But all four words φ_i in φ cannot be given their conventional meanings because cities do not sleep. Intuitively, its three possible meanings could be *the residents are asleep* or *the city is quiet* or *the residents are quiet*, all of which are modulated contents. It is reasonably clear from u that Batman could have meant any of these three meanings. How might such meanings be derived?

17.1 The Semantic Constraint

As indicated in Section 12.1, there are homonyms like BANK with unrelated conventional meanings and polysemes like EYE or SCHOOL with related conventional meanings, the former taking more time to process than the latter. Indeed, one sense of BANK is itself also polysemous because there can be many different kinds of financial institutions that are all called BANK. These more specific senses can also be conveyed in an utterance. The different related senses of a polyseme can all be thought of as conventionalized modulations, that is, modulations that become conventionalized with use.

Recall the extended Semantic Constraint from Section 11.5,[4]

$$\omega \longrightarrow C^{\omega} \xrightarrow{u} C'^{\omega} \xrightarrow{u} P^{\omega}$$

Any modification of this schema must preserve the possibility of penumbral shift that occurs with vague concepts. Remember that vagueness is ubiquitous

[3] Recanati (2004: 34-36) discusses this example.

[4] See especially page 205.

and it is very likely that it is a vague property (subjective or intersubjective) that undergoes modulation.

Earlier, in Part III, C^{ω} was assumed to be the full conventional meaning of ω even when it was polysemous. Now, based on Frisson's (2009) experimental findings, it can be said to be the common, underspecified core of all the related senses when such related senses exist. For example, in the sense of BANK which is a financial institution, just this core, that is, just *financial institution*, will be designated as one homonymic conventional meaning C_1^{BANK}, where the subscript 1 is used to distinguish this sense from the other homonymic sense of *river bank*.[5] Refinements of this core such as a retail bank or a savings and loan institution or a credit union or a shadow bank or whatever will not be included in this core. This core is a concept in the agent's head. So are the refinements all concepts in the agent's head.

If BANK is used in an utterance with this first sense rather than *river bank* then the shifted concept $C_1'^{\text{BANK}}$ may be such that $C_1'^{\text{BANK}} = C_1^{\text{BANK}}$, that is, no penumbral shift may occur. There is likely to be a great deal of indeterminacy here as it may not be clear whether a shift occurs owing to, say, a different set of exemplars being taken into account for the vague concept *financial institution* or whether its meaning is being modulated. But let us suspend the possibility of a penumbral shift for now and take $C_1'^{\text{BANK}} = C_1^{\text{BANK}}$.

The last step in the extended Semantic Constraint is from $C_1'^{\text{BANK}}$ to P_1^{BANK}, the corresponding subjective or intersubjective property that is a *possible* constituent of the proposition conveyed. There will be as many such properties as there are conventionalized refinements of the core concept available either in the agent's head or intersubjectively. For example, the concept *financial institution* may get refined to the property *shadow bank* in some particular situation. The "null" possibility of *no* refinement, that is, just the property *financial institution* can also occur because in many utterances this core is all that may be intended. Which one of these is selected will depend on the Flow Constraint. This game-theoretic process of handling polysemy is identical to that of handling homonymy except that there are usually many more polysemous senses.

Let us take it, then, that in cases of both homonymy and polysemy, there is a possibly underspecified core concept that is the conventional meaning, which may or may not undergo penumbral shift, that is then transformed by the referential map into a possibly refined subjective or intersubjective property. There will be as many such possibilities as there are conventional meanings and their refinements.[6] These possibilities then enter the Flow Constraint where they get disam-

[5]There are several other conventional meanings of BANK but I will deal with just these two.

[6]This actually depends on the kind of word being considered. There are different things that happen when, for example, the word is a quantifier like THE or EVERY as discussed in Parikh (2010: Chapter 6), but I am not concerned with such words here.

biguated by interdependent partial information games as discussed in Part III. All I have done so far is to change the conventional meaning from its full meaning to an underspecified core in the case of polysemy. The rest of the process is the same as before.

To understand modulation, return to the example of Batman and Robin and focus, for the moment, on the word φ_2 = CITY. How might its sense get modulated to *residents*? Assume that the latter content is *not* a conventionalized refinement of the core concept even though it may well be. If it were, it would just be handled as described above. But, for the sake of the argument, assume it is not.

The locutionary Semantic Constraint applies as follows:

$$\text{CITY} \longrightarrow C^{\text{CITY}} \xrightarrow{u} C'^{\text{CITY}} \xrightarrow{u} P^{\text{CITY}}$$

I now make three assumptions to keep things simple. First, assume that CITY has just one conventional meaning. Second, assume that it is not polysemous so that C^{CITY} is fully specified rather than being just a common core. (The reason for tackling polysemy above is that modulation can apply to either a fully specified or partially specified concept so the analysis must apply to both cases.) Third, assume there is no penumbral shift. Notice that *city* is a vague concept because it just stands for a "large" town so penumbral shifts are certainly possible, but in the current utterance situation u, they can be ignored. In such a situation, then, the property P^{CITY} will be just the usual vague property of being a city. This is one *possible* content of the word that will enter the Flow Constraint.

That is, the (locutionary) Flow Constraint would be activated with the unmodulated locutionary meaning P^{CITY} as one interpretation. I will look at this presently. Before I do so, I want to look at what enables the modulation of P^{CITY} to *residents*.

In order to get to *residents*, it becomes necessary to avail of the *illocutionary* Semantic Constraint involving the sub-Constraints of Relevance and Distance. Recall that Robin had earlier uttered, "People were busy today." This means the concept *residents* is derivationally *near* the baseline concept C^{CITY} (actually C'^{CITY}, but the two are the same here) and the former is also *relevant*. Why can this be asserted? It is because Batman's goal, which can be assumed to be cooperative, is to respond to Robin's observation about the people or residents of the city. So, the property *residents*[7] can be *derived* from the corresponding property P^{CITY} within the open ball $\text{Ball}_u^{\mathcal{B}}(P^{\text{CITY}}, \epsilon_{d,u}) = \{x \in \mathcal{I} \mid d_u^{\mathcal{B}}(P^{\text{CITY}}, x) < \epsilon_{d,u}\}$. Such synecdochic derivations from whole to part are just routinely available reasoning strategies. For example, the agent may first observe that Robin has just talked

[7]We deal with the corresponding properties rather than concepts as there is no guarantee that the final destination infon will be in the agent's head from the outset. As he reasons, he is led to discover the result and this result may lead him outside his head, that is, outside his initial memory.

about the residents being busy. Then he may suppose that if Batman were also talking about the residents the conversational goal of discussing more or less the same topic would be fulfilled. And, so, partly by abduction, he would conclude that Batman is in fact talking about the residents. The length of such a derivational chain might be just two or three steps. Moreover, *residents* is also relevant because it has positive value for the interlocutors on account of adding information to the dialogue. So *residents* is both relevant and sufficiently close and could be admitted as a *possibility*.

But why would the local search for this content be triggered at all? Why would the agent not be satisfied with just P^{CITY} which is available from the locutionary Semantic Constraint? This is where the back and forth between the Semantic and Flow Constraints comes in. In other words, the missing first step in the short derivation above is that P^{CITY} is not adequate, as I will demonstrate below. That is why the agent has to search for something that is derivationally related to P^{CITY} within the ball. This something will turn out to be *residents* because of the rest of the derivation suggested above.

Before we get to the back and forth, I want to make a couple of observations.

First, a city has many parts and has many aspects, not just its residents. So why is *residents* near but not some other related aspect of a city? Why is only *residents* within the open ball? This is because it is the only aspect of a city that would, in *u*, contribute to the conversational goal as made clear by the abductive step above. In some other setting, a speaker may wish to refer to a city's buildings as that meaning may contribute to the goal there. And so on. So only *residents* is found as a possibility after a search is triggered by the inadequacy of P^{CITY}.

Second, consider φ_4 = ASLEEP. Again, a similar argument can be employed to show that the property *quiet* or *inactive* is close to the baseline property *asleep* because these properties contrast with Robin's use of the word BUSY. These meanings also have relevance because Batman could be seen as drawing a contrast between the residents being quiet or inactive at night and busy or active during the day, and thereby contributing to his dialogue with Robin. Intuitively, the physical city's being quiet or inactive does not seem to fulfill any "direct" conversational goal but it could also be a possibility if Robin's earlier utterance implied the physical city was noisy and full of movement during the day. In such loose conversations especially, just like the sentence BILL RAN in Part III, relevance or the value of information is itself measured quite liberally.

I will therefore assume that the illocutionary Semantic Constraint can, if triggered, make meanings such as *residents* and *quiet* available as possibilities for the Flow Constraint. Thus, both the locutionary and illocutionary Semantic Constraint may play a role in generating the possibilities, the latter only when triggered. When this happens, the resulting meaning may no longer be clearly classifiable as locutionary or illocutionary, reinforcing the fuzziness of all these categories that attempt to partition meaning.

17.2 The Flow Constraint

As I said above, the locutionary Semantic Constraint is activated upon encountering φ_2 = CITY and it results in P^{CITY}, which is just the vague subjective or intersubjective property of being a city. This assumes there was just one conventional meaning C^{CITY} and there is no polysemy and no penumbral shift so that $C^{\text{CITY}} = C'^{\text{CITY}}$. This generates the game in Figure 17.1 where $\sigma_2 = P^{\text{CITY}}$.

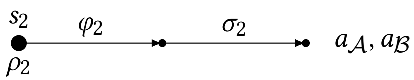


Figure 17.1: The locutionary partial information game for P^{CITY}

Likewise, the locutionary Semantic Constraint operates on φ_4 = ASLEEP and results in P^{ASLEEP} and generates the game in Figure 17.2 where $\sigma_4 = P^{\text{ASLEEP}}$.

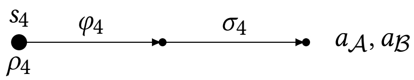


Figure 17.2: The locutionary partial information game for P^{ASLEEP}

Recall from Part III that σ_4 is a conditioning variable in ρ_2 and σ_2 is a conditioning variable in ρ_4. (I am ignoring the other conditioning variables that arise from the games corresponding to the words φ_1 = THE and φ_3 = IS as they do not play any significant role in this example.[8]) The prior probability ρ_2 is the probability that $\mathcal{A}$ = Batman is conveying $\sigma_2 = P^{\text{CITY}}$ *given* that he is conveying $\sigma_4 = P^{\text{ASLEEP}}$ (and the other semantic and syntactic contents). Similarly, the prior probability ρ_4 is the probability that $\mathcal{A}$ is conveying $\sigma_4 = P^{\text{ASLEEP}}$ *given* that he is conveying $\sigma_2 = P^{\text{CITY}}$ (and the other semantic and syntactic contents). But, obviously, $\mathcal{A}$ could not possibly be conveying either content given that he is conveying the

[8]For more on such games relating to THE and IS, see Parikh (2010: Chapter 6).

other(s) because cities and sleep are incompatible and even a partially rational agent would not so contradict himself.

This mutual incompatibility of $\sigma_2 = P^{\text{CITY}}$ and $\sigma_4 = P^{\text{ASLEEP}}$ is what provides the desired trigger that activates the local search for a modulated meaning for either or both contents. Observe that I have used Grice's and Searle's idea that there should be a trigger for modulation based on some kind of inadequacy together with the idea that this inadequacy should be detected without computing the entire locutionary content as they believed. In the mechanism I have described, the mutual contradictoriness of $\sigma_2 = P^{\text{CITY}}$ and $\sigma_4 = P^{\text{ASLEEP}}$ is realized locally, without computing the full locutionary proposition.

Some trigger is required because, otherwise, agents would be trying to modulate every word they encountered and would take unconscionably long to process every utterance. This is where Sperber & Wilson (2008) and Wilson (2011) err because in their system, vaguely specified as it is, there is no trigger and what they call "mutual adjustments" occur ubiquitously even when an unmodulated meaning is perfectly acceptable. The term "mutual adjustment" is fine as an informal description but a *theory* needs to spell out what this back and forth consists in. That is precisely what Equilibrium Linguistics offers. In addition, this agrees with Frisson's (2009) empirical assertion that the core conventional meaning is instantaneously accessed but the full, refined – or modulated – meaning is computed at the end of the sentence if at all.

Because the interpretive process as it occurs within the speaker's Generation Game and the addressee's Interpretation Game runs up against this block of contradictoriness, it backtracks to the Semantic Constraint from the Flow Constraint. This time the illocutionary Semantic Constraint is activated and the Distance and Relevance criterion is put to use. Now that P^{CITY} and P^{ASLEEP} have been found to be mutually inadequate, two local searches ensue based on derivational proximity and relevance relative to the goals of the agents as inferred from the Setting Game. The result of the two searches is the *discovery* of the property *residents* starting from P^{CITY} and *quiet* or *inactive* starting from P^{ASLEEP}. Thus, these two properties become new possibilities as the computation returns to the Flow Constraint from the Semantic Constraint.

In the new games that emerge, P^{CITY} and P^{ASLEEP} are still viable contents because they are only mutually contradictory and so cannot *both* be present in the equilibrium content. Either of them singly may well persist. The next round of games is shown in Figures 17.3 and 17.4 with $\sigma_2' =$ *residents* and $\sigma_4' =$ *quiet* or $\sigma_4' =$ *inactive*. (Figure 17.3 and Figure 9.3 look the same but stand for completely different things.)

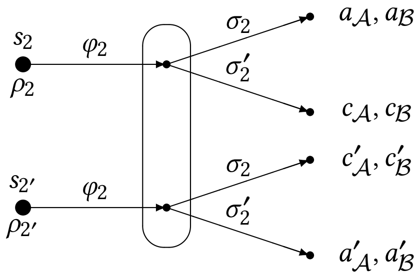


Figure 17.3: The partial information game for the modulation of P^{CITY}

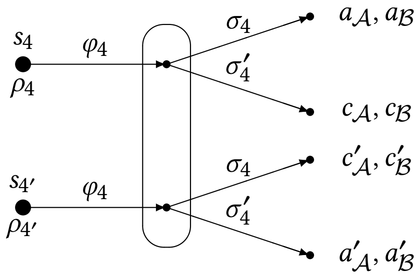


Figure 17.4: The partial information game for the modulation of P^{ASLEEP}

Interestingly, these games are neither wholly locutionary or wholly illocutionary because one content in each is locutionary and the other is illocutionary. By now the reader is likely an expert in solving such games and it should be easy to see that the joint solution can be either $(\sigma_2', \sigma_4) = (\textit{residents}, P^{\text{ASLEEP}})$ or $(\sigma_2, \sigma_4') = (P^{\text{CITY}}, \textit{quiet})$ or $(\sigma_2', \sigma_4') = (\textit{residents}, \textit{quiet})$. It cannot be $(\sigma_2, \sigma_4) = (P^{\text{CITY}}, P^{\text{ASLEEP}})$ as the prior probabilities will not allow it.[9]

So we started with the locutionary Semantic Constraint and then went to the locutionary Flow Constraint. The latter involved a mutually contradictory pair of contents P^{CITY} and P^{ASLEEP} and so we backtracked to the illocutionary Semantic Constraint, found possible modulations of these two meanings through local searches of the infon space, and then returned to a mixed locutionary and illocutionary Flow Constraint where the games could be solved with mutually compatible solutions. This is the back and forth or mutual adjustment between the

[9]See Section 8.1 to see how such games are solved. I have dropped the locutionary contents σ_1 and σ_3 of THE and IS for convenience. Also, I have dropped the meaning *inactive*, although it should be evident that such indeterminacies among possible modulations will be rife.

Semantic and Flow Constraints mentioned earlier. All three possible meanings that I referred to as meanings Batman could have conveyed emerged naturally as solutions. If desired, the equilibria can be augmented with probabilities to show that each of the solutions has some range of probability based on u and on how close and relevant the modulations are. The equilibrium content of the whole utterance φ = THE CITY IS ASLEEP is then partly locutionary and partly illocutionary because the words THE and IS have locutionary equilibria and one or both of CITY and ASLEEP have illocutionary contents.

Just to make this process crystal clear, I set it out more graphically:

Utterance Situation
⟶ Individual words of sentence become available
⟶ Locutionary Semantic Constraint activated
⟶ Possible referential contents become available
⟶ Locutionary Flow Constraint activated
⟶ Mutual incompatibility among locutionary contents detected
⟶ Need for modulation triggered
⟶ Illocutionary Semantic Constraint activated
⟶ Possible modulated contents found and made available
⟶ Mixed Flow Constraint with locutionary and illocutionary contents activated
⟶ Mixed equilibrium contents

What is the range of possibilities for modulations of locutionary meanings? Consider the example of metaphor from Sperber & Wilson (2008: 27) where a woman says to an uncouth suitor, "Keep your paws off me." The authors claim that "the ad hoc concepts constructed to carry these implications will then at least overlap with the concepts encoded by the utterance (otherwise we would be dealing with purely associationist rather than inferential relations). Since the concepts PAW and HAND have disjoint extensions, we claim that 'paw' could not be used to convey the meaning HAND." But this is patently false and shows that their conception of inferential relation is too narrow. Indeed, as I discussed in the case of HE IS A FINE FRIEND in Chapter 14, human reasoning is sophisticated enough to *revise* and *reject* one or more of its premises. In their example, it is quite clear that the utterance has the rough meaning *keep your clumsy, grasping hands off me* and the locutionary starting point *paw* has been discarded. It is true, as Sperber and Wilson claim, that the exact meaning conveyed may not be paraphrasable but this is because the *number* of properties constituting the modulated meaning is indefinite and because these properties are also vague, making it hard to reproduce the same effects with a paraphrase.

Returning to Cohen's example of THE STONE LION and his concern about how which word dominates which is to be determined, consider the examples of THE LARGE MOUSE and THE LARGE ELEPHANT. In the latter two phrases, it is clear that the adjective is modulated whereas in the former the noun is modulated. So one can say that it is unlikely that any syntactic determination of what is modulated is at play. In THE STONE LION, in most situations, no modulation of *stone* will be found in the set of possible modulations $\{x \in \mathrm{Ball}_u(stone, \epsilon_{d,u}) \mid \mathrm{Relevance}_u(x) > \epsilon_{R,u}\}$. That is, this set will be empty. But *lion* will get modulated successfully to something like *representation of a lion*. The opposite will happen with THE LARGE MOUSE in most contexts and *large* will be modulated, not *mouse*. But, given the highly situated nature of language, one should always expect that some rare context will invert these possible modulations.

17.3 Recapitulation

I have now shown in detail how modulation works. When it occurs, unlike free enrichment and implicature, the partial information games in the Flow Constraint are part of the locutionary global game $LG_u(\varphi)$ even though they contain illocutionary modulated contents uncovered by the illocutionary Semantic Constraint. As such, they interact with the games corresponding to all the other semantic, syntactic, and phonetic possible contents thrown up by the sentence uttered as described in Part III. The only reason modulation was analyzed in this Part is that we needed the illocutionary Semantic Constraint.

Earlier, I discussed polysemy and its associated conventional meaning which is an underspecified core that gets shifted and refined if required. In THE CITY IS ASLEEP, this issue did not come up explicitly as I assumed no polysemy was involved. But, in practice, all three kinds of lexical meaning – homonymy, polysemy, and vagueness – can occur simultaneously, as evinced by a word like BANK, which has two homonymic meanings, of which one is polysemous, and of which both could also be vague. So, in attempting a full account of the mechanism of modulation, it is necessary to ensure that none of these possibilities are neglected. When there is polysemy and vagueness, the underspecified core C^{ω} gets shifted to C'^{ω} and it is the latter concept from which a corresponding subjective or intersubjective property P^{ω} is formed. This property then provides the starting point for a local search for a modulated meaning if such a need arises.

Additionally, as Frisson (2009) says, the comprehension of the underspecified core sense of a polysemous word occurs instantaneously but the full meaning is realized only at the end of the sentence if at all. Modulations behave similarly to polysemous refinements in this respect except that they are not conventionalized but arrived at through inferential search. Indeed, polysemy can be viewed as conventionalized modulation as I remarked earlier. If it turns out that Frisson's result cannot be so extended to modulation, then a revision of the theory I have presented might be required. But Equilibrium Linguistics is a framework which allows multiple theories to be developed within it and I have developed just one option.

I have also shown how all three classical aspects of lexical meaning – homonymy, polysemy, and vagueness – would be handled by Equilibrium Linguistics.[10] The first and third of these were analyzed in Part III and polysemy required just a small modification of the conventional meaning. Modulation, what might be called the fourth aspect though it applies not just to words but also to phrases, builds on the first three.

Figures of speech like metaphor, metonymy, synecdoche, hyperbole, and others can all be seen as instances of modulation. For example, an utterance of THE PEN IS MIGHTIER THAN THE SWORD would result in a modulation of both nouns. In this case, the whole sentence would have to be evaluated locutionarily as there is no direct incompatibility between *pen* and *sword*. As discussed in Section 10.4, the truth of the utterance may also play a role in this determination. The meaning of this particular aphorism is now obviously conventionalized and so agents do not actually need to infer its meaning any more. Other tropes might result from implicature as happens with some examples of irony. So by giving an account of all these types of indirect meaning, I have also provided accounts of such phenomena.

More complex examples such as THE CITY WHICH IS POLLUTED IS UNHAPPY can also be tackled. Here, the word POLLUTED is about the city itself and the word UNHAPPY possibly makes a reference to its residents. I leave it to the reader to apply the theory I have described to this example.

As always, we have to contend with indeterminacy. I have already alluded to the indeterminate modulation of P^{ASLEEP} to either *quiet* or *inactive* or both as well as to other possible properties that may not be easy to make verbally explicit. All such modulations will lie within the set of possible modulations

[10] See Murphy (2010: Chapter 5) for a discussion of these three aspects of lexical meaning. Also, look back to Footnote 9 in Section 3.1 to see how CAN YOU PASS THE SALT? would be handled as an instance of polysemy or conventionalized modulation.

$\{x \in \mathrm{Ball}_u(P^{\text{ASLEEP}}, \epsilon_{d,u}) \mid \mathrm{Relevance}_u(x) > \epsilon_{R,u}\}$ though an agent may not always be able to articulate them. The other indeterminacy, also discussed earlier, is between modulation on the one hand and free enrichment and implicature on the other. As the example of the garage being open shows, one agent may handle it as a modulation, another as an implicature.

A psycholinguistic issue is whether all these indirect meanings are stored in the agent's memory or lie outside and are discovered afresh. I have tried to emphasize that the process of local search will sometimes lead an agent outside his head, that is to say, it will lead him to entertain an entirely new idea that he has never contemplated before. This is also presumably how creativity occurs and an understanding of modulation-like processes may lead to deep insights into creativity.

A computational issue is how the complex and flexible human reasoning involved in search can be modeled. The usual tack is to model some subset of the infon space and some subset of possible reasoning strategies and thereby succeed with a subset of indirect meanings. It seems unlikely that the full range of possibilities can be mapped although, with machine learning techniques, one may be able to annotate training data sufficiently to allow such algorithms to work fairly well. In any case, I have at least shown how, in principle, the realm of illocutionary or indirect meaning can be brought within the ambit of science via a model that is philosophically sound, mathematically solid, computationally tractable, and empirically adequate.

Having made these observations, I now point out perhaps the most important one. Does the account of the *fixed point principle* which generalizes Fregean compositionality and the context principle as discussed in Section 10.2 need to be altered in any way to accommodate modulation? As modulation involves just the locutionary global game $LG_u(\varphi)$, and as it therefore involves interdependence among the various local partial information games and among all the possible semantic, syntactic, and phonetic contents of the sentence uttered, no change needs to be made to it. All the results of Section 8.1 continue to apply including Theorem 8.5 and Equation 6 from Section 8.1.5 in particular.

This completes my discussion of modulation.

18 Overview of illocutionary meaning

18.1 Review of the argument for indirect meaning

I have now described in full detail how the notions of relevance and distance that constitute the illocutionary Semantic Constraint can, together with the illocutionary Flow Constraint, yield the indirect meanings of an utterance. I have also shown how these meanings are generally indeterminate and involve multiple infons with different probabilities, thus justifying the picture of meaning as smeared out on the infon lattice, one region for each interlocutor in the exchange.

Combining this model with the one in Part III for locutionary meaning gives us a way to compute both direct and indirect meanings. In particular, I have discussed how lexical and structural disambiguation and reference-fixing or saturation on the side of locutionary meaning and free enrichment, implicature, and modulation on the side of illocutionary meaning are handled. This also provides insight into how many tropes work so figurative speech also falls within their scope. The one aspect of meaning I have not explicitly considered is direct and indirect illocutionary force, that is, the stating, commanding and requesting, asking, promising, and other forms of doing things with words, to use Austin's felicitous phrase, that accompany every utterance. Force does not pose any special obstacles and can be resolved by the same partial information games as worked with other direct and indirect meanings.

A small adjustment that needs to be made in the model for illocutionary meaning is that I have assumed that the baseline from which distances are measured is either the locutionary meaning or one of its proper parts. But when all types of indirect meanings are evaluated together, the baseline may change to, for example, a modulated meaning or an enriched meaning or both. I will show one example of this in the next section. In any case, such changes to the model are straightforward to make.

One unfinished matter mentioned in Chapter 15 is whether informational proximity implies sufficient relevance. Because the account of distance is difficult to formalize it may be hard to give a proof of this. But one might argue as follows. First, it could be noted that some information τ is relevant just in case it fulfills the shared conversational goal because that is what leads to an increased *value* for the agent. If τ is already adequately near the locutionary meaning (or its proper part), then it (or information derivable from it or information it is derived from) must abductively or otherwise fulfill the shared conversational goal within the open ball. Thus, proximity implies fulfillment of goals. And fulfillment of goals implies increased value of information, which is relevance. So relevance is superfluous to the derivation of indirect meaning. I am not sure if this informal argument can stand up to greater scrutiny but if it does it would greatly simplify the account as then only distances within the open ball need to be calculated to find illocutionary meanings. That would also emasculate the problems raised by intensionality as well as other difficulties for the notion of relevance. And it would sever ties completely with the Gricean maxims while retaining the Cooperative Principle and also the Gricean reasoning involved in the analysis of particular examples.

In the Gricean scheme it is the maxims that provide a "standard" against which literal meaning is measured. If it fails, an implicature is triggered to restore adequacy. In my account it is the goals of the agents as they emerge in the Setting Game that provide the standard against which locutionary meaning is measured. When such goals are not met, illocutionary meanings are triggered to restore adequacy. So the maxims can be seen in a somewhat new light, as approximations to goals or as high-level goals that can be *overridden* by the more specific goals, as I argued in Section 5.2.[1] When viewed this way, it becomes clear why maxim-based explanations of implicature are often unsatisfactory and vague. It is because maxims can only partially account for the breadth and variety of goals that interlocutors can have and it may often seem that we are force-fitting them into one or another maxim. Horn (2014) offers a very rich set of examples to show how maxims such as being orderly are, in fact, realized in many different ways in different subcommunities and sets of situations, which helps to particularize otherwise "universal" maxims, thereby bridging the gap somewhat between abstract and concrete goals.

Another issue is to what degree the requirements of common knowledge can be said to be met in computing illocutionary meaning. This is a delicate matter and in theoretical work it is best to just assume it. But its lack is often responsible for all kinds of miscommunication and partial communication as captured by the overlapping regions of the infon space representing each agent's content. So when our interest lies in such aspects of communication it may become necessary to more carefully specify the nature of the shared knowledge as I have discussed

[1] Earlier, in Parikh (2001: Section 7.7), I had seen them as approximations to rationality which seems a bit unspecific relative to goals.

in *The Use of Language* (2001: Chapters 5 and 6) in some detail. There, I also tackled other flows of information, such as *suggesting*, where full common knowledge does not obtain.

I remind the reader how thoroughly pragmatic and practical my model for both locutionary and illocutionary meaning is as opposed to the more epistemic inclinations of most other researchers as mentioned in Section 5.1. The goals and preferences of the agents play a crucial role in all the derivations. This is completely missing in Relevance Theory and is relatively absent in Grice as well despite his talk about purposes and despite some of his actual analyses. As most others have followed Grice, they also share the same limitations.

18.2 A complete example

I now deal relatively briefly with a more or less complete example bringing especially the Content Selection Game back into focus. This will help connect what we have accomplished in Parts III and IV.

Suppose a seven-year-old child is playing quietly by himself in a room. His activity may partly be described by his ongoing actions and partly by choices he may be considering. His mother is working on her book on communication, *Deconstructing Derrida*, in the same room. Her activity may likewise be described by her thoughts and possible actions like writing. Now, the child cuts himself on his finger very slightly and a drop of blood appears. He begins to cry loudly. The mother sees it is a very minor cut and, as she is absorbed in her thoughts, she wants him to stop crying while she interrupts her work to attend to the cut. This forms the context or the background situation u for the communication that transpires between them. The Setting Game SG_u incorporates the partly convergent and partly divergent interests of the two agents. From all the examples I have looked at so far it should be apparent that such situations can be extremely varied. What is remarkable is that situation theory and situated game and decision theory allow us to represent this wide range of possibilities quite compactly.

Informally, what happens next is that she utters the sentence, “The cut is not serious.” And the child hears her and stops crying. How is this exchange to be analyzed and understood?

Based on SG_u the mother $\mathcal{A}$ wants to stop the child $\mathcal{B}$'s crying. She could bring this about in a number of ways. One way is to directly convey to the child that he should stop crying. This includes both the locutionary content to stop crying and the corresponding illocutionary force of a “command” or firm request. Another way is to convey a more circuitous content that the cut is not serious and *so* the child should stop crying. There are many other possibilities depending partly even on the creativity of the mother.[2] If telepathy were possible, the mother would convey these direct or indirect contents *im*mediately, that is, without the mediation of language. But before she can do any conveying, either via telepathy or via language, she has to select what to convey, either the direct or indirect content or some other content altogether.

Such contents do not come ready-made into $\mathcal{A}$'s mind. They have to be individuated based on her *goal* to get $\mathcal{B}$ to stop crying because the cut is not serious. As I said in Section 13.2, the contents possess what I call *act-equivalence* since they have the same purpose. This is a kind of intensionality of choice and action as equivalent contents may induce different responses.

On the one hand, the human brain is extraordinarily fast, but on the other, we are not usually aware of such content selection when we speak, let alone all the other decisions that are involved. What actually happens in the body is not known, but we can try to identify the broad structure of constraints – the *philosophical* psychology – that must be respected in a *rough* way. For example, I am presenting these constraints as a problem of *choice*. It is not necessary that the choices be made in the detailed way I am about to describe, but the constraints predict that choice of certain kinds must be present.[3]

The Content Selection Game that results is shown in Figure 18.1.

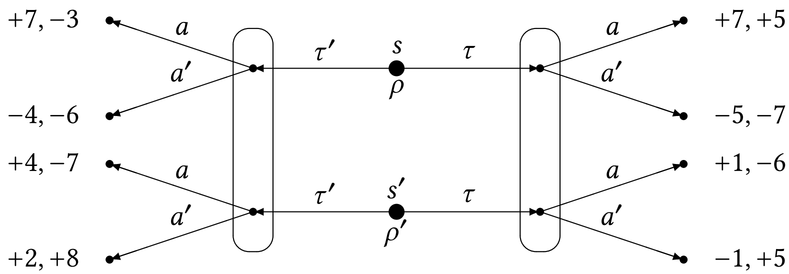


Figure 18.1: The mother and child Content Selection Game

[2]For example, she could try to distract the child.

[3]See Glimcher (2004) for some preliminary evidence that supports this view.

The utterance situation u contains a great deal of information, some of which is shared and some of which is not. For example, the fact that the child is crying loudly is shared, in fact, it is common knowledge between them in the nonwell-founded way described earlier. But the child isn't sure whether the mother has the mental space to fuss over him or merely attend to him and return forthwith to her work. This uncertainty on the child's part may be represented by a set of two possible situations, s itself where she will not fuss and s' where she will fuss. Both these situations are a part of u. As far as he is concerned, he could be either in s or in s', the two situations being identical except for this single uncertainty about fussing. The mother knows that she cannot spare the time to fuss and so knows that she is in s. But she is also aware that the child does not know if she will fuss or not, that is, if the situation that obtains is s' or s. So she is forced to consider s' even though she knows it is not a factual situation. Because enough of the foregoing is shared between them, these two possibilities become common knowledge between them. Since there is no further information about the situations between them, it is common knowledge between them that the probabilities ρ and ρ' of these initial situations s and s' being factual are the same.[4]

I will restrict myself to just the two contents explicitly considered above, the indirect content labeled τ and the direct content labeled τ'. These are the two contents the mother could convey in both situations and they are represented by four branches emanating from them. At this stage, there are different possibilities. We can assume the child is aware of these possible contents, though the argument for that in this setting is somewhat thin. Or we can assume that only the mother has these possibilities in mind and chooses between them and so only she represents the full strategic situation between them. Different settings will dictate different local assumptions: both can be accommodated within the choice situation I am developing.

In response to conveying τ or τ', the child can either stop crying – the action represented by a – or continue crying – the action represented by a'. And once both mother and child have acted, certain consequences ensue for each of them that are most conveniently represented by utilities. The particular numbers have been chosen to reflect their relative preferences for these consequences in the

[4] As I also pointed out in Section 11.5, there is a certain subtlety involved here because the mother actually knows that s is factual and so that $\rho = 1$. One could say either that the game is played prior to her knowing this or, what I prefer, that they have common knowledge of the child's belief of equiprobability. That is, the mother adopts the child's ignorance for the sake of the situation.

setting. These preferences are dictated by the complexities of the background situation which involve the following salient features:

1. The shared goal of getting the child to stop crying by attending to him

2. The partly convergent, partly divergent personal subgoals of the participants because the mother does not want to fuss and the child wants the mother to fuss

3. The effort of conveying and interpreting a content

4. The psychological fact that if the child has a reason to stop crying, he will be more inclined to do so. Such a reason presumably functions like an Aristotelian enthymeme.

5. The desire of the mother to reassure her child and to reinforce her parental bond, that is, their relationship

The first three of these have been amply examined so far but the last two have been looked at less thoroughly. The latter two facts involve a moral dimension as mentioned in Taylor (1985/1999) and in several places in Parts I, II, and III. Such moral facts, whether of a major kind or a minor kind, are often present in interactions involving communication and this normative dimension of language forms its second fundamental aspect, the first being its acquisition of meaning in the first place. But the two are intertwined as is evident from this example and from Chapter 11.

For example, if the mother chooses to convey τ in s and if the child chooses a, then the mother gets a payoff of 7 and the child gets 5. These payoffs include both costs and benefits. If the mother chooses the direct content τ' in s, and if the child once again chooses to stop crying, then they get 7 and −3, respectively. Here, the mother's net payoff stays roughly the same owing to two opposing pressures: the lower effort of conveying τ' reduces the cost but the ignoring of the third and fourth factors raises the cost, with the result that the payoff remains roughly the same. On the other hand, the child's reward plummets as he is not offered any basis for his action and is not reassured by his mother's self-centered response. Remember that the situation s involves the mother's not wanting to fuss over the child and it is in this context that the child makes his evaluations. The other payoffs can be analyzed similarly keeping in mind that s' involves the alternative situation where the mother is inclined to fuss. The key thing to note is that the payoffs reflect both cooperation and conflict and also costs and benefits.

By now, having seen the examples in Section 8.6 and Chapter 11, it should be clear how quite complex psychological, social, anthropological, and moral features of the utterance situation become part of the Communication Game. The idea is that the *structural* essences of these attributes are *abstracted* from the concrete situation and inserted into the abstract game-theoretic model, that is all. This is what would allow the so-called *thick* aspects of the utterance situation and even the larger environment, the focus of so much inquiry in the social sciences, to be accommodated by the framework of Equilibrium Linguistics.

It turns out that the solution to the choice situation in Figure 18.1 is that the mother should choose to convey τ in s, which is the situation that matters since it is factual. But now the mother, having solved the content selection problem, has to figure out how to convey the indirect content that the cut is not serious and so the child should stop crying. For this, she turns to language as one modality among others. So far, the child knows nothing of what is brewing except perhaps that she has noticed his crying.

As we know, this figuring out of how to convey a content via language is also a choice situation called the Generation Game. In a sense, I got ahead of myself because I expressed τ in English. Needless to say, τ has to have some more abstract representation in the brain. And the mother's task is to convert this abstract representation into an English sentence. Part of this task is what to make explicit and what to leave implicit, that is, what words to utter and what to convey as an illocutionary content (e.g. implicature). For example, the mother could say any of the following:

- The cut is not serious so please stop crying
- The cut is not serious
- The cut is small
- It's a small cut

⋮

⋮

We are not usually conscious of how such choices get made, but the brain nevertheless does have to make them. The second, third, and fourth options, for example, say something about the cut but leave the child to figure out the implicature that he should stop crying *as a result.* As mentioned in Section 8.4, there are objective (e.g. length, complexity) and subjective (e.g. style, maintaining or altering relationships, avoiding errors) aspects of cost that determine the optimal sentence and what is made explicit and left implicit. In addition, it is not just the sentence that has to be selected but also how it should be produced, with what intonation and so on. Once the best sentence is uttered the Interpretation Game emerges.

Overall, the explanation for the mother's utterance of "The cut is not serious" and the child's stopping crying runs as follows. First, there is the setting u within which everything takes place and an associated Setting Game. Next, the mother plays the induced Content Selection Game and selects the indirect content τ over τ' as that is the equilibrium of that game. This is so because their preferences were based on the fact the child would be more likely to respond as desired if given a reason. Without it, he is unlikely to be persuaded not to mention the reassurance he receives. After this selection, the mother chooses a sentence φ ("The cut is not serious," possibly) by playing the Generation Game. Then this sentence φ is uttered and the child plays the Interpretation Game. He infers the full meaning of φ (i.e. the literal content + implicature) which is τ and then plays the Content Selection Game where he chooses the appropriate response to τ, namely, a, and so stops crying. In inferring τ the methods and results of both Parts III and IV have to be used.

Thus, the mother (i.e. the speaker) starts with the Content Selection Game and then plays the Generation Game. The child (i.e. the addressee) starts with the Interpretation Game and ends with the Content Selection Game which could also be called the Response Game from his point of view. This double set of choices that both agents have to make – selecting a content and then the utterance for the speaker and choosing an interpretation and then an action for the addressee – are quite reasonable once one realizes that one choice in each set of choices would have been required even if telepathy were possible. Both players then return to the Setting Game.

As we have been seeing throughout, this is more or less the full structure of communication and it constitutes what I have called the Communication Game.

A more complex example that builds on this analysis can also be considered. Assume u is the same as before but now there is a slight edge in the mother's response to the child's crying, maybe because she is concentrating on a particularly challenging point in her writing. She might then utter a slightly sharper sentence like "You are not going to die." This would result from a somewhat different but related Content Selection Game where the content τ'' is that the child will not die *from the cut* and so he should stop crying because the cut is not serious. This time there is not only a locutionary content and an implicature but also a completion *from the cut* and the saturation of the pronoun YOU. Indeed, as mentioned in Chapter 16, the implicature is discovered starting from the baseline of the enriched meaning, not the locutionary meaning. The payoffs in the Content Selection Game would also be a little different as they would account for the chance the mother gets to express her annoyance at the interruption. The reader should now be able to supply the details for such an utterance.

Small variations of this kind in the exact choice of content and sentence will always be possible in a situation and this makes the process of content selection and generation a fundamentally creative process to some degree even when the goals from the Setting Game and *u* are fully specified. This is what makes it difficult to fully spell out the structure of the Communication Game. But I have shown that it is possible to go quite far in this process nevertheless.

If I had to find one line to summarize the ground traversed, it would have to be that the influence of logicism and truth-conditional approaches to meaning more generally have kept the field from appreciating a fuller idea of agency and therefore of communication and language and perception and metaphysics. In a sense, the field has more to do with rhetoric than logic, and game theory and situation theory provide the right means to approach it. Utility is perhaps more important than truth in semantics and communication.

19 Beyond illocutionary meaning

An utterance is an extraordinarily complex event. We underestimate and are usually unaware of this complexity because we are able to handle not just one but multiple utterances in quick succession with relative ease. If you tell a layperson about the study of meaning and content, they are generally puzzled by exactly what difficulties the field could possibly pose when a certain level of understanding appears so readily and transparently available. On further thought, they might feel that more intricate texts of the kind found in literature may raise some real questions but then wonder how something as amorphous as content could be explored rigorously, even mathematically.

These are legitimate questions. So far I have been explaining the part of semiosis I have called micro-semantics, something that has been attempted in modern times since the inaugural work of Frege and Russell and in classical times in at least four different traditions.[1] This puzzle of meaning, its relative ease of partial access despite its immense complexity, should now be at least partly resolved. It is helpful at this stage to step back and survey just what the range of phenomena are that have to be explained. A variety of contents are associated with even a single utterance and what these are often remains in an elusive background with just the particular content of interest becoming the focus of inquiry, whether it is literal meaning or implicature or some other type of meaning.

In this chapter, I try to identify just this scope of the content of an utterance. My *geography* of content is theoretically underpinned by Equilibrium Linguistics. Its key feature is that many types of content can be derived in a more or less uniform manner using the tools of situation theory and situated game theory. This computation typically involves two stages within a larger Communication Game: first generating the *possible* interpretations and then choosing one or more among them. Sometimes, we can avoid carrying out these steps separately and linearly by allowing them to interact in a circular process that is more efficient.

Importantly, there are less acknowledged contents that I address now that require somewhat different methods from the ones I have described so far. They go beyond illocutionary meaning. My goal here is less to describe detailed derivations than to identify as much as possible of the entire field of meaning.

[1] See van Bekkum et al. (1997).

To start with, I have shown that syntactic and phonetic contents are to be treated as legitimate contents of an utterance on par with semantic contents such as locutionary and illocutionary meaning. This insight is often obscured because the processes that determine all three are seldom seen to be analogous and interdependent in the way suggested by Equilibrium Linguistics. So this identification depends very much on the underlying theoretical framework that tackles the three problems in a uniform way.

I have also suggested in the past that higher-order implicatures (and weaker flows of information that do not require common knowledge) may exist. As I say in Parikh (2001: Section 7.9):

> In principle, there is nothing to stop us from considering higher-order implicatures. These would arise when a first-order implicature gives rise to a second-order implicature, the second-order implicature to a third-order implicature, and so on. Such higher-order implicatures are more difficult to derive and so are not commonly found in ordinary conversation. They require more explicit calculation and so are more often found in literary texts like novels and poems, where part of the point is not to state things too explicitly, but to leave them to the imagination of the reader ("show, don't tell"). This leaves room for creative readings and misreadings and enriches the text. A reader has to work harder to derive higher-order implicatures, and common knowledge at this relatively rarefied level becomes increasingly more tenuous. As a result, the reader has to bring in his own private background of belief and knowledge, and fill in the text to create a dense world of meanings and associations. As this begins to happen, implicature shades off into suggestion, and we are no longer in the realm of communication and shared understandings, but rather in the fluid world of imagination and other transformations of thought.

When a literary critic says that *Crime and Punishment* is about the absence of god (or a moral order) in the modern world, that is an example of a higher-order meaning or even possibly a higher-order implicature. While the passage above refers to higher-order implicatures that are communicated, there is no reason to suppose that corresponding higher-order meanings that are not common knowledge cannot arise.

Quite ordinary utterances can also give rise to higher-order meanings. In the example of the mother and child in the previous chapter, it is possible for the child to infer not only that the mother is conveying τ in the Content Selection Game but also that the situation is s and not s'. This tells him that the mother is in a situation where she will not fuss over him. Should this inferred content be seen as a second-order implicature? It is not something the mother intends to convey perhaps but she may be aware that the child can infer it and, in fact, the information becomes common knowledge between them. Alternatively, the content can also be seen as so-called *inverse information*, a point I return to below.

I now introduce certain types of contents that are perhaps less recognized and are seldom admitted into the favored circle of the objects of semantical inquiry. This may partly be because Grice largely defined the terms of the discussion and for reasons that are hard to surmise the field has stuck to what he charted out as the scope of semantics and pragmatics. These Gricean terms have involved a more or less exclusive focus on what is *communicated* via an utterance rather than on the information available to be extracted from an utterance. While this restriction makes sense for Grice's own project of naturalizing meaning, it is not the only aspect of meaning that can and should be studied and the goals of the field need to be broadened. To a certain degree, it is continental philosophy that has looked at such meanings, although with no clear understanding of their scope and limits or of how they arise.

19.1 Significance

I have argued that there is already a fair bit conveyed by an utterance as direct and indirect contents. The following question now arises. If σ is a locutionary or illocutionary meaning of φ uttered in u, and if $\sigma \Rightarrow_{\ell} \sigma'$ as defined in Section 2.1, should σ' also count as part of what is conveyed, whether explicitly or implicitly? If, for example, the content *I see the murderer* "logically" implies the content *I see a person*, should the latter also be counted as part of what is conveyed by the utterance? To repeat the quote from Thomason (1990: p. 349) from Chapter 15:

> All this shows, I think, is that 'Did you mean to say' is used to query people about the consequences of what they have said, without distinguishing these clearly from what they said. More generally, 'Did you mean to do' does not distinguish between intentions and foreseeable side effects, which is exactly the distinction we need to make here. To achieve some progress in these negative cases, it seems that we will need to say more about the intensionality of intentions in general, and of intentions to assert in particular, and also to find tests that extract more reliable conclusions from the evidence. The philosophical background, of course, suggests that it won't be easy to make progress on these matters.

As Thomason asks, how can we distinguish between what is conveyed and its consequences? My answer is that it may not always be possible to distinguish between intended and unintended consequences and so, depending on our interests, such consequences and others could be included in what is conveyed.

Just as I raised a question about $\Rightarrow_\ell$, if σ and σ' are both part of the content, should $\sigma \wedge \sigma'$ also be part of the content? The response should again be positive even if an agent does not quite realize this and the meet of the two infons remains implicit.

Based on this, we can say how this more capacious content should be characterized. If $\mathcal{C}_u(\varphi)$ is the combined direct and indirect content of an utterance, then the (logical) *significance* of the utterance is the smallest filter containing $\mathcal{C}_u(\varphi)$.[2] Unfortunately, things are not so simple because there are probabilities present in $\mathcal{C}_u(\varphi)$ and I need to first specify how these should be handled.

We can tentatively define $(\sigma, \pi) \wedge (\sigma', \pi')$ to be $(\sigma \wedge \sigma', \min(\pi, \pi'))$ and stipulate that if $\sigma \Rightarrow_\ell \sigma'$ and $\pi' \leq \pi$ then it follows that $(\sigma, \pi) \Rightarrow_\ell (\sigma', \pi')$. To simplify things, we could just set $\pi' = \pi$ for the latter. With this understanding, the definition of logical significance is complete.

Recall that if $s \vDash \sigma$ and $\sigma \Rightarrow_\ell \tau$ then $s \vDash \tau$ and also that $s \vDash \sigma \wedge \tau$ if and only if $s \vDash \sigma$ and $s \vDash \tau$. These facts tell us that propositions can naturally be broadened to include significances as well.

Intuitively, it makes sense to include all the consequences of the direct and indirect meanings even though agents may not explicitly realize many of them. In an ongoing conversation, such significances can often be taken for granted as part of the understanding among interlocutors. Significance should therefore also be treated as part of what I have called the *referential* meaning of an utterance in addition to direct and indirect contents. If one refers to a murderer, one has also referred to a person and this property is therefore also part of the meaning even if an agent doesn't quite work it out. All the referential meanings dealt with so far can be thought of as *conveyed* by the speaker to the addressee and some subset of this may even be *communicated* in Grice's special sense. In addition, the intentions attaching to such meanings may be either explicit or implicit because it may not be possible for a speaker to foresee all the logical consequences of his utterance. It can happen that a person makes a claim without grasping one or more of its consequences; if some untenable consequence is pointed out to him, he may want to withdraw his claim and this lends some weight to the intuition that such significance be included in the referential meaning of an utterance.

[2]Recall that filters were defined in Section 2.1.

Also, as we have seen in the derivation of illocutionary meaning, situated human reasoning is much broader than this narrower notion $\Rightarrow_\ell$ allows. So the idea of significance can also be broadened to encompass such more flexible inferences (relative to u) as might be involved in getting to what *Crime and Punishment* is about, for example. This kind of content becomes more difficult to readily admit into what is conveyed by the speaker or author as remote consequences of utterances are harder to compute. Some interpretations may be foreseen by speakers, others not. The matter is highly indeterminate. But this is precisely why the characterization of what is conveyed by an utterance is not the only goal semantics should pursue. A broader goal is to characterize all the information extractable from an utterance relative to some reference situation such as u. In interpreting fiction, moreover, the context u is often highly indeterminate and this is what gives rise to multiple interpretations of the same text. As long as a valid argument exists for such a conclusion it has to be admitted as legitimate and the criterion for assessing alternatives shifts to whether an interpretation is insightful or interesting.

Consider the pithy significance of the simple but profound statement by the poet Wallace Stevens that art is an attempt to see the world again with fresh eyes. Its direct and indirect meanings are quite straightforward to determine but its significance is very rich and different recipients will get it to different extents depending on their own depth of experience.

19.2 Associations and extended meaning

It is possible to go even further if we allow mental processes that are less constrained to operate on the referential meanings derived from an utterance. I call such meanings *associations* and they are almost entirely subjective and depend on all kinds of connections that may be triggered by a referential meaning. A lot of creative thought is associative in just this way and associations should also be included in the range of meaning. To be sure, there is no way to reproduce exactly the same associations in the same external circumstances and they are not as "logical" as referential meanings but so what? The logical bias of the field has prevented the recognition of associations as legitimate meanings even though they form a very important part of our experience especially when we read some more complex text like a novel or poem. Much of the richness of literature and other art relies on evoking associations in people however much these may differ from person to person. A mere mention of a beautiful sunset can conjure all kinds of mental images.

Consider the following examples from Fernández & Cairns (2011: 244–245).

(31) a. He was pounding the nail when …

b. He was looking for the nail when …

The first of these is likely to lead an addressee to infer that an instrument like a hammer was being used. This is less likely with the second utterance. Such inferences are stored together with the direct and indirect meaning of the utterance and may not be easily separable from it. In fact, addressees are frequently mistaken about precisely what they have heard because such associations are mixed in with the meaning.

Can associations be inferred from first principles with the same rigor that Equilibrium Linguistics brings to the derivation of direct and indirect meanings? The key thing to realize here is that each association is a connection between one or more referential meanings and a cluster of other infons in an agent's brain generally based on world knowledge. In other words, all associations are a proper subset of an agent's memory which in turn is a proper subset of the infon space $\mathcal{I}$. The second thing to notice is that what gets activated depends to a great extent on how the particular agent's memory is *organized.* If σ is a referential meaning and it involves items that are stored near τ_i, $i = 1, 2, 3, \ldots\ n$, then some or all of the τ_i could be triggered. But different agents will inevitably have different arrangements of the same infons and that is why their associations will, in general, be different. Part of this structure of memory, as the study of neural networks has shown, is the relative strengths of the linkages between infons. As science progresses, there is no reason why we should not be able to probe the memory of an agent, form a model of it, and then *predict* its associations when subjected to an utterance. For this task, situation theory should have an important role to play in understanding mental representation and its role in evoking associations.

Another type of expanded meaning is the implications of referential meanings and associations when combined with the knowledge and beliefs of an agent. If σ is part of the referential meaning (including significance) and if σ' is an associated belief of the agent, then the consequences of putting σ and σ' together is part of what I am calling the *extended* meaning of the utterance. It, too, is largely subjective although, since many beliefs are shared, two or more agents may arrive at some overlap in their extended meanings. We may wish to restrict the set of beliefs taken into account to just the set of activated beliefs or something a little broader such as beliefs activated upon later reflection as well. Extended meanings are roughly the smallest filter containing all referential meanings including significances, all associations, and all considered beliefs *and* all the associations of this filter, and so on. In other words, it is the closure of both operations of association and extension carried out simultaneously.

Associations and extended meanings are generally not part of what is intended by an utterance even if they may be intended to be evoked abstractly by an author, especially in fiction. But they are nevertheless part of the content of the *utterance* and this is part of why I started this chapter by saying that even a single utterance is an extraordinarily complex event. So far, I have been trying to show that the meanings of an utterance can cover a potentially wide and even infinite territory. But this should not mislead one into thinking that this kind of meaning is *always* vast. It can occasionally be fairly simple and quite small.

19.3 Inverse information

Barwise & Perry (1983) introduced the idea of inverse information which is the information you can infer from so-to-speak *external* or, better, *inverse* aspects of an utterance. For example, the accent of a speaker may reveal his Indian origins or a statement may reveal a speaker's location (e.g. if he says what time it is). This kind of information can be quite subtle depending on the circumstances and on the actual sentence uttered especially if it is combined with public knowledge or private knowledge. Occasionally, such information may qualify as both part of the (intended) indirect content as well as part of inverse information. This can happen if, for example, a speaker emphasizes his accent or other characteristic in some way available to him in order to actively *convey* something about his background. Certain presuppositions of an utterance may also sometimes belong more to inverse information than to the referential meaning. Since it is well known that the United States has a president then saying something about the president would carry a presupposition that such a person exists but since it is taken for granted anyway by the interlocutors, this information may best be seen as being part of the inverse information of the utterance.

I want to now briefly look at a particular kind of inverse information that Barwise and Perry did not bring to light in the context of the mother and child example. In Figure 18.1 in the previous chapter, the mother elects to convey the content τ to the child. After she utters an appropriate sentence and he figures out the intended interpretation, the child is able to form a model of the Content Selection Game. He solves it and chooses the action a to stop crying. But he also learns that s is the factual situation and not s'. By digging into s he infers that the mother does not wish to fuss as she is busy. This is the reason *why* she is conveying τ. This kind of answer could be called inverse information as it is generally not intended and so is not a second-order implicature. Such inverse information is ubiquitous, in fact, because we often try to infer why an agent did something.

This is a more complete description of how inverse information of certain kinds might arise. Inverse information can be treated as part of the content of an utterance even though it may not actually be communicated or conveyed to the addressee.

19.4 Latent meaning

There is the old joke of the three psychoanalysts stepping into an elevator one morning with the first saying "Good morning" to the other two, and the second asking the third, "What do you think he meant?" Freud and his followers revel in such hidden and repressed meanings and literary and art critics (e.g. Barthes 1957; Bordwell 1989; Eagleton 2008) have made this kind of meaning their special preserve although Bordwell is "analytic" in his approach. These are meanings that get disclosed even *despite* their author. For example, the mother's not wanting to fuss could conceivably also be a latent meaning.

Latent meanings are often derived in the context of a set of extraneous beliefs somewhat akin to what I called associations and extended meanings earlier. Indeed, the two categories may overlap. When critics engage with a text they may bring a theoretical framework to bear on it and this framework would then combine with the narrower contents of the text to allow such meanings to emerge. One reason why Continentals and critics focus on these meanings is because the more mundane referential meanings are relatively straightforward for human beings to discern: it is no feat to ascertain that the mother is asking the child to stop crying because the cut is not serious. Of course, *theoretically*, it *is* an accomplishment to show how such a meaning can be derived from first principles and this has occupied researchers for millennia. But critics and Continental philosophers are not interested in this challenge and often also employ frameworks that eschew referential meanings altogether (Eagleton 2008). On the other hand, semanticists have exhibited their own insularities and have failed to appreciate these richer aspects of meaning. What is required is to marry the rigorous methods of semantics with the broader scope of meaning I have been delineating. Otherwise all one gets is the excesses of Continental philosophy and the utter silence of Anglo-American philosophy on dimensions of meaning that are of central importance.

19.5 Discourse meaning

I have been alluding indirectly to discourse meaning when I talked about the meanings of texts but here is a specific example that goes beyond the meaning of a single utterance. In *Getting Even*, Woody Allen says, "Can we actually 'know' the universe? My God, it's hard enough finding your way around in Chinatown." The humor in this works partly because of the contrast between the scale of the universe and the scale of Chinatown and partly because of the bathetic fall from the loftiness of epistemological inquiry to merely getting around. But this contrast and bathos are not directly present in either sentence taken by itself. It is their *juxtaposition* as part of a single discourse that gives rise to such discourse meanings.

It is possible to develop a rigorous model of discourse meaning based on Equilibrium Linguistics by creating a sequence of Communication Games and linking them appropriately. The challenge would be to understand the kinds of relations (e.g. contrast, bathos, etc.) that can occur among the meanings of single utterances in the discourse. These relations would form part of the discourse meaning – so-called *inter-utterance meaning* – that transcends single-utterance meaning.

In addition, just as the content of a word depends on the contents of other words in an utterance, so the content of an utterance depends on the contents of other utterances in a discourse. Perhaps the most commonly investigated area of discourse meaning is anaphoric reference,[3] and there are many subtleties even in this seemingly simple problem.

19.6 Emotive meaning

All the different types of meaning discussed above have a propositional form. But there is often a non-propositional emotive content conveyed in an utterance. If a speaker says, "Man! George W. Bush was elected again!" then either a positive or negative emotion may be conveyed depending on the utterance situation and especially on the identity of the speaker. McCready (2012) provides an interesting game-theoretic analysis of how such meanings are communicated. However, the particular formulation he chooses based on signaling games appears too complex to be psychologically plausible. In addition, the nonmonotonic logic he uses to determine whether the emotion is positive or negative makes no mention of conversational goals.

[3] See Clark & Parikh (2007) and the references therein.

A simpler way to see emotive meanings is via my account of modulation in Chapter 17. For the word MAN in the utterance above, there would be an underspecified conventional core which would be modulated by the illocutionary Semantic Constraint consisting of Distance and possibly Relevance either to a positive or negative content depending on the situation and especially the goals and preferences of the interlocutors. Such an analysis would also connect the intended emotive meaning with the other locutionary meanings of the utterance through the prior probabilities in the partial information games of the Flow Constraint. This kind of connection and mutual reinforcement in equilibrium by the contents of other words in the utterance is also missing in McCready's paper. The degree and tone of the emotion may also be inferred via the relevant Communication Game, partly by taking the Phonetic Constraint into account.

Such meanings can be locutionary or illocutionary and may also be extracted from an utterance without being intended. I discuss the related notion of effect in the next chapter.

I have now described phonetic and syntactic contents which might be called *linguistic* meanings, and earlier I described conventional meanings and referential meanings, the latter of which include direct and indirect meanings as well as significances, and finally I have described associations, extended meanings, inverse information, and latent meanings. As just mentioned, there can also be emotive meanings. *All* of these together may be described as comprising *utterance meaning*. Discourse meaning goes beyond single utterance meaning. Some of these meanings are communicated, some merely intended, either explicitly or implicitly, and the rest simply occur privately in the mind of each agent depending on what *other* information is applied to the problem.

I may have missed some elements of utterance meaning but it should be possible to insert such items into the framework I have sketched. The key fact about this entire range of meaning is that it is accompanied by a high degree of indeterminacy which implies that each component is often probabilistic and may differ from speaker to addressee, and it may not be possible to draw sharp boundaries between different varieties of meaning. And yet, as I have tried to argue, a fair amount of order and rigor can be brought to the task. Also, just because there can be so much to a single utterance does not imply that meaning is always vast and complex. For many routine utterances it may be relatively simple.

Gricean communication is too restrictive a focus of inquiry. The flow of information where one studies what may be intended explicitly or implicitly (e.g. Hirsch 1967; 1978; some of the essays in Iseminger 1992) is more permissive but even this is too conservative in the end. The field of meaning ought to cover *all* the information extractable from an utterance by an agent. When such a program is carried out, semantics begins to connect with the kinds of meaning studied in much literary and art criticism and the hermeneutic sciences. I believe the sketch above shows how such a link may be forged that may help bring greater rigor to the humanities broadly conceived. It also recovers its connection with the history of semantics as can be seen from van Bekkum et al. (1997).

Grice's project of naturalizing meaning is nevertheless necessary to preserve, as without it the entire realm of meaning would remain ungrounded. The hermeneutic sciences miss this element in their account. It is only by combining the two sides, intended flows of meaning and extracted meanings that are unintended, that the split in the study of meaning can be healed and the discipline made whole.

However, as I have argued in this book, Grice's own approach to his project is unfortunately flawed. I see Equilibrium Semantics as a more comprehensive and demonstrably superior approach to meaning that remains grounded in communication and therefore in the Gricean project.

20 Classifying meaning

I have now laid out more or less the entire field of meaning. Picturing this takes a few diagrams. The first basic division is of utterance meaning into intended meaning and extracted but unintended meaning. Intentions may be explicit or implicit. This is shown in Figure 20.1.

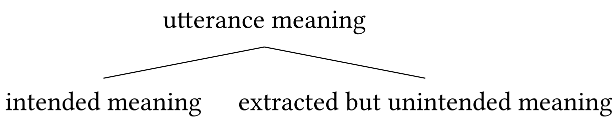


Figure 20.1: Classification of utterance meaning

I could have added a third branch showing syntactic and phonetic contents. Dividing the left branch further, we get the tree in Figure 20.2.

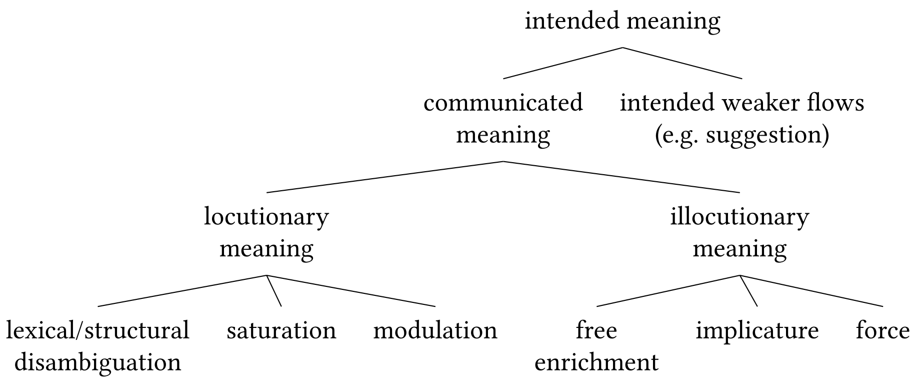


Figure 20.2: Classification of intended meaning

I have placed modulation under locutionary meaning even though it would in fact straddle locutionary and illocutionary meaning. Implicature includes higher-order implicatures as well. Many if not all figures of speech fall under modulation or implicature.

The category of intended weaker flows represents transfers of information that do not require full common knowledge. These were studied in some detail in my previous two books and mentioned in Section 7.4. There is an entire infinite lattice of flows of information with communicative flows involving common knowledge being the strongest. When something is suggested, for example, we get a flow that is weaker than communication. Much art and literature involve suggestion and similar weaker flows.

I have touched upon Grice's distinction between literal meaning (or *what is said*) and implicature now and then and this is just a different way to carve up direct and indirect contents. Literal meaning includes locutionary meaning and enrichments. The direct/indirect distinction has more to do with *how* these contents are derived and Grice's distinction has more to do with *what* contents we lump together psychologically. There has been a great deal of discussion about the latter division, especially by Recanati (2004), and he tries to argue that literal meanings are in some sense *available* to our conscious perception in a way that locutionary or direct meanings are not because the latter have to be teased apart from modulations and enrichments after the fact. So an alternative way to classify intended meaning is shown in Figure 20.3.

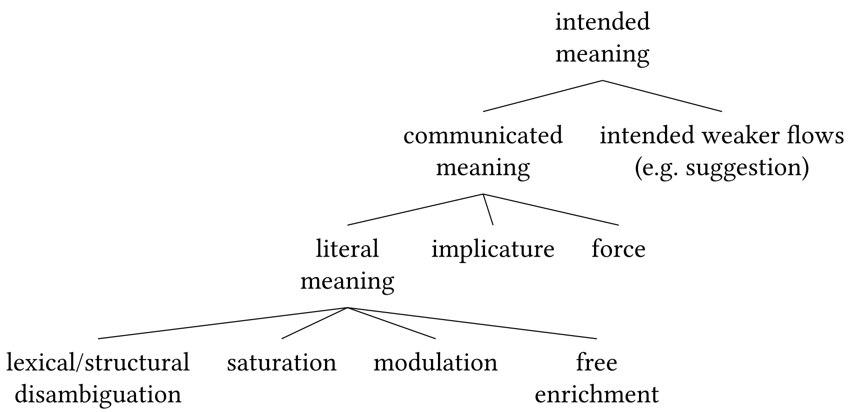


Figure 20.3: Alternative classification of intended meaning

I have emphasized throughout that many of these categories are not separated by strict black-and-white divisions but rather that it is often indeterminate where a particular content is placed by an agent. Such a classification is useful nonetheless because it tells one what mechanisms might have produced the content or whether it was consciously available to the agent or not.

The category of extracted but unintended meaning can be divided as shown in Figure 20.4.

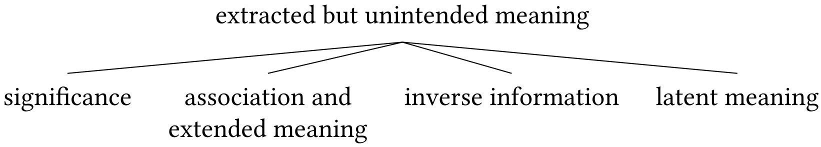


Figure 20.4: Classification of extracted but unintended meaning

Recall that some parts of significance, especially logical significance, are intended so again these classes are not watertight. The same is true of associations, extended meanings, and inverse information. Latent meanings, by definition, cannot be intended. I have tried to show in the previous chapter how the two basic aspects of utterance meaning, intended and unintended meaning, fit together into a natural whole. If we wish, we can draw a final diagram for discourse meaning shown in Figure 20.5. The category of inter-utterance meanings that transcend individual utterance meanings was discussed in the previous chapter.

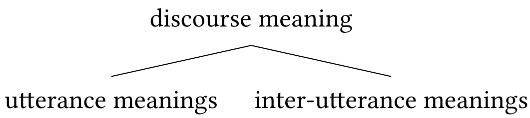


Figure 20.5: Classification of discourse meaning

As a test case, consider Stalnaker's (2006) discussion of the following remark of US Treasury Secretary John Snow in May 2003: "When the dollar is at a lower level it helps exports, and I think exports are getting stronger as a result." The announcement caused the dollar to drop precipitously in value even though only a commonplace economic insight had been stated. What meaning issuing from this utterance created this effect? There might be more than one plausible way to analyze this example but it seems to me that it belongs to possibly unintended inverse information of the same kind as the inverse information about the mother's not wanting to fuss.

At the very start of the book, I said communication involves the relaying of content and has an *effect* and that while the former is cognitive, the latter can involve the whole person. I have so far not said much about this effect although it is related to the emotive meaning of an utterance discussed earlier. Effects accompany most utterances though they are not part of their meaning. Happy or sad news may affect the addressee differently, possibly even creating palpable bodily effects. These may be intended or unintended. Much communication is colored in this way even if there are no actual linguistic markers of the emotion transmitted. When understanding an utterance, it is very important to capture the textures of such effects as well. These are nonpropositional but nevertheless part of semantics. This is possibly where it is best to resort to concrete descriptive techniques if such detail is required but, as I said in Section 12.2, sentiment analysis in natural language processing deals with this phenomenon as a classification problem. I have already discussed in that section how Equilibrium Linguistics can be extended to such classification tasks so I will not say more about it here.

Finally, I repeat my definition of communication, Definition 5.3, from Section 5.3, as we have come full circle. Recall that communication is the right notion to reduce and not speaker meaning as Grice believed. Also keep in mind that the proposition p contains probabilistic contents as described in Section 15.2.

Definition 20.1. $\mathcal{A}$ communicates p to $\mathcal{B}$ by uttering φ in u if and only if $\mathcal{A}$ intends (possibly partly implicitly) to convey p to $\mathcal{B}$ in u, $\mathcal{B}$ intends (possibly partly implicitly) to interpret $\mathcal{A}$'s utterance of φ in u, and the games $G_u^{\mathcal{A}}(\varphi)$, $G_u^{\mathcal{B}}(\varphi)$, $G_u(\varphi)$ induced thereby are equal and common knowledge and their solution is p.

I have now developed the full details of the global games in the definition as well as of the larger picture of micro-semantics sketched in Chapter 3. This should make it possible to see how the definition naturalizes meaning *modulo* the category of conventional meaning which has so far been taken as externally given. In Part V on macro-semantics, I provide an internal account of conventional meaning after which the naturalization of meaning will be fully realized assuming the framework I have described is correct.

Before I turn to that task, we take a brief detour to see what Equilibrium Linguistics has to offer the problem of translation.

21 Translation

My purpose here is to highlight just one contribution my framework can make to the task of translation. Translation requires not only a grasp and command of the source and target languages but also a deep understanding of the text to be translated. Jurafsky & Martin (2009: Chapter 25) describe a computational approach to the problem.

Different languages individuate the world in different ways. And different languages affect addressees of discourses in different ways. These simple facts have a profound consequence: perfect translations do not exist. This compels us to seek approximate translations and to view translation as a process of approximation. Moreover, languages differ in several other ways as well, in the information they make explicit (e.g. definiteness and number information), in their syntax, and so on. In this section, I describe in broad outline how Equilibrium Linguistics can bring precision to the idea of translation as approximation.

The basic idea is very simple. Let the source language be $\mathcal{L}$ and the target language be $\mathcal{L}'$. Consider an utterance of $\varphi \in \mathcal{L}$ in u. Then $\varphi' \in \mathcal{L}'$ relative to u' will be a translation of φ in u if their meanings are sufficiently close to each other. Suppose σ is the (locutionary and illocutionary) meaning of φ in u and σ' is the (locutionary and illocutionary) meaning of φ' in u'. That is, $\mathcal{C}_u(\varphi) = \sigma$ and $\mathcal{C}_{u'}(\varphi') = \sigma'$. (Alternatively, it is possible to let σ and σ' be the intended meanings of the respective utterances that include weaker flows that are not common knowledge.) Then σ and σ' must be sufficiently near each other. How should nearness be measured? There are two possible ways, each with somewhat different properties. One is to use the idea of distance d_u from Chapter 14 and the other is to use the idea of distance δ_u from Section 2.1. The former is more practical as it takes account of the interlocutors' goals in u[1] and the latter is more logical. I personally think d_u is likely to be more useful than δ_u. Presumably the goals in u' are aligned with the goals in u so it does not matter which situation is used.

[1] This needs some interpretation as σ presumably already meets the goals of the agents as it is both the locutionary and illocutionary content of φ in u, so what does it mean to traverse the infon space from σ to σ'? In this context, it would mean that σ' also fulfills the goals. That is, goal fulfillment is preserved in going from σ to σ'.

Assume we have a translation threshold $\epsilon_{t,u}$. Then we can give the following definition.

Definition 21.1. φ' in u' is a translation of φ in u if and only if $d_u(\sigma, \sigma') < \epsilon_{t,u}$ (alternatively, $\delta_u(\sigma, \sigma') < \epsilon_{t,u}$).

There will generally be multiple translations available and the task would then be to identify the best one according to some criterion. One possibility is to minimize the distance but this may not always yield the ideal outcome.

This still leaves open how an actual translation would be derived as the definition merely characterizes the set of translations. I have shown in detail how to go from φ in u to σ. Then reasoning of one sort or another may be used to go from σ to σ'. But it still remains to go from σ' to φ' in u'. This is the inverse direction of generation rather than interpretation and as the generation problem remains partially solved, so does the problem of translation. Indeed, the task is to find a nearby σ' such that it is the content of some $\varphi' \in \mathcal{L}'$ in u'. The last two steps are thus interdependent and involve a back-and-forth interaction between them.

This definition of translation is a little narrow as it omits any consideration of the intended effects of the source and target texts. As just discussed, such effects involve classification or detailed description and the former approach can be used to broaden the definition to include effects that are also *near* each other.

I believe Part IV gives a more or less complete solution to the problem of illocutionary communication that is, as advertised in Chapter 1, philosophically sound, mathematically solid, reasonably computationally tractable, and empirically adequate. The Setting Game, Content Selection Game, Generation Game, and Interpretation Game that constitute an (illocutionary) Communication Game are developed in detail. So are the two illocutionary Constraints: Semantic and Flow. The universality of locutionary partial information games demonstrated in Part III can be readily extended to illocutionary partial information games. I consider meanings that lie beyond direct and indirect meaning and classify the entire realm of meaning, both intended and unintended. I complete this conceptual geography of meaning by approaching the effects of communication as either a classificatory or descriptive undertaking. I take a brief look at translation. Again, as promised in Chapter 1, I connect arguments in philosophy, linguistics, psychology, and computer science, unifying them through the framework of Equilibrium Linguistics and relating them to the hermeneutic sciences.

In the next Part, I look at Language Games and at how conventional meanings originate and change.

Part V

Language Games

22 Defining and solving Language Games

So far I have assumed that the locutionary Semantic Constraint and conventional meanings in particular were given and available in the computation of the referential meanings of an utterance. This was the natural task of micro-semantics and I have shown how Equilibrium Semantics provides a more or less complete account of it and, indeed, of the whole process of communication in the small. But micro-semantics does not by itself explain how conventional meanings emerge in the first place nor how these meanings change over time. Both the origin and dynamics of conventional meaning must themselves reside in communication as there is no other possible source. Indeed, the Semantic, Syntactic, and Phonetic Constraints that are used in the Generation and Interpretation Games are themselves the products of communication. This is not to deny the initial endowment of capacities people are born with, only to highlight the role communication plays in activating these capacities.

I sketched a skeletal framework for macro-semantics in Section 3.2 and in this chapter I will give it some flesh. As I said earlier, my goal is only to offer a toy model of both origins and change that reveals the essence of these processes. I will abstract from syntax and phonetics altogether and look only at imaginary single-word sentences in studying the origins of conventional meaning. I will also make other simplifying assumptions. Once this core model of a Language Game is developed, it will be relatively easy to see how it might be extended to accommodate more realistic scenarios. I will do the same sort of thing with semantic change, but before I get to it, I will briefly discuss Grice (1968), Lewis (1969), Skyrms (2010), and Tomasello (2010) on origins and compare them with Language Games. A byproduct of this network of Communication Games is that it shows a way to think about convention more broadly and realistically than Lewis (1969) did.

I remind the reader that my approach is analogous to the idea of general equilibrium in economics where one has multiple interacting markets in society. In our situation, there are multiple interlocking conversations in society and they lead to what I call a meaningful equilibrium. I start with a particularly simple Language Game and then progressively generalize it.

22.1 A simple Language Game

Recall that the locutionary Semantic Constraint has the general form $\omega \longrightarrow P \xrightarrow{u} \sigma$ where the first map is called the conventional map and the second the referential map. That is, the locutionary Semantic Constraint consists of the Conventional and Referential sub-Constraints. The property P is the conventional meaning and the infon σ is the referential meaning.[1]

To start with, assume there are just two properties P_1 and P_2 of interest in a community of three agents $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$. These two properties yield the two contents σ_1 and σ_2. That is, the referential map $P_1 \xrightarrow{u} \sigma_1$ and $P_2 \xrightarrow{u} \sigma_2$ is taken as given. In these one-to-one correspondences, u is just a parameter and takes on different values, say, u_1 and u_2, for different utterance situations. For example, P_1 could be the property *apple* and P_2 the property *banana* and σ_1, σ_2 could be a particular apple and banana in the environment relative to the corresponding utterance situation in which the respective property is used by the agents.

Where do these referential maps come from? They are purely ontological and are naturally given as part of the individuation of reality. I had discussed such ontological transforms in *Language and Equilibrium* (2010: Section 2.4) and they need not detain us here. But, just as an example, the *instantiation* map is as follows: $ind(P, s) = \langle\langle\ (\ x \mid s \models \langle\langle\ P;\ x\ \rangle\rangle\ )\ \rangle\rangle = \langle\langle\ a\ \rangle\rangle$ where s is the situation relative to which a property P is instantiated in the individual a.[2] In this map there is an infon embedded inside another infon and it is read as "the individual x such that x has the property P in situation s." The use of the variable x indicates that the result of this operation need not be unique. It is via instantiation that one goes from the property *apple* to a particular apple in some situation. I am taking such referential/ontological maps as given but in general such metaphysical individuation of reality and the emergence of conventional meanings occur side by side in a larger process I have called Equilibrium Metaphysics in the book just cited.

Assuming the Referential sub-Constraint, a language $\mathcal{L}$ is required in order to communicate contents such as σ_1 and σ_2. This is because the properties P_1 and P_2 cannot be used directly as they are abstract. So an agent has to use a property *by* using a word, something that can be physically produced. Assume

[1] I have deliberately used P instead of the concept C to keep things simple.

[2] See Section 2.1.

$\mathcal{L}$ is made up of just two single-word sentences ω_1 and ω_2.[3] Initially, at time $t = 0$ as it were, neither word is conventionally attached to either property. There is no conventional map such as, for example, $\omega_1 \longrightarrow P_1$ and $\omega_2 \longrightarrow P_2$ or $\omega_1 \longrightarrow P_2$ and $\omega_2 \longrightarrow P_1$. Assuming this pairing emerges in a one-to-one way, these are the only four correspondences possible.

In this very simple scenario, there are exactly two words, two conventional meanings, and two referential meanings. Why are both conventional and referential meanings required? Why isn't it possible to go directly from word to object? This was discussed in my previous book and in Section 1.2 and will be fully addressed in Chapter 26, but the short answer is twofold: both are required to avoid Frege's (1892/1980) problem of informative identities such as "Hesperus is Phosphorus" and to account for the full range of meaning, especially as encountered in the case of noun phrases involving names, descriptions, and generalized quantifiers.

For semiosis to occur, one of the two conventional maps above must result from the communicative exchanges in the community, either $\omega_1 \longrightarrow P_1$ and $\omega_2 \longrightarrow P_2$ or $\omega_1 \longrightarrow P_2$ and $\omega_2 \longrightarrow P_1$. Suppose $\mathcal{A}_1$ talks to $\mathcal{A}_2$ and to $\mathcal{A}_3$ via two separate utterances. The notation I have used so far for a Communication Game is Γ_u. This was convenient when there were just two interlocutors $\mathcal{A}$ and $\mathcal{B}$ and just one utterance situation u being considered. Now there are multiple agents and multiple utterance situations so it is better to drop the parameter u, keeping in mind that all communication always occurs inside a particular utterance situation, and to add superscripts for the two agents. So the Communication Game between $\mathcal{A}_1$ and $\mathcal{A}_2$ would be called Γ^{12} or Γ^{21}, both being the same game, the first symbol representing the game from $\mathcal{A}_1$'s point of view and the second symbol representing the game from $\mathcal{A}_2$'s point of view. Similarly, the Communication Game between $\mathcal{A}_1$ and $\mathcal{A}_3$ would be $\Gamma^{13} = \Gamma^{31}$. For our purposes it does not matter who the speaker is and who the addressee is. Both agents in both Communication Games can assume either role. This situation can be represented visually as shown in Figure 22.1.

The figure depicts a *connected* graph with the vertices labeled by the three agents and the edges labeled by the corresponding Communication Games.[4] Such an object consisting of a network of agents and Communication Games

[3]For those puzzled by such single-word sentences, think of Wittgenstein's (1953/1968) "Slab!" or Quine's (1960) "Lo! A rabbit."

[4]A graph here is a set of vertices representing the agents in the community together with undirected edges linking them which represent Communication Games between them. The edges are undirected because it does not matter who the speaker is and who the addressee is. A connected graph is one in which there is a path from any vertex to any other vertex.

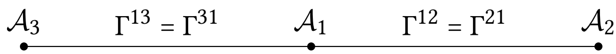


Figure 22.1: A simple Language Game Γ

between them is a *Language Game*. Its symbol is just Γ (or $\Gamma(n)$ for n agents). Importantly, Language Games are almost never common knowledge; only their component Communication Games may be. This fact will figure in Chapter 23 when we discuss convention.

In general, there will be many more agents and many more Communication Games as well as many more words, conventional meanings, and referential meanings. When conventional meanings emerge through such a Language Game, the community of agents will become a *linguistic* community. It is possible in principle to envisage a larger graph encompassing the entire population of the planet that comprises many linguistic communities with some agents being members of more than one such community. Such a graph may not be connected. As populations move about, coming closer or drawing apart either spatially or culturally, the links among them will change and so will their languages.

As words do not have conventional meanings initially, the Communication Games in Figure 22.1 will have a special form which I now specify. Consider $\Gamma^{12} = (SG^{12}, CSG^{12}, GG^{12}, UG^{12})$, the four component games. Here the superscripts can be ordered as 12 or 21 as both pairs of corresponding games are the same. When considering a Communication Game as part of a wider Language Game, it is convenient to abstract from the details of the various Setting Games and Content Selection Games. Alternatively, we can just assume that SG^{12} induces the trivial CSG^{12} shown in Figure 22.2.

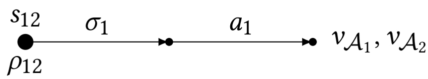


Figure 22.2: Content Selection Game CSG^{12}

I have assumed that $\mathcal{A}_1$ is the content selector and, later, speaker but it could easily have been $\mathcal{A}_2$. And I have assumed that the content he wishes to convey is σ_1 rather than σ_2 just to keep the indices aligned. The upshot of this is that $\mathcal{A}_1$ chooses to convey σ_1 to $\mathcal{A}_2$ in order to get her to do some action a_1. This could easily have been assumed at the outset but I want to emphasize that the whole Communication Game with all its component games is involved in the Language Game.

Now we come to GG^{12}, the Generation Game. Since $\mathcal{A}_1$ has chosen to convey σ_1 – that is his only option in the trivial CSG^{12} – he has also implicitly selected P_1 in virtue of the Referential sub-Constraint $P_1 \xrightarrow{u} \sigma_1$. At this stage, neither ω_1 nor ω_2 are attached to either conventional meaning P_1 or P_2 so $\mathcal{A}_1$ has to consider both words as possible carriers of the intended meaning. Keep in mind that if, say, ω_1 had been conventionally connected to P_1 rather than ω_2, the shape of GG^{12} would also have been trivial, essentially ω_1 as his choice followed by σ_1 as her choice, mediated by P_1 via an already existing locutionary Semantic Constraint. But part of the difficulty the agents face at the dawn of (linguistic) semiosis is that they have to not only choose a linguistic expression and a corresponding referential meaning via a conventional meaning, they also have to connect the linguistic expression with a conventional meaning in the first place. This leads to a slightly more complicated game of partial information as shown in Figure 22.3.[5]

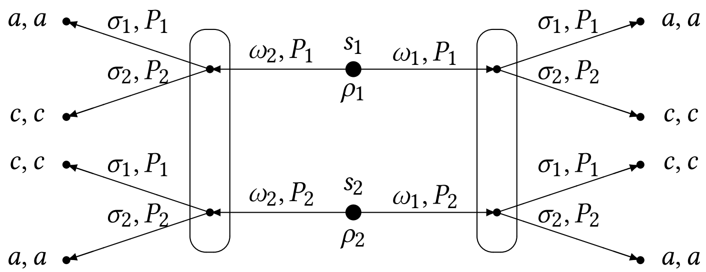


Figure 22.3: Partial information game $g^{12} = g^{21}$

In this game, $\mathcal{A}_1$ wants to convey σ_1 by using P_1 in the situation s_1. He not only makes a choice between ω_1 and ω_2 but also chooses P_1 in s_1, thereby connecting one of the words with P_1. And $\mathcal{A}_2$ has to choose not only between σ_1 and σ_2 but also between P_1 and P_2 because she does not know whether he is conveying P_1 or P_2. This compels $\mathcal{A}_1$ to consider an alternative situation s_2 in which he has to choose between attaching ω_1 and ω_2 to P_2 and $\mathcal{A}_2$ has the same choices as before.

[5]The shape of the game tree is identical to the shapes of the Content Selection Games in Figure 11.1 in Section 11.5 and in Figure 18.1 in Section 18.2. The latter games, however, are completely different as they involve the selection of contents and responses and the former involves the selection of utterances and meanings.

Also, since σ_1 and P_1 are linked and σ_2 and P_2 are linked via the referential map, $\mathcal{A}_2$'s choice is really between (σ_1, P_1) and (σ_2, P_2). The payoffs have been made symmetric to keep things simple and we have $a > c$ because (σ_1, P_1) is the correct interpretation in s_1 and (σ_2, P_2) is the correct interpretation in s_2.

There is no question of anything else in GG^{12} as there are no other linguistic expressions available that might lead to alternative locutionary games. And there are no other games of partial information either, because the utterance consists of just one word, either ω_1 or ω_2. It could be said that $g^{12} = g^{21}$ is a semantic game and there must be a corresponding syntactic game for both words and this is true but I will omit it for now as we are dealing just with single words. So GG^{12} contains just g^{12} and nothing else.[6]

It should be obvious that the Interpretation Game UG^{12} also contains just g^{12} and nothing else. So we have completed the description of Γ^{12}.

There are two Nash equilibria in g^{12}, one corresponding to $\mathcal{A}_1$ choosing (ω_1, P_1) in s_1 and (ω_2, P_2) in s_2 and $\mathcal{A}_2$ choosing (σ_1, P_1) at the (upper) intermediate node on the right side of the figure and (σ_2, P_2) at the (lower) intermediate node on the left side of the figure, and the other corresponding to $\mathcal{A}_1$ choosing (ω_2, P_1) in s_1 and (ω_1, P_2) in s_2 and $\mathcal{A}_2$ choosing (σ_1, P_1) at the (upper) intermediate node on the left side of the figure and (σ_2, P_2) at the (lower) intermediate node on the right side of the figure. In plain English, $\mathcal{A}_1$ can optimally attach either word to P_1 and the other word to P_2 and $\mathcal{A}_2$ must always choose based on whether P_1 is being conveyed or P_2. Because the payoffs are symmetrically chosen, both equilibria are equivalent. In other words, either ω_1 or ω_2 could attach to P_1 as one would expect in the absence of any differential costs of the two words. Note that since s_1 is actual and s_2 is counterfactual, only the relevant part of either equilibrium corresponding to s_1 will be played.

This completes my discussion of Γ^{12} which brings us to $\Gamma^{13} = (SG^{13}, CSG^{13}, GG^{13}, UG^{13})$, the Communication Game between $\mathcal{A}_1$ and $\mathcal{A}_3$. In this game, too, it can be said that SG^{13} induces a trivial CSG^{13} which is similar to CSG^{12} except that σ_1 is replaced by σ_2 and a_1 by some other action a_2. It is not necessary that the two Content Selection Games differ in this way as $\mathcal{A}_1$ may have the same exchange with both the other agents. But in that case the conventional meanings corresponding to the language $\mathcal{L}$ consisting of the two words would

[6] The form of GG^{12} is actually similar to the form of the Generation Games we saw in Figure 7.8 and Sections 7.3 and 8.4. The speaker has a choice between ω_1 and ω_2 because both can express σ_1 via P_1 and both words lead to the same game g^{12}. The difference occurs in the possible different costs of ω_1 and ω_2 and therefore in the net values received by the speaker with each choice. But I have deliberately abstracted from such complications because they obscure the essence of Language Games.

only be partially learned, which is an important case I will return to later. For now, assume that the exchanges involve different contents so that the whole language may be learned.

CSG^{13} then leads to the same partial information game shown in Figure 22.3 that is also a part of GG^{13} and UG^{13} but this time the situation s_2 is actual and not s_1 so the two partial information games are not, strictly speaking, identical. It is called g^{13} or g^{31}. Again, there are two symmetric Nash equilibria as before but only the parts corresponding to s_2 rather than s_1 get played.

I had said earlier that it does not matter which agent is the speaker and which the addressee in these two Communication Games and I have taken $\mathcal{A}_1$ to be the speaker in both for now. I will soon account for the other possibilities.

If Γ^{12} and Γ^{13} and in particular g^{12} and g^{13} were played completely independently of each other, then the same word could end up attaching to both conventional meanings, P_1 and P_2. But the agents are at least partly rational and this suffices for $\mathcal{A}_1$ who is involved in both sets of games making his choices so as not to overload any word as that would be suboptimal and would lead to inefficiencies in future exchanges. So if he chooses to connect ω_1 with P_1, then he simultaneously chooses to connect ω_2 with P_2, and vice versa. In this way, he makes the best use of the linguistic resources available to him. This is nothing but the Consistency Condition across utterances mentioned in Section 3.2 in its simplest form. This condition is not something externally or arbitrarily imposed on agents, it is just a product of rationality because future ambiguity incurs an avoidable processing cost.

Thus, the way $\mathcal{A}_1$ solves g^{12} and g^{13} and, therefore, Γ^{12} and Γ^{13} is by choosing *complementary* Nash equilibria in these games in keeping with the Consistency Condition. Out of the four possibilities $\{(\omega_1, P_1), (\omega_2, P_2)\}$, $\{(\omega_1, P_2), (\omega_2, P_1)\}$, $\{(\omega_1, P_1), (\omega_1, P_2)\}$, and $\{(\omega_2, P_1), (\omega_2, P_2)\}$, the last two are eliminated. This still leaves two equilibria on the speaker's side. On the addressees' side, it is clear that the first game yields (σ_1, P_1) and the second game yields (σ_2, P_2).

How does the speaker choose between the two Nash equilibria that still remain? It is only when ω_1 and ω_2 have identical or inversely proportional costs and expected frequencies of use that the two equilibria will be equivalent. In all other cases, there will be a natural bias in payoffs that will favor one equilibrium over another.[7] For the relatively rare cases when they are equivalent, there are three ways to break the symmetry.

One purely internal way is to introduce some dynamics into the space of strategies as, for example, Skyrms (2010) does. This is generally unsatisfactory for reasons discussed in Section 22.3.

[7] It is worth referring back to footnote 4 in Section 8.1 at this stage.

A second is to say that there is a short process of trial and error and feedback through which $\mathcal{A}_2$ learns $\mathcal{A}_1$'s choice of equilibrium which is determined by the flip of a fair coin. That is, the Communication Game Γ^{12} is repeated until $\mathcal{A}_2$ gets it right in a few steps. Certainly this must happen in some cases. For example, if $\mathcal{A}_1$ is conveying something about a particular apple rather than a particular banana but $\mathcal{A}_2$ thinks the opposite, then she will fail in her understanding and therefore in her response to $\mathcal{A}_1$ in CSG^{12}. But this failure is an opportunity to learn the right equilibrium and therefore the intended meaning of the utterance.

The third possibility is to look historically at what must actually have happened in many situations. Gestures preceded verbal language and it is likely that early humans both pointed at objects in their visual environment and uttered accompanying grunts. In other words, language could build on gesture and early attempts at verbal communication were probably overdetermined: the addressee could infer the correct verbal response from what the speaker was pointing at. Thus, much learning must also have happened in this one-shot way, instantaneously. Of course, pointing and other gestures also involve solving games but here context and salience play some role in eliminating most possibilities. There is also an element of naturalism as opposed to conventionalism in gestural language, just as there is with visual language, that is absent in verbal language. So semiosis uses quasi-naturalistic symbols to ascend to purely conventional symbols.

Let us assume that $\mathcal{A}_1$ chooses $\{(\omega_1, P_1), (\omega_2, P_2)\}$ rather than $\{(\omega_1, P_2), (\omega_2, P_1)\}$. And we already know that $\mathcal{A}_2$ chooses (σ_1, P_1) in g^{12} and $\mathcal{A}_3$ chooses (σ_2, P_2) in g^{13}. Note that the presence of the two conventional meanings in the addressees' choices tells one that $\mathcal{A}_2$ has associated ω_1 with P_1 and $\mathcal{A}_3$ has associated ω_2 with P_2.

This implies that the entire Language Game Γ has been solved with two results. One is that the words ω_1 and ω_2 acquire the conventional meanings P_1 and P_2, respectively. The other is that the individual communications get across the referential meanings σ_1 and σ_2. So there are both private and public successes. The so-called meaningful equilibrium of Γ can be formally expressed as $\{[(\omega_1, P_1), (\sigma_1, P_1)]; [(\omega_2, P_2), (\sigma_2, P_2)]\}$.

Now, $\mathcal{A}_1$ has participated in both games and so he knows the whole meaningful equilibrium. But $\mathcal{A}_2$ has learned only that $\omega_1 \longrightarrow P_1$ and $\mathcal{A}_3$ has learned only that $\omega_2 \longrightarrow P_2$. Each of the addressees has encountered only part of the language and thus shares in only part of the success. In the special case where both addressees know there are only two words and conventional meanings in the language, they could infer the other connection as a possibility but in more realistic situations this cannot be done. They will learn only that part of the language that they experience. This is quite reasonable. If someone hasn't come across the word SESQUIPEDALIAN, they may not know its conventional meaning and be unable to use it.

This more or less winds up the description of Γ and how it is solved via the Consistency Condition being added to the Nash criterion. It results in the Conventional sub-Constraint being established and through it the locutionary Semantic Constraint because the Referential sub-Constraint is given.

One matter that remains is that I made $\mathcal{A}_1$ the speaker in both the games. What if he had been the addressee in, say, g^{13}? How would $\mathcal{A}_3$ know which equilibrium to discard as she plays just that one game? She cannot choose arbitrarily as $\mathcal{A}_1$ may already have made his choice in the other game g^{12} which could conflict with her choice. In such cases what happens is that $\mathcal{A}_1$, being a player in both games, gives $\mathcal{A}_3$ feedback via repetition or gestural overdetermination that a certain word has already been used for another conventional meaning. So it does not matter who the speaker is and who the addressee is, in general. All permutations are allowed.

This may strike the purist as unsatisfactory as they may want something fully internal to the setup. But my goal is empirical adequacy and in actual fact such original acts of semiosis are not likely to happen in a pristine way but in a somewhat messy back-and-forth, repetitive, and overdetermined way. Also, there are likely to be a few more Communication Games with the same two words among the three members of our tiny community before the conventional meanings are solidly cemented. But the simple Language Game I have presented captures the heart of the matter. And it is easy to create more complex Language Games even with just three members because we can add multiple links between any two members as they can talk to each other multiple times.[8]

Indeed, it is almost as easy to generalize the foregoing to n agents connected by multiple Communication Games in a new Language Game Γ together with m words, m conventional meanings, and m referential meanings of interest to the community. Let $S^{\star}_{ij}$ be the equilibrium strategy of agent $\mathcal{A}_i$ in game g^{ij} (or, equivalently and a little loosely, in Γ^{ij}) which is a part of Γ. This could be a speaker strategy or an addressee strategy and there will be as many such strategies as there are games $\mathcal{A}_i$ plays with the other agents in the community. $\mathcal{A}_j$ will naturally play the same game and in that case it will be called g^{ji} and the corresponding equilibrium strategy will be $S^{\star}_{ji}$. If $\mathcal{A}_i$ is the speaker then $\mathcal{A}_j$ will be the addressee and if $\mathcal{A}_j$ is the speaker then $\mathcal{A}_i$ will be the addressee. Keep in mind

[8]Such a graph with multiple possible links between any two vertices is called a *multigraph*.

that both the speaker strategy and the addressee strategy *in the initial situations that are actual* are ordered pairs, either $(\omega_{l'}, P_l)$ or (σ_l, P_l) with $l = 1, 2, \dots, m$ and $l' = 1, 2, \dots, m$.

The first index i in the strategy $S^\star_{ij}$ goes from 1 to n because the Language Game is a connected graph[9] and each of the n agents participates in at least one game. The second index j belongs to an index set $K_i \subset \{1, 2, \dots, n\}$ because the agent $\mathcal{A}_i$ may not talk to every other agent and also not to himself. So there will be index sets K_1, K_2, ..., K_n for each of the n agents. For example, $\mathcal{A}_1$ may talk to $\mathcal{A}_4$ and $\mathcal{A}_7$ via Communication Games Γ^{14} and Γ^{17} and to no one else in which case $K_1 = \{4, 7\}$. And so on, for each of the n agents.

Lastly, let the intention to convey a particular content (σ_l, P_l) in g^{ij} be denoted by intention^{ij}_l. A meaningful equilibrium may now be defined as follows.

Definition 22.1. A Language Game $\Gamma(n) \equiv \Gamma$ involving m words, m conventional meanings, and m referential meanings relative to appropriate utterance situations is given. Let $S^\star_{ij}$ be a component strategy of g^{ij} in Γ with $i = 1, 2, \dots, n$ and $j \in K_i$ where $K_i \subset \{1, 2, \dots, n\}$ is an index set containing indices j corresponding to each game Γ^{ij} in Γ. Then $S^\star_{ij}$ is a component of a meaningful equilibrium of Γ if and only if

1. Nash Condition: $S^\star_{ij}$ is a component of a Nash equilibrium of g^{ij}, either the speaker component or the addressee component depending on whether $\mathcal{A}_i$ is the speaker or addressee of g^{ij}.

2. Consistency Condition: $\text{intention}^{ij}_l \neq \text{intention}^{ik}_{l'}$ if and only if $1^{st}(S^\star_{ij}) \neq 1^{st}(S^\star_{ik})$ and $2^{nd}(S^\star_{ij}) \neq 2^{nd}(S^\star_{ik})$ for all $k \in K_i$.[10]

The first condition just says that every meaningful strategy must be part of a Nash equilibrium in the relevant component partial information game. The second condition says that the same speaker strategy or the same addressee strategy can be employed in two different component games an agent plays if and only if the intentions in those component games are the same. This is somewhat compactly expressed and I urge the reader to work out what happens when $S^\star_{ij}$ and $S^\star_{ik}$ are both speaker strategies or both addressee strategies or one is a speaker

[9] For notational reasons, it is easier to stick to a graph than a multigraph. But the definition and theorem below carry over to multigraph Language Games without any difficulty.

[10] Here, $1^{st}(a, b) = a$ and $2^{nd}(a, b) = b$, that is, 1^{st} picks out the first component of an ordered pair and 2^{nd} picks out the second component.

strategy and the other is an addressee strategy. All four cases are covered by the above statement of the Consistency Condition.

Since the number of words, conventional meanings, and referential meanings are all m, all Definition 22.1 says is that the set of words is mapped via a bijection onto the set of conventional meanings in equilibrium. There are $m!$ such permutations if the payoffs are all symmetric as they were in Figure 22.3 and any one set of meaningful equilibrium strategies can then be selected.

Since such a bijection always exists, and since Nash equilibria always exist, a meaningful equilibrium always exists, and we can record this important fact below.

Theorem 22.1. A meaningful equilibrium always exists in a Language Game $\Gamma(n)$ involving m words, m conventional meanings, and m referential meanings relative to appropriate utterance situations.

Thus, the foregoing explains with respect to a relatively simple Language Game how words acquire conventional meanings and, therefore, how semiosis gets off the ground. If required, it may be assumed that such pairings are further cemented by repeated use via Γ played at subsequent times.

Combined with the analyses of Parts III and IV, this completes the reduction of meaning, how both conventional meanings and referential meanings attach to utterances, assuming nothing more than the partial rationality of finite agents. In this sense, the main problem of semantics identified in Section 1.1 – how language acquires meaning – has been solved at an abstract level. It required first solving the problem of communication in the small – micro-semantics – assuming conventional meaning was fixed, and then solving the problem of communication in the large – macro-semantics – allowing conventional meaning to vary. This top-down way of approaching the main problem of semantics via communication provides a fairly tight constraint on how more detailed accounts of particular expressions in language are constructed. As I said in Section 1.2, it brings a *uniformity* to the many subquestions of semantics.

Though the model actually constructed is very simple, it has elements of Peirce (1867–1913/1991), de Saussure (1916/1972; 1960), and the later Wittgenstein (1953/1968), the first because it shows how the connections between words and conventional meanings are arbitrary, the second because it shows they are simultaneous and systemic, and the third because it shows how all meaning arises ultimately from use or communication.

22.2 Some generalizations

I now briefly describe some generalizations mentioned in Section 3.2.

The most obvious one is to allow lexical ambiguity by making the number of conventional and referential meanings of interest m' with $m' > m$. Since the number of words m would be less than the number of conventional meanings m', some or all of the words would have to carry more than one meaning. This is relatively straightforward to do by simply adding more initial situations to the games of partial information such as g^{12} and g^{13} and making the game trees a little more complex. For example, if $m = 2$ and $m' = 3$ then in the simple Language Game with $n = 3$ agents we would add one initial situation to both g^{12} and g^{13}, each corresponding to some third conventional meaning P_3 and referential meaning σ_3, and add more branches to each situation. This slightly complicates the definition of a meaningful equilibrium and makes it a little harder to see its simple essence so I do not pursue it here.

In actual fact, the number of possible conventional meanings, that is, properties, is infinite. So, in this case, one has to first identify a finite number of conventional meanings that have the highest frequency of use and then follow the procedure above.[11] This can be done in one fell swoop by simply assuming anticipated frequencies abstractly and selecting some finite number of conventional meanings at the start, or it can be done via a dynamic Language Game played at discrete times $t = 0, 1, 2, \ldots$. The latter is naturally more difficult to work out and involves semantic change, something I will consider in Chapter 24.

It is also not too hard to extend the utterances to full multiple-word sentences and allow a pre-given grammar. All it involves is adding syntactic partial information games to the semantic ones we already have. This is notationally cumbersome but not conceptually so, especially for relatively simple sentence structures, and, again, the definition of a meaningful equilibrium has to be modified to include the syntactic games.

Something much more adventurous is also conceivable. It is possible not only to generate the (locutionary) Semantic Constraint but also the Syntactic Constraint from nothing but rationality. That is, we can show how sentences acquire syntax and how a language acquires a grammar. This could result in contradicting Chomsky's account of a universal grammar that is innate, something that is being questioned anyway.

[11]Incidentally, this double fact, of identifying a finite number of possibilities for conventional meanings based on frequency of use, and then identifying the equilibrium of the relevant finite game, seems to solve Wittgenstein's skeptical paradox (of plus and quus) as discussed by Kripke (1982).

A similar strategy can be deployed with phonetics but as I have not said much in detail about phonetic games I just mention the possibility. The reason why all three Constraints – Semantic, Syntactic, and Phonetic – can be handled in analogous ways is that, as I've said earlier, they *are* analogous. All involve attaching semantic, syntactic, or phonetic values or contents to words and then allowing such couplings to be disambiguated in identical ways via the Flow Constraint.

This indicates that really only the Flow Constraint corresponding to partial rationality is fundamental. The other three are a consequence of communication viewed macro-semantically. I haven't said how the Referential sub-Constraint of the Semantic Constraint might be derived. Suffice it to say here that it is a result of rationality and ontology. I discuss it partly in *Language and Equilibrium* (2010: Section 2.4). Thus, the partial rationality of agents engaged in communication is responsible for all of semiosis and communication itself arises from the needs and goals of agents.

Vagueness is harder to handle because different agents will come to attach slightly different concepts to words as discussed in Chapter 11. This means conventional meanings are not fully shared and it requires relaxing the Consistency Condition slightly and allowing an extended locutionary Semantic Constraint to operate. Modulation, implicature, and free enrichment are also not easy to incorporate. It might be reasonably conjectured that early communication with new words is free of such complications. The referential map is more complex than I have assumed because a conventional meaning can lead to more than one referential meaning (e.g. for words like THE and EVERY). This flexibility also needs to be accommodated and the map derived from first principles as mentioned above. Conflict can also be introduced into the relevant Content Selection Games of each Communication Game in Γ. Since conflict usually does not enter into the corresponding partial information games, this does not pose insuperable problems. Finally, just as it is possible to derive grammar from rationality in principle, it should also be possible to pull rational ontologies out of seemingly thin air.

These last few extensions are undoubtedly more challenging. In any case, once such models have been elaborated it should be possible to apply them empirically as suggested in Section 3.2.

22.3 Comparisons

Grice (1968) sketches an account of word meaning on the basis of an agent having a procedure in his repertoire to convey a certain meaning. It has the following drawbacks:

1. It leaves out the element of choice that inevitably arises when deciding which word to pair with which meaning. It focuses on a single word-object pair and so leaves out the key to symbolic meaning that Deacon (1997) emphasizes in the first part of his book: that symbolization is *systemic* and the link between word and referent is mediated by the link between word and word. This, incidentally, is exactly what happens in a Language Game via the bijection referred to in Section 22.1.

2. It fails to show how individual decisions result in a group conventional meaning. Though Grice (1989a: 127) does say that each member of the group acts in a certain way on the condition that other members do likewise, he does not address how such a circular situation might get off the ground. In a much later publication, Grice (1989b: 290–297) constructs a kind of mythical scenario for how nonnatural meaning might emerge from natural meaning. As I wrote in Chapter 5, because he lacked the apparatus of game theory, he could not quite envisage how such nonnatural pairings might arise naturally through rational choice although they may build on quasi-natural gestural pairings.

3. A less significant missing element is the two-tier system of conventional and referential meaning as opposed to simple word-referent connections. This could presumably be added to Grice's account without too much difficulty.

Lewis (1969) is able to overcome the first shortcoming above with his invention of signaling games. But, short of everyone playing one large communal game rather than participating in a network of two-person games, he shows no mechanism to go from micro-semantics to macro-semantics, from individual to group. What is missing is the idea of multiple, interlocking games in society that enable the micro level to be linked with the macro level. This is what makes Lewis's notion of convention too strict as I will argue in the next chapter. And the third flaw mentioned above is also present.

Because there are perfectly symmetrical multiple equilibria in Lewis's signaling games, Skyrms (2010) adds adaptive dynamics of a special sort (e.g. replicator dynamics, reinforcement learning) to break the symmetries to realize unique equilibria in most cases given arbitrary starting points in the space of possible strategies. This conclusion does not appear to be so remarkable because rather special sorts of dynamics are assumed without much argument. In a lucid paper, Huttegger (2007)[12] suggests that such dynamics may be supported by assuming

[12] See also the references therein.

that agents imitate one another though this argument is made in a very general way and is also rather vague: why would one kind of dynamics be copied rather than another and how does the process get off the ground? It is not enough to say that success is copied because it may be difficult for the agents to gauge what is successful in many circumstances. Also, the method seems to allow only the simplest cases to be considered where there may be two words and two meanings or possibly *m* words and *m* meanings. Apart from the fact that such dynamical movements in strategy space may take a relatively long time (e.g. a thousand iterations) to converge to an equilibrium implying nonequilibrium behavior for an excessive duration, the approach is implemented at such a high level of abstraction that it is unrealistic and unclear how to get to more empirically interesting results.

Skyrms's model is behaviorist and makes the intentions and intelligence of human agents superfluous. That the same framework applies to both bees and humans is a flaw, not a virtue, because something important has been left out and the special nature of human conventional and referential meaning is lost. Surely humans do not need to blunder about a thousand times the way bees may have to in order to reach an equilibrium.

He misses the role of gestures that are likely to have accompanied speech at an early stage of evolution. As I explained, linguistic meaning follows gestural meaning temporally, the latter being much easier to establish, and overdetermination enables one-shot pairings to occur quite often. This kind of bootstrapping allows agents to avoid potentially questionable and extended dynamics to reach equilibrium.

In trying to avoid assumptions like common knowledge and salience that Lewis made, Skyrms seems to ignore the role of the situation in which communication occurs. Early signals were likely meant to identify salient objects in the environment such as food or predators and these could have been steppingstones to later signals where such salience was not available. And, as I have argued, more complex communication often relies on aspects of the situation that are invisible such as shared background information.

He also focuses exclusively on pure coordination. As Gintis points out in his review of Skyrms's book:[13]

> As it turns out, there are very few zero-sum games in reality, and there are equally few pure coordination games. Thus Skyrms misses most of the real action, which lies in understanding the conditions under which signaling

[13] https://www.amazon.com/gp/customer-reviews/R1TCQS6NFB5M3K/

> equilibria emerge even when agents do not have identical interests. This is most acute in human societies in which individuals routinely stand to gain from transmitting untruthful messages. But it is equally important in animal societies in which males gain from transmitting incorrect messages concerning their reproductive fitness to females, who use such information to guide their choice of mates. There is a brilliant literature on this topic, based on contributions by Ronald Fisher, Michael Spence, Amos Zahavi, and Alan Grafen, but this literature is not touched upon in Skyrms's exposition.

As I said in the previous section, conflict can be introduced in Content Selection Games and partial information games if required more or less painlessly and so Language Games can readily handle both coordination and conflict.

Lastly, Skyrms does not address the issue of the need for a network of interconnected games. This suggests that he operates within the same ambit as Lewis as far as linking the micro and macro levels is concerned.

As an aside, I point out that Skyrms generalizes the notion of semantic content by making it probabilistic but my notion is more general yet because there are multiple distributions in a content as discussed in Sections 15.2 and 15.4. Also, Skyrms's content is holistic and does not allow for partial propositions (e.g. the content of an uttered noun phrase). Lastly, it is not Austinian as it does not incorporate the described situation or what he calls the state. As a result it is not fine-grained enough and succumbs to logical omniscience and to the Fregean problem of informative identities.

Thus, my basic verdict is that signaling games in the Lewis-Skyrms tradition are on the right track but do not go far enough toward explaining the origins of meaning.

Tomasello's (2010) account is more satisfying, both theoretically and empirically. He emphasizes the precedence of gestures as compared with words and uses the term "piggybacking" to describe what I have called overdetermination. His argument about so-called recursive Gricean "mindreading" is less convincing and I have argued in earlier chapters that it is generally not required. He too underscores the situated and intelligent nature of communication as I have throughout this book and in my earlier publications, but because he does not avail of game-theoretic ideas, he misses the possibility of arguing for his conclusions in greater detail and with greater rigor through the use of mathematical models.

23 Convention

What justifies calling conventional meanings conventional?

Lewis (1969: 58) says there are three basic conditions for a regularity in the behavior of members of a population to be a convention: the regularity itself, a system of mutual expectations, and a system of preferences. On top of this, he demands that these three conditions be common knowledge in the population.

There are two undesirable characteristics in this account, one explicit and one implicit, that are best brought out by an example.

First, if a population is reasonably large, no member is likely to know all of it. That is, some people will be strangers to one another. Intuitively, we know that conventions can and do exist in such communities but common knowledge of any regularity will be impossible because the members do not even know of one another's existence in a complete way.

Second, many conventions (e.g. a handshake) are instantiated locally (e.g. between two agents). Lewis modeled convention with a single large game played by all the members simultaneously. This represents an extreme case and Lewis's implicit use of a single game is a special degenerate case of a network, especially when seen in the context of Language Games that involve a network of games.

These observations suggest that actual conventions are weaker than Lewis thought. That is, his explicit condition of common knowledge and implicit condition of a single large game are sufficient but not necessary for a regularity to be a convention. His notion thus needs to be generalized. We can amend his definition as follows, keeping his original notation and terminology intact.

Definition 23.1. A regularity *R* in the behavior of members of a population *P* when they are agents in a recurrent network of games *S* is a convention if and only if it is true that, and members' probability estimates of other members belonging to *P* are sufficiently high that, in any instance of *S* among members of *P*,

1. everyone conforms to *R*;

2. everyone expects everyone else to conform to *R*;

3. everyone prefers to conform to R on condition that the others do, since S is a network of games and uniform conformity to R is the preferred equilibrium of S among at least two equilibria.

There are *two* changes to Lewis's definition: one is the weakening of the requirement of global common knowledge and the other is the replacement of a single large game by a network of games.

If we substitute a Language Game Γ for S in the definition, we can establish why the meaning that results from a meaningful equilibrium is *conventional* in this generalized sense of convention. In this case, the regularity R is just the pairing of a word with a property. I record this important fact.

Theorem 23.1. The conventional meanings that result from a meaningful equilibrium of a Language Game are in fact conventional. In other words, the name given to this kind of meaning is appropriate.

This also implies that other conventions – like shaking hands to greet someone – are best analyzed via networks of interconnected games rather than as a single large game as Lewis did. To the extent that norms exist as conventions, this applies to norms as well.

Interestingly, then, our study of the origins of linguistic conventions has led to a broader notion of convention in general and to a revision of Lewis's classic definition.

Sometimes, it appears that we get stuck in suboptimal conventions, that is, in equilibria that are not Pareto-optimal. A notable example is the QWERTY keyboard. In such cases, it often happens that a closer look reveals that the supposed suboptimal convention was, in fact, optimal at a given time in the past, and later, when new technologies arose, other conventions became optimal but inaccessible owing to the high switching costs. In other cases, it may happen that the incremental process of reaching an equilibrium is piecewise optimal but not globally so, resulting in an inferior convention in the end that becomes sticky. This kind of thing is likely to prevail in many historical conventions, such as the overall outcome of an incremental peace process, for example.

The foregoing addresses some of Burge's (1975) criticisms of Lewis's notion. First, I replace Lewis's global common knowledge by probabilities and local (generally nonconscious) common knowledge. Second, I consider the issue of switching via a consideration of costs. Burge is also concerned about an over-zealous reliance on rationality in situations where "inertia, superstition, and ignorance" may prevail. This observation may be accommodated by using nonconscious partial rationality and awareness (e.g. an agent may know only a fragment of the objective partial information game and so may not realize there are other alternative equilibria).

Gilbert (1981) makes some interesting points about the details of Lewis's framework but they are not relevant here as my framework is very different.

24 Semantic change

I referred in Section 3.2 to the fact that not only do conventional meanings emerge through communication, they also evolve over the centuries. Such semantic change is part of the broader changes that occur in all aspects of language. My goal here is to sketch the mechanism by which such changes occur and show how it fits in with the rest of Equilibrium Semantics. As such, it merely scratches the surface of the large field of historical linguistics and, in particular, historical semantics and historical pragmatics, an instance of which may be found in Traugott & Dasher (2007) for example.

An important constraint on almost all language change is that it must occur during communication. A speaker may utter a sentence to communicate some information to an addressee. Usually, the utterance merely reinforces standard ways of speaking (or writing). Occasionally, one or more aspects of the utterance – either semantic, syntactic, or phonetic – are altered. For example, a single word of the sentence uttered may involve a broadened meaning. The addressee then has to comprehend the utterance despite the alteration. If she is successful, the relevant change gets established between the two agents and can be repeated between them. Such a change may then also be transmitted to other agents by either of the two agents via similar utterances and may thereby spread through the linguistic community and become a change in the language.

Uttering sentences thus serves at least two ends: the first, generally conscious and intended, is to enable the flow of information[1] from speaker to addressee, and the second, generally unconscious and unintended, is to either reinforce or change aspects of language. Because communication and language change must occur together, the constraints on the former also help to constrain the latter.

A key constraint that makes change possible was called *overdetermination* in Parikh (2010: Section 6.1.5). Overdetermination in this second sense[2] is the presence of an element in an utterance that provides information about another element in that utterance. For example, in the sentence, "The dog didn't bark," the verb phrase DIDN'T BARK constrains the noun phrase to refer to something that

[1] I mean to include all kinds of performative utterances of the kind Austin (1975) introduced under this broad description.

[2] The first sense was discussed in Section 22.1 in the context of Language Games.

can bark – such as a dog or, less likely, a seal – in most ordinary utterance situations. Even if the word DOG had been eclipsed by some noise, the addressee would still have understood the utterance as conveying the information that the *dog* didn't bark because in such everyday situations it is only dogs that bark.[3] We can say here that the verb phrase overdetermined the meaning of the noun phrase given the context. In general, the elements that do the overdetermining can be semantic, syntactic, phonological, or contextual.[4]

The second constraint that enables change is the mechanism through which overdetermined contents are transmitted, namely, *modulation* or *free enrichment* or *implicature*.

I will propose a model of language change based on overdetermination and modulation and on the propagation of an initial change through the community by looking at an example of semantic broadening. Many ideas in the theory of language change (e.g. Crowley & Bowern 2010) appear to be qualitative and large-scale, that is, they do not examine language change in the small and then relate it to the community-wide scale. In my view, only this kind of communicative cum propagative process can really explain it. Large-scale evolutionary models by themselves abstract too much from the essence of the process. A better way is to see the evolution as a sequence of small steps, each of which is modeled in detail.

24.1 Semantic broadening: The communicative aspect

The modern English word DOG derives from the middle English word DOGGE (which comes from the old English word DOCGA). Earlier, the word referred to a powerful breed of dogs (such as Great Danes) that originated in England. It was later semantically broadened (and transformed morphologically). This change must have come about through communicative interactions and I now look at the details of the semantic transformation. I abstract from the morphological change and so will use the word DOG to refer to both its narrower and broader meanings.

Presumably, the word was used by some particular speaker in a sentence to *communicate* the broader meaning for the first time in the history of English. We can reconstruct the kind of original situation in which the change occurred and

[3]Consider the sentence "His boss didn't bark at him." This and similar examples show that overdetermination works only when the utterance situation is also sufficiently constraining to rule out other alternatives.

[4]This notion of overdetermination is a generalization of the idea of selectional restriction.

was communicated successfully by the following thought experiment. Middle English existed from roughly 1150 CE to 1470 CE and the paradigm situation I consider is relatively more modern. A more authentic example would require delving into the relevant sociology of late medieval households.

Suppose $\mathcal{A}$ and $\mathcal{B}$ have a poodle[5] and they engage in the following dialogue.

> $\mathcal{B}$: When did you get home last night?
> $\mathcal{A}$: Around midnight. The dog didn't bark.

It is not necessary to know *why* $\mathcal{A}$ used DOG to refer to their poodle. Maybe it was meant affectionately. Be that as it may, in the scenario sketched, the word was used to refer to their poodle and not to some larger dog which would have been the then-standard use.

Consider the second sentence $\mathcal{A}$ utters: "The dog didn't bark." ($\omega = \omega_1\omega_2\omega_3\omega_4$)

There are two possibilities for ω_2 = DOG. The first is the old conventional meaning of ω_2, the property $P_1^{\omega_2}$ of being a Great Dane, say. The second is the new conventional meaning $P_2^{\omega_2}$ the utterance somehow succeeds in establishing. Initially, the addressee considers just the first of these.

Now, first the locutionary Semantic and Flow Constraints are activated for ω_2 and the other words. This leads to a problem as there is no large dog in the resource situation accompanying the utterance. This triggers the illocutionary Semantic Constraint which results in finding $P_2^{\omega_2}$ and the corresponding referent by using overdetermination, and then a mixed Flow Constraint resolves any ambiguities that remain.

How does $\mathcal{B}$ decide whether the new conventional meaning $P_2^{\omega_2}$ is the property of being a dog (in its broadened modern sense) or the property of being a poodle? Here, it is arguable that there is a linguistic constraint that involves a preference – in *some* situations – for semantic broadening (from Great Danes to any dog) over a shift to a parallel and coordinate property (from Great Danes to poodles). In the first case of broadening, the earlier meaning is superseded by the new meaning; in the second case, the earlier meaning remains and the new meaning becomes a second meaning of the word. The latter is costlier in general and so the first option is preferred in *certain* situations.[6]

[5] Apparently, the poodle was imported into Great Britain from France in the 1870s but treat the time as the 15th century.

[6] Consider the (semantically related) term "bitch." It derives from the old English word "bicce" and originated as a neutral term for a female dog, a meaning it still has. But when it was used to refer to a human, apparently around the fifteenth century, with an intended meaning like *bad woman* or *virago*, that meaning stuck. It was not broadened to something like a female mammal. See the relevant entry in the Oxford Dictionary of English Etymology. My thanks to Tom Wasow for this example.

At this stage, we can say that in the communication between $\mathcal{A}$ and $\mathcal{B}$ a new conventional and therefore referential meaning get conveyed via modulation. There is no guarantee that the new conventional meaning will stick, even just between the two interlocutors. Indeed, most new meanings are short-lived. But any initial change has the potential to be propagated through the community and become a lasting change.

24.2 Semantic broadening: The propagative aspect

Recently, evolutionary models that abstract from the details of the propagation and study how suitable fractions of the population converge to a steady state based on positive or negative reinforcements have become popular.[7] In my view, such models are too abstract, and assumptions involving reinforcements function more or less as mathematical fictions that enable the derivation of desirable empirical results. An alternative approach that builds on the preceding model of language change in the small is as follows.

After $\mathcal{A}$ has successfully communicated the new meaning to $\mathcal{B}$ above, the change may last in either person's mind or it may fade away. So we may assume there is some probability q of a change being communicated and lasting between the two interlocutors given that a conversation takes place between them. If the change lasts, then either of them can pass it on to a third party via a similar modulation. And so on.

Start with a finite community of n persons 1, 2, 3, …, n.[8] Then let 1 have some altered meaning for a word in mind at time $t = 0$.[9] The people in the community who accept the new meaning are collectively called the change set. Thus, at $t = 0$ the change set is just {1}. Assume that conversations take place with probability $p > 0$ between some pairs j and k and with probability zero for other pairs. If a communication is successful and the change is lasting, it cannot be reversed. This gives us a sequence of random graphs $G_t(n, p)$ at each subsequent stage $t = 1, 2, 3, \ldots$ where the n vertices represent the n members and an edge between any two vertices represents a conversation between the two corresponding persons and occurs independently for some pairs of vertices with probability $p > 0$. I will assume that each graph in the sequence is connected in the same way, that is, there is a path between any two nodes where a path is a sequence of distinct

[7] See Yang (2009) and the references therein.

[8] Thus, n does double duty: it serves as a name of a member and also indicates the total number of members.

[9] The member 1 plays the same role of originator as $\mathcal{A}$ did in the previous section.

nodes with edges between successive nodes having $p > 0$.[10] If at least one of the vertices belongs to the change set prior to the time of communication and the communication succeeds and the change is lasting, then the other vertex is added to the change set. In this way, the change set keeps expanding and hopefully converges to the entire community of n vertices.

The basic question is: what are the conditions on n, p, and q under which this happens? This is not the most general formulation of the situation but it suffices to state a simple version of the problem.

Let C denote the entire community $\{1, 2, 3, \ldots, n\}$ and C_t denote the change set at the end of period t. We would like to ask under what conditions $\lim_{t \to \infty} C_t = C$.

If u, v in C are two distinct vertices, let $u \sim v$ represent an edge between u and v or, in other words, a conversation between them. Then the probability of a conversation between u and v is $P(u \sim v) = p$ as specified by the statement of the problem. Also, if E is the event of a communication between u and v whose effect persists, then $P(E \mid u \sim v) = q$ again as specified.

As a result, the probability of a successful conversational event with a lasting communication at any given instant is given by $P(u \sim v \;\cap\; E) = P(u \sim v)P(E \mid u \sim v) = pq$. Thus, the probability of the complementary single failure at any instant is $P([u \sim v \;\cap\; E]') = (1 - pq)$.

Now let $k \in C$ be an arbitrary member distinct from 1. Because the graphs are connected, there is always a path between 1 and k. If we are able to compute the probability of k entering C_t, then we would be able to see what happens to it in the limit as $t \longrightarrow \infty$. Let this probability be denoted by $P(k \in C_t)$.

First consider a path $(1, k)$ of length $l = 1$ where the vertices 1 and k are adjacent. If k fails to be in C_t given a path of length $l = 1$, this must have the probability of a single failure $P([u \sim v \;\cap\; E]')$ repeating itself for instants at $1, 2, 3, \ldots, t$. This allows us to write:

$$P(k \notin C_t \mid (1, k)) \;=\; P([u \sim v \;\cap\; E]')^t \;=\; (1 - pq)^t$$

The probability of the complementary event, that is, the probability of success

[10] A connected graph is usually just one where there is a path between any two vertices. In the case of a random graph, one has to add the requirement that $p > 0$ over the edges constituting a path. Moreover, in our case, the connections must be the same over time in the sequence of random graphs $G_t(n, p)$. I have deliberately used a slightly more restrictive notion of path as it is more convenient for our purposes. It is possible to work these random graphs into the Language Games considered earlier as the two are related but this formulation is simpler. Keep in mind that we know empirically that there are on average six degrees of separation between any two people on the planet.

over this path, is:

$$P(k \in C_t \mid (1, k)) = 1 - (1 - pq)^t$$

Next consider a path of length $l = 2$ and let the intervening node be j. That is, let the particular path be $(1, j, k)$ where $j \in C$ is different from 1 and k. The key thing to be aware of here is that if k is to become a member of C_t, then it can do so only after j has first become a member of the change set. Let $t_0 = 0$, let t_1 be the time at which j succeeds, and let $t_2 = t$. Then, it is possible to write the following:

$$P(k \in C_t \mid (1, j, k)) = [1 - (1 - pq)^{t_1 - t_0}][1 - (1 - pq)^{t_2 - t_1}]$$

which is nothing but the success of j at t_1 followed by the success of k at t_2.

Now suppose a particular path of arbitrary length l is $(1, j_1, \dots, j_{l-1}, k)$, where each element of the sequence is a distinct member of C. Then for k to become a member of the change set, it is necessary that each j_m has to have become a member of the change set first and in the right order, starting with j_1 and ending with j_{l-1}. If we let $t_1, t_2, \dots, t_{l-1}$ be the successive times at which the corresponding nodes $j_1, j_2, \dots, j_{l-1}$ become members of the change set, and if we let $\Delta t_m = t_m - t_{m-1}$ for $m = 1, 2, \dots, l$ with $t_0 = 0$ and $t_l = t$, the probability of success over this path is given by:

(32) $$P(k \in C_t \mid (1, j_1, \dots, j_{l-1}, k)) = \prod_{m=1}^{l} [1 - (1 - pq)^{\Delta t_m}]$$

As $\Delta t_m \longrightarrow \infty$ for all m from 1 to l, the conditional probability of success tends to 1 as long as $pq > 0$, that is, as long as $p > 0$ and $q > 0$.

If there is more than one path between 1 and k, the (unconditional) probability of success $P(k \in C_t)$ can only be greater. That is:

$$P(k \in C_t \mid (1, j_1, \dots, j_{l-1}, k)) \leq P(k \in C_t) \leq 1$$

which implies that:

(33) $$\lim_{t \longrightarrow \infty} P(k \in C_t) = 1$$

as well by the so-called squeeze theorem from elementary calculus.

Since k was arbitrarily chosen, this shows that:

$$\lim_{t \longrightarrow \infty} C_t = C$$

as long as $p > 0$ and $q > 0$. The convergence is independent of n, the size of the community.

In other words, as long as a path exists between the originator and every member of the community and the probability of a lasting communication is positive, all members will eventually fall into the change set. That is, an initial change between two members becomes a change in the language. This result fits the example I have considered as the narrow meaning of DOG has completely disappeared.

There are many ways of relaxing the simplifications assumed earlier. Many potential chains of change could be broken by an errant member, that is, q is likely to vary from pair to pair rather than be constant across the community. Also, one could assume there are two originators of two distinct meanings for a word and study what happens in a community whose graph forms two relatively isolated but connected components. This may be how linguistic variation occurs, creating distinct communities and languages over time when a sufficient number of changes accumulate in one component of a community. Labov's (2001) detailed study of social factors influencing linguistic change may also suggest other ways to expand the model.

24.3 Generalizing the model

This account of semantic broadening can be adapted to other kinds of language change. For example, a syntactic change involving the present continuous (e.g. "I am liking your shoes") would initially appear odd but would, via syntactic modulations, recover the intended meaning that *I like your shoes*.[11]

I have added a philosophical and pragmatic dimension to the problem of language change by connecting it to a theory of communication. Does it account for all types of language change? Probably not. Even if it does apply to, say, phonetic change, it must operate at a less conscious level than our semantic example.

I have also tried to propose a new graph-theoretic model that builds on the idea of change occurring step by communicative step in a community. It appears to offer certain advantages over large-scale evolutionary models involving segments of populations by sticking more closely to the knitting, as it were.

As both these aspects can be generalized in many directions, they together amount to an outline of a framework for the study of language change.

[11] The example is an instance of a variant of Indian English and is influenced by the corresponding syntax of many Indian languages.

25 Beyond language

There are broadly two ways to go beyond language. One is toward other symbol systems such as images and gestures. The other is toward actions that also function as utterances. I have dealt with both directions in *The Use of Language* and to a lesser extent in *Language and Equilibrium* but I briefly examine an instance of the second in a little detail to show how Equilibrium Semantics can connect with the wider social sciences.

Consider an example from the field of international relations which may be thought of as applied political science and sociology not to mention economics. Imagine a situation where there are two neighboring countries X and Y and X begins massing troops near their common border. This is not at all an uncommon situation and one that can be quite ambiguous.

First, there is the physical action itself which could be an offensive or defensive move in the ongoing relations between X and Y. It could also be a response to internal developments in X (e.g. to divert attention from them). Which of these options is actually the case would depend on the context.

But beyond this level, there could also be a second level where the physical action functions as an utterance and conveys a meaning to the government of Y and possibly to other parties such as the citizens of X and Y. This meaning could involve any of the possibly many contentious elements in the relations between X and Y. For example, the massing of the troops could serve as a warning or intimate an intent, depending on the concrete profile of X and Y and their interactions. And it could be highly ambiguous and unclear.

Such a meaning is essentially illocutionary and, indeed, like an implicature because it is calculable, cancelable, indeterminate, nondetachable, and reinforceable. It can be derived by the same methods involving the illocutionary Semantic Constraint consisting of Relevance and Distance and the corresponding Flow Constraint.

There are innumerable such examples where an action's purpose is not restricted to its immediate consequences but also includes a layer of illocutionary meanings. And Equilibrium Semantics applies to all of them. Thus, its relevance for anthropology, sociology, political science, economics and allied fields is not limited to the direct role of language and meaning in the emergence, evolution, and transformation of social institutions but also includes the entire realm of action itself.

Part V shows how Language Games and meaningful equilibria may be used to explain the origins of conventional meaning from the ground up and compares my approach with that of others. Once such models are developed in greater detail, they could begin to yield empirical results. An unexpected bonus is a new insight into the nature of convention broadly conceived. A communicative approach to historical linguistics is forged, one that establishes an initial semantic change via modulation that is then propagated step by step to the whole community via a random graph. Lastly, language is just the first rung of a tall ladder that leads to other symbol systems and social action and ultimately to all of civilization.

Part VI

Conclusion

26 Communication, Frege's puzzle, and reference

We have now traversed the whole path of Equilibrium Semantics. Looking back, it is possible to see many different dimensions of the problem of meaning – philosophical, linguistic, psychological, computational, and even humanistic, encompassing many aspects of the human sciences as discussed especially in Chapters 19 and 25. Throughout, I have tried to demonstrate that the idea of equilibrium pervades all of semiosis.

Over the course of its modern history, many have felt the intuitive appeal of a use-based or communication-based foundation for semantics but the content of this Wittgensteinian insight has always been obscure, even mysterious. What can one possibly do with *use*? It seems like such a structureless, anti-theoretical idea. Indeed, even connecting use with communication appears ineffectual because the latter looks like it has little to offer beyond Austinian and Gricean informalities. And so the relative – but viciously circular and question-begging[1] – safety of a Fregean and logicist truth-based foundation has continued to beckon.

In my view, neither use-based unsystematicity nor truth-based systematicity offer a scientific approach to semantics. The only remaining possibility is to follow Grice, whose singular attempt to understand communication in the small and in the large was systematic but informal. I have tried to present a very thorough critique of practically his entire oeuvre together with that of his many followers whether they are orthodox Griceans, neo-Griceans, or post-Griceans.

Equilibrium Semantics may be seen as a mathematical alternative to Grice that organizes the raw material of semantics quite differently. It rests on four fundamental ideas: reference, use, indeterminacy, and equilibrium, the first taken from ideal language philosophy which was mainly truth-based, the second taken from ordinary language philosophy which was mainly use-based, and the last two supplied by my framework itself.[2] These are developed into Communication Games and Language Games and into four Constraints, Semantic, Syntactic, Phonetic,

[1] See the quote by Dummett in Section 1.2.

[2] I have discussed these four ideas in a general way in some detail in my previous book and so will not do so here.

and Flow, based on the mathematical frameworks of situation theory and game theory. It offers in principle a unified approach to all of semantics in a way that is philosophically sound, mathematically solid, computationally tractable, and empirically adequate.

26.1 Four philosophical benefits

As I mentioned in Chapter 1, there are at least four advantages to considering meaning via communication as Equilibrium Semantics does.

The first is that because it tackles the main question of semantics – how language acquires meaning – directly, the framework's many results impose a top-down *constraint* on possible answers to its various subquestions involving particular linguistic constructions. Indeed, the subtasks of semantics devolve into building alternative theories *within* the framework for the same phenomena that can be straightforwardly compared because they share a common foundation. In the same way, theories within the framework for different phenomena can be synthesized and integrated owing to a common foundation. I tried to demonstrate this latter kind of integration in Parikh (2010: Chapter 6) where I developed accounts of descriptions, names, and generalized quantifiers that all shared the same basic assumptions about how noun phrases work.

The only other possibility is to be bottom-up and piecemeal and hope that the proliferation of foundational assumptions will somehow get reconciled. It should be obvious that this is a hopeless task. In fact, many theorists end up relying on foundational Gricean assumptions, not pausing sufficiently to ask whether relegating certain things to "pragmatics" and to the Gricean maxims is a sound move.[3] If my critique of Grice is correct then this recourse is no longer available.

Thus, if Equilibrium Semantics is accepted as a constraint on the formation of theories of meaning, it will be much easier to tackle the challenges more complex constructions pose with an elaboration of uniform tools so that the many subquestions of semantics all cohere into an integrated solution to its overarching problem of the relation of language to reality and knowledge. I illustrate this in the next section by returning to Frege's puzzle of informative identities that was posed in Section 1.2.

As we saw, Dummett gives a further reason to pursue meaning via communication: to sidestep the interdependence of truth and meaning. Those relying on truth and truth-conditional semantics or pragmatics, the most popular approach today, never show how the truth conditions are derived from first principles. A

[3]Even Kripke (1977) is guilty of this.

moment's thought will reveal that it is the theorist who already understands the example being discussed who provides the truth conditions. But how do the interlocutors arrive at them? The truth conditions specify the meaning of an utterance but this very meaning is also presupposed in identifying the truth conditions.

My approach has described use directly and derived meaning from use. Once this is done, it is possible to use truth conditions or infon-based contents, the latter representing more fine-grained information conditions. This procedure dovetails with the first advantage listed above because when the content of an utterance is assumed, it is assumed *relative* to a background framework like Equilibrium Semantics. In such cases, one knows whereof one speaks because the assumptions can be made good by simply deriving the content from first principles when required. It is only then that the relevant theory can be said to be complete.

The third philosophical benefit is to reduce meaning to communication and then rely on Dretske's further reductions of mental representations to physical facts. This allows us to see meaning as part of the natural world. I believe I have developed this reduction fairly completely throughout the book. With the truth-conditional strategy, this possibility remains murky at best. Lewis (1972: 170) writes:

> I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and, second, the description of the psychological and sociological facts whereby a particular one of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics.

As I have tried to show, especially in Part V, these two topics are inextricably linked because Language Games depend on interlocking Communication Games and the conventional meanings of Communication Games result from their participation in Language Games. That is, macro-semantics and micro-semantics cannot be separated from each other. Indeed, it could not be otherwise because language and meaning are just one large social institution that hangs together *only* because its different parts and levels cohere. This is precisely what the burden of this book has been, to show how the main problem of semantics, how language acquires meaning in the small and in the large, can be solved.

Lastly, I have argued that the vagueness of language is an important *natural* source of its normativity, arising as it does from "the great variation of things in the world combined with the limitations on our concepts" as quoted from

Murphy (2004: 21) in Chapter 11. Thus, this elusive property, too, becomes clear in one of its aspects via understanding the communication of vague utterances.

26.2 Solving Frege's puzzle

I will develop just the part of Equilibrium Semantics required for the solution. I omit the four different uses of names, referential, attributive, generic, and predicative, referred to earlier and analyzed in detail in Parikh (2010: Chapter 6). Apart from the particular theory I offer, I want to emphasize that Equilibrium Semantics is a framework and as such is compatible with other possible accounts although it also rules out some. This is the sense in which it provides a constraint on semantic theories.

It is often not realized that the cognitive significance of identity statements involves *two* items of information. In the case of Hesperus and Phosphorus, not only does the addressee learn that the names HESPERUS and PHOSPHORUS co-refer but also that one is the heavenly body seen in the evening and the other is the heavenly body seen in the morning and that they are the same.[4] Most people subscribe instinctively to a direct reference theory where each name is just a label for the referent. And then neither item of information is available in the trivial content of the utterance and its cognitive significance becomes a mystery.

Direct reference theories are also held by sophisticated contemporary theorists who argue for quite counterintuitive consequences of their theory such as Hesperus's identity with Phosphorus being a metaphysically necessary truth that is nevertheless not epistemologically a priori as it is an empirical discovery.

I now solve Frege's puzzle in a way that survives challenges posed by more complex constructions and by other constraints such as Putnam's (1975) arguments about the nature of meaning.[5]

[4] It seems even Frege himself may not have realized this because his *Begriffsschrift* solution attends to the first item involving names and his *On Sense and Reference* solution attends to the second item involving modes of presentation. Subsequent scholarship, as far as I know, has also attended to one or the other but never to both in a single solution.

[5] I do not discuss how my solution offers a way of solving Putnam's problems about meaning but anyone familiar with them should be able to see how. Indeed, I feel my solution is more satisfactory than his because he includes the referent in the conventional meaning or intension of the term, which is a bit artificial. If it is kept separate in a resource situation and both the conventional meaning and resource situation jointly determine the referent then we get a more natural solution.

26.2.1 Preliminaries

26.2.1.1 Causal chains

Kripke (1980) and others developed the notion of a causal chain to account for how the reference of a name occurs. The idea is that someone in a linguistic community starts using a name to refer to an entity, say a heavenly body, and then this use spreads through the community. Korta & Perry (2011: Chapter 7) call the structure of references a network. A causal chain can in fact be rigorously defined as following a certain path within the network of Communication Games called a Language Game.[6] In general, there will be multiple causal chains for any name and the network that results will be a subgraph of a Language Game because not all parts of the Language Game will involve the name. One consequence of such a rigorous definition is that it allows us to construct the infon $\langle\langle R^c;\ a;\ N;\ u \rangle\rangle$ where a is the referent and source of the chain, N is the name that refers in the situation u, and R^c is the relation linking a to N in u. There will be cases, however, where there is no real referent a and then a will be replaced either by a corresponding fictional object drawn from a fictional universe or by the contradictory infon **0**.

26.2.1.2 Resource situations and modes of presentation

The idea of resource situations comes from Barwise & Perry (1983). At least one resource situation r_u is associated with every noun phrase via u. Every resource situation associated with a name N contains an infon like $\langle\langle R^c;\ a;\ N;\ u \rangle\rangle$ *by definition*. That is, $r_u(N)$ is a resource situation induced by u only if $r_u(N) \models \langle\langle R^c;\ a;\ N;\ u \rangle\rangle$ for some R^c, a relative to N and u. Not all resource situations involve such causal chains; some may be perceptual or may depend on an attribute.

Frege (1892/1980) conflated the distinct notions of sense and mode of presentation. Sense is analogous to conventional meaning, which for names is just the property P^N = *named* N. The mode of presentation is what enables a speaker or addressee to fix a reference and is analogous to a resource situation.[7]

[6]Incidentally, so can Putnam's (1975) division of linguistic labor.

[7]I myself conflated these two notions in Parikh (2010: 291) and posited the hybrid notion of an indexical property which was a property that contained a link to an external object. But such a combination is not required as the link can be housed within a resource situation. If we wish to avoid Kripke's (1980: pages 68–70) noncircularity condition even for the property taken by itself, we can reinstate the hybrid notion. See footnote 14 in Chapter 3.

26.2.1.3 The extension function

In Section 2.1 I had defined the following function:

$$(34)\quad e(P,s) = \begin{cases} a & \text{if there is exactly one object } a \text{ having } P \text{ in } s \\ \{x \mid s \vDash \langle\!\langle\, P;\ x \,\rangle\!\rangle\} & \text{otherwise} \end{cases}$$

In the context of causal chains and resource situations for names, I substitute P^N and r_u for P and s:

$$(35)\quad e(P^N, r_u) = \begin{cases} a, & \text{if there is exactly one object } a \text{ having } P^N \text{ in } r_u \\ \{x \mid r_u \vDash \langle\!\langle\, P^N;\ x \,\rangle\!\rangle\}, & \text{otherwise} \end{cases}$$

$e(P^N, r_u)$ will provide the content of names like HESPERUS and PHOSPHORUS: $e(P^{\text{HESPERUS}}, r_u)$ = Venus and $e(P^{\text{PHOSPHORUS}}, r'_u)$ = Venus where r_u, r'_u are the two corresponding resource situations.[8]

26.2.1.4 Functions treated intensionally and extensionally

Generally, functions are treated extensionally. For example, $f(x) = x + 2$ is taken to be equal to $g(x) = x + 4 - 2$ because the "external" values both functions yield are the same. However, they are not intensionally equal because their other properties such as the time taken to compute their outputs may differ. So functions taken intensionally may be different objects even though extensionally they are the same.

In the same way, we can treat the extension function above intensionally or extensionally. Thus, $e(P^{\text{HESPERUS}}, r_u)$ and $e(P^{\text{PHOSPHORUS}}, r'_u)$ are different objects intensionally because their first and second arguments differ. Specifically for our purposes, the causal chains their second arguments contain are different. Extensionally, however, they are the same and just have the value Venus.

26.2.2 Setting up the analysis

Now consider an utterance situation u involving a conversation about astronomy where $\mathcal{A}$ says "Hesperus is Phosphorus" to $\mathcal{B}$. As our interest is in solving

[8] I am omitting certain complications involving conditioned infons that are considered in Parikh (2010) as the present definitions suffice for our purposes.

Frege's puzzle, we can abstract from many features of the Communication Game Γ_u and focus just on the locutionary Semantic Constraint. We can also ignore the mythological meanings of "Hesperus" and "Phosphorus" as they would be eliminated by the Flow Constraint.

This may seem like a trivial step as theorists do it almost unconsciously but the difference is that now one knows precisely what is being abstracted from so that the missing details can be filled in if required. It is just this kind of awareness that has led many researchers to consider utterances instead of sentences as they realized that certain contextual aspects of utterances could not be ignored.

26.2.3 Semantic Constraint

Let the sentence uttered be $\varphi = \varphi_1\varphi_2\varphi_3$ where φ_1 = HESPERUS, φ_2 = IS, and φ_3 = PHOSPHORUS. Also, $R^=$ is just the relation of equality.

(36) HESPERUS:
Referential Use: $\varphi_1 \longrightarrow P^{\varphi_1} \xrightarrow{u} e(P^{\varphi_1}, r_u) = \text{Venus} = \sigma_1$

(37) IS:
Predicative Use: $\varphi_2 \longrightarrow P^{\varphi_2} = R^= \xrightarrow{u} \langle\langle\, R^= \,\rangle\rangle = \sigma_2$[9]

(38) PHOSPHORUS:
Referential Use: $\varphi_3 \longrightarrow P^{\varphi_3} \xrightarrow{u} e(P^{\varphi_3}, r'_u) = \text{Venus} = \sigma_3$

26.2.4 The analysis

As we have abstracted from the different possible uses of names and also from the lexical ambiguities in φ, the Flow Constraint involves trivial games and yields the following locutionary content for the utterance:

$$\sigma^\ell = \langle\langle\, R^=;\ e(P^{\varphi_1}, r_u);\ e(P^{\varphi_3}, r'_u) \,\rangle\rangle$$

Since we can treat the arguments of this infon both intensionally and extensionally, it follows immediately that the content carries nontrivial information because only the extensional equality holds. Indeed, both items of information

[9]Actually, there are two other uses of IS that I discuss in Parikh (2010), the so-called auxiliary and membership uses. The current predicative use just picks out the equality relation. However, it is conditioned by the requirement that it refer to a time period overlapping with the time of utterance and also belonging to a finite time interval.

that make up the utterance's cognitive significance are conveyed. There is no particular reason why the different causal chains in r_u and r'_u should involve the same object Venus. This requires, in fact, an empirical discovery.

In an utterance of "Hesperus is Phosphorus" it is usually the intensional content that is intended as the extensional content is trivial. The elimination of the latter would also be effected by the Flow Constraint.

If we consider the sentence "Hesperus is Hesperus" instead then its content would be:

$$\tau^{\ell} = \langle\!\langle\, R^{=};\ e(P^{\text{HESPERUS}}, r_u);\ e(P^{\text{HESPERUS}}, r_u)\,\rangle\!\rangle$$

where $\varphi_1 = \varphi_3 =$ HESPERUS and the two resource situations are now the same. This equality is trivial as expected and, indeed, it holds both extensionally and intensionally.

Suppose someone says, "Woodchucks are groundhogs" in some situation u. Now this is a case where both terms have the same conventional meaning because the property of being a woodchuck is the same as the property of being a groundhog. That is, their senses are the same so Frege's solution will not work. But the resource situations will be different and contain different causal chains terminating in the same kind of rodent, and this allows nontrivial information to be learned from the identity.[10]

As I said at the start of this section, one item of information is that the first is the heavenly body seen in the evening and the second is the heavenly body seen in the morning and they are the same. The first part of this information is contained in r_u both via the causal chain $\langle\!\langle\, R^c_1;$ Venus; HESPERUS; $u\,\rangle\!\rangle$ and via background information about when Hesperus is observed (i.e. in the evening). This latter fact is also embedded in the relation R^c_1 because the causal link involves Hesperus in the evening. The second part of this information is contained in r'_u via the different causal chain $\langle\!\langle\, R^c_2;$ Venus; PHOSPHORUS; $u\,\rangle\!\rangle$ and via different other facts such as when Phosphorus is observed (i.e. in the morning). And the third part, the identity relation, is of course contributed by IS. So one nontrivial significance of "Hesperus is Phosphorus" is clearly available to the speaker and addressee. (The same kind of difference occurs in the resource situations for woodchucks and groundhogs. One term may have been encountered in the field and the other in a dictionary. And so on.) The second item involving the coreference of the two names is also readily available from the distinct causal chains. So my solution is able to account for both items of cognitive significance.

[10] I would like to thank Ed Zalta for extended conversations about this topic over the years. See Zalta (1983: Chapter VI) and Zalta (1988: Chapters 9–12).

Many proposals exhibit some difference between "Hesperus is Phosphorus" and "Hesperus is Hesperus," either that the two names co-refer or that the two modes of presentation result in the same object, without, however, deriving both and so miss the former's full cognitive significance.

Imagine something like an Agatha Christie murder mystery in which the characters' identities have been scrambled and in solving the mystery, Hercule Poirot grandly announces, "Smith isn't Smith but Jones is Jones." Here, the first conjunct isn't contradictory but informative and the second conjunct isn't trivial but also informative just like "Hesperus is Phosphorus." My approach easily handles such utterances because the relevant resource situations for each occurrence of the two names are different and contain different facts. Most other solutions, whether Fregean or of the direct reference kind, would find them hard to account for.

26.2.5 Wider considerations

Direct reference theorists and others who treat the contents of identity statements only extensionally are led to conflate the trivial and informative contents mentioned above, and from there, inescapably, to the startling conclusion that they are synthetic necessary truths known a posteriori. Kripke's (1980) discovery that it is important to distinguish the metaphysical notion of necessity from the epistemological notion of aprioricity and also from the linguistic notion of analyticity is certainly valid. In general, though, as Kripke shows himself, whether or not there are contingent a priori truths or necessary a posteriori truths is an open question and one that should have little to do with how we use our language; it should have to do with how the world is and with how agents might learn its many truths. For example, we may discover through computational means that a certain large number is a prime: this is then a necessary a posteriori truth.

Perhaps our most important means of acquiring information is through linguistic communication; not only that, much of the information we have could not be had without language. But the closer scrutiny initiated by Kripke himself should caution us against taking one of the relatively simple uses of language as evidence for rather counterintuitive claims about what are essentially metaphysical and epistemological concepts.

Possibly, mathematical identities such as $7+5 = 12$ can also be seen in this dual intensional-extensional way and then their being informative would follow. Of course, they still remain necessary truths unlike "Hesperus is Phosphorus" and the question of how necessary truths can be informative – like learning that a large number is a prime – is a different one.

26.3 Defining reference

I briefly point out an immediate consequence of the framework: it becomes easy to define reference. For Equilibrium Semantics the notion is broader than the standard notion because the content of any legitimate expression that is uttered is a referent. In an utterance of "Bill ran," both "Bill" and "ran" refer.

There are two natural notions of reference, "subjective reference" corresponding to speaker meaning and "reference" to communication. The former allows a referent not to be communicated whereas the latter involves the successful transmission of the referent to the addressee. The expression α below can be subsentential or a whole sentence and σ is a partial or full infon.

Definition 26.1. $\mathcal{A}$ subjectively refers to σ by uttering α in u if and only if there is a situation c such that $\mathcal{A}$ means $p = (c \models \sigma)$ by uttering α in u.

Definition 26.2. $\mathcal{A}$ refers to σ by uttering α in u if and only if there is a situation c and an addressee $\mathcal{B}$ such that $\mathcal{A}$ communicates $p = (c \models \sigma)$ to $\mathcal{B}$ by uttering α in u.

Here, $p = (c \models \sigma)$ is a partial or full proposition based on the partial or full infon σ. When $\sigma = \langle\langle\, a \,\rangle\rangle = a$ where a is an individual, we get the familiar case of singular reference.

Appendix A: Situated games

The basic task of this Appendix is to derive the locutionary semantic map $g_u : \mathcal{L} \longrightarrow \mathcal{G}$ introduced in Part III from first principles. This derivation will provide the mathematical foundation for our informal discussion of situated games of partial information. It will also enable readers to see how all the other games such as syntactic games of partial information, illocutionary games of partial information, locutionary and illocutionary global games, Setting Games, Content Selection Games, Generation Games, Interpretation Games, Communication Games, and Language Games can be defined either analogously or by further construction.

Recall that the elements of situation theory have already been spelled out as fully as required for the purposes of this book in Section 2.1. Certain basic facts relating to agents and language were also identified in Section 2.2 and Chapter 4. I will take this material as given here.

The key innovation in my presentation, based on Parikh (1987b; 1990; 2001; 2010), is that games are constructed from situation-theoretic objects just as numbers can be constructed from sets. The motivation for preferring such an embedding of game theory in situation theory is discussed in Appendix A of *Language and Equilibrium* and so will not be repeated here.

A.1 The background

Let the sets of infons, situations, and propositions be $\mathcal{I}$, $\mathcal{S}$, and **PROP** respectively. These sets are assumed to be finite and represent the range of infons, situations, and propositions of interest relative to a small part of reality called the environment $\mathcal{E}$ within the overall informational space or ontology $\mathcal{O}$. There are various operations on these entities partially described in Section 2.1 and these should be borne in mind as well.

While any action can be communicative in the right context, I restrict my focus to linguistic communication. The Conventional sub-Constraint **C** and the Referential sub-Constraint **R** originate in communication as discussed in Part V and together provide the set of possible (semantic) contents of an utterance. Another way of describing the task of the Appendix is to say that the Flow Constraint **F** will be developed mathematically assuming **C** and **R** are given.

Since $\mathcal{A}$'s actions are utterances of sentences, they can be modeled by ordered pairs $\langle \varphi, u \rangle$, where φ is a sentence in $\mathcal{L}$. As the utterance situation u is a parameter, it is convenient to simplify the notation by specifying utterances by their sentences alone.

$\mathcal{B}$'s actions are interpretations of $\mathcal{A}$'s utterances, and can be modeled as pairs $\langle \sigma, c \rangle$, where σ is an infon and c the situation described partially by the infon. Once again, this can be simplified to just σ, because c is held fixed for all the infons that serve as interpretations relative to the utterance situation u. I have taken infons rather than propositions as the relevant contents. Clearly, both are transmitted in an utterance.

Initially, lexical games $g_u(\varphi_i)$ will be constructed, where φ_i is a word in the sentence φ. The corresponding possible contents of φ_i obtained from **C** and **R** will be denoted by σ_i^y where y stands for zero or more primes. To avoid the proliferation of generic indices, I choose to replace i with just the number 1 with the understanding that the same construction would apply mutatis mutandis to all the words in the sentence uttered. Thus, φ_i is replaced by φ_1 and σ_i^y by σ_1^y.

It is convenient to define a function m_u that maps elements of $\mathcal{L}$ into their possible contents σ_1^y relative to u. $m_u(\varphi_1) = \{\sigma_1^y \mid y = 0, 1, 2, \dots \text{primes}\}$ is assumed to be finite and generated via **C** and **R**. Indeed, m_u is just the composition of the conventional and referential maps suitably defined. An added advantage of using this composed function is that the construction of $g_u(\varphi_1)$ can be applied beyond language to other symbol systems and other domains like social systems if m_u is reinterpreted for those other contexts. In other words, **C** and **R** and their corresponding maps may be viewed as just the *theory* of m_u for the case of language. Further, different parts of language such as nouns and verbs require different theories for each corresponding part of the Referential Constraint **R**, something that was considered in greater detail in my previous book. A similar function m'_u that maps elements of $\mathcal{L}$ into their possible parse trees based on the Syntactic Constraint may be assumed if we wish to consider syntax but I will just abstract from it here.

The way the game-theoretic model works is that given an utterance situation u, the speaker $\mathcal{A}$ forms an intention to convey a (partial) content σ_1 (like *Bill Smith*) as part of a full utterance and chooses a possibly ambiguous locution φ_1 (like "Bill") such that $\sigma_1 \in m_u(\varphi_1)$. If σ_1 is the only member of $m_u(\varphi_1)$, a trivial game ensues. If φ_1 is ambiguous then a nontrivial game results. In either case, the number of initial situations in the game equals the number of possible contents of φ_1.

A.2 Some prior basic elements

Some basic relations, infons, situations, and functions pertaining to communication are now defined.

The first is the relation itc of intending to convey something. This is a three-place relation with a speaker who does the intending, an infon the speaker intends to convey, and an addressee to whom the content is addressed. Such intentions can be either explicit or implicit. The second relation hu is the relation of having uttered something. Its first argument is a speaker and its second argument is a parameter that is anchored to an expression. However, instead of writing $\langle\!\langle\, hu;\ \mathcal{A};\ \dot{\varphi}\,\rangle\!\rangle$ with $f(\dot{\varphi}) = \varphi_1$, I will write simply $\langle\!\langle\, hu;\ \mathcal{A};\ \varphi_1\,\rangle\!\rangle$ to keep the notation uncluttered. The third relation is the relation hi of an addressee's having interpreted an utterance as communicating something. This is a three-place relation with arguments an addressee, an utterance, and an infon. From these three relations, corresponding infons and situations can be constructed.

We start with a specification of the possible initial situations in a game of partial information.

Definition A.1. $s_{1^y} \subset u$ and $s_{1^y} \vDash \{\langle\!\langle\, itc, \mathcal{A}, \sigma_1^y, \mathcal{B}\,\rangle\!\rangle\}$, where $\sigma_1^y \in m_u(\varphi_1)$.

This says simply that the initial situation s_{1^y} is a subset of the utterance situation u and supports an infon representing the appropriate intention $\mathcal{A}$ would have to convey σ_1^y to $\mathcal{B}$.

I had earlier used t to refer to various parse trees but I will be abstracting from syntax here so I will use it instead for situations that result from an initial situation after $\mathcal{A}$ has uttered something.

Definition A.2.

$$act_{\mathcal{A}} : \mathcal{S} \times \mathcal{L} \longrightarrow \mathcal{S}$$
$$act_{\mathcal{A}}(s, \varphi_1) = s \cup \{\langle\!\langle\, hu, \mathcal{A}, \varphi_1\,\rangle\!\rangle\}.$$
$$act_{\mathcal{B}} : \mathrm{ran}(act_{\mathcal{A}}) \times \mathcal{I} \longrightarrow \mathcal{S}$$
$$act_{\mathcal{B}}(t, \tau) = t \cup \{\langle\!\langle\, hi, \mathcal{B}, \varphi_1, \tau\,\rangle\!\rangle\} \text{ where } \varphi_1 = 2^{nd}[act_{\mathcal{A}}^{-1}(t)].$$

Here $\mathrm{ran}(act_{\mathcal{A}})$ is just the range of the function $act_{\mathcal{A}}$. These two functions give general descriptions of the consequences of $\mathcal{A}$'s uttering something in a situation and of $\mathcal{B}$'s subsequently interpreting it. All that happens is that the initial situations s get augmented first with $\mathcal{A}$'s utterance and next with $\mathcal{B}$'s interpretation. It should be easy to see that if s were taken to be one of the initial situations s_{1^y} above and if appropriate utterances and interpretations were specified, then we would have the basic mechanism for generating the game tree via the updates given by $act_{\mathcal{A}}$ and $act_{\mathcal{B}}$. Thus, having started with the elements of situation theory and with some basic relations pertaining to communication, games of partial information will naturally follow.

This is partly where using situation theory enables us to make the information available to each agent at each stage explicit. Uttering φ_1 in s has the consequence that $\langle\langle\, hu, \mathcal{A}, \varphi_1 \,\rangle\rangle$ gets added to s, a fact that is explicitly available to both agents. Without this kind of underlying construction, this fact would remain implicit. This kind of articulation is especially useful in contexts where artificial agents that communicate have to be designed because such information can then be used in making situated and strategic inferences.

We first remind ourselves of the form of games of incomplete information which will be the approximate target of our construction.

A.3 Games of incomplete information

As I have said, traditionally, games are approached directly. Although Harsanyi (1967) was the first person to define games of incomplete information, he did so in normal or strategic form. What we need is their extensive form definition because that is where various aspects of the choice structure are made transparent. This was made clear by Kuhn (1953) who was the first to define extensive form games for games of perfect and imperfect information. Kreps & Wilson (1982) extended Kuhn's definition to games of incomplete information. We use their formulation of a game of incomplete information as a tuple of sets and functions as the approximate goal of the situation-theoretic construction. This goal is approximate because games of partial information are different in certain respects from incomplete information games.

Kreps and Wilson Kreps & Wilson (1982) define a game of incomplete information directly as an extended tuple $\langle T, \prec ; ACT, act ; N, \eta ; P ; H ; v \rangle$. T is a set of nodes and $\prec$ is a partial ordering on T that makes the pair $\langle T, \prec \rangle$ a tree (more precisely, a forest). ACT is a set of actions and act is a function that maps every noninitial node of $\langle T, \prec \rangle$ into some action in ACT. This is intended to be the action that leads to this node. N is a set of agents (or players) and η is a mapping from the set of nonterminal nodes onto N. η establishes whose turn it is to act. P is a vector of probabilities on the set of initial nodes. This much of the tuple gives a tree with decision nodes connected by actions with an agent identified for each decision node. H is a partition on T that consists of subpartitions, one for each player. It is meant to capture the information sets of each agent, the sets of decision nodes of an agent that cannot be distinguished by the agent. Accordingly, each agent's subpartition is a collection of those sets of nodes that are his or her information sets. Finally, v is the payoff function, a mapping from the terminal nodes into the set $\mathbb{R}$ of real numbers.

A.4 Games of partial information

I now construct locutionary semantic partial information games as described in Chapter 7. This is a straightforward matter of defining the elements of the tuple above one by one.

A.4.1 Situations and choices

As was said above in Section A.1, given an utterance situation u, the speaker $\mathcal{A}$ forms an intention to convey a (partial) content σ_1 (like *Bill Smith*) as part of a full utterance and chooses a possibly ambiguous locution φ_1 (like "Bill") such that $\sigma_1 \in m_u(\varphi_1)$. The first element of the tuple that has to be constructed is the tree of possible situations and to do this we start with the initial situations identified in Definition A.1 in Section A.2. Then we define $\mathcal{A}$'s choice sets in each of these initial situations which just contain φ_1. Next, we generate the intermediate situations that result from φ_1 based on $act_{\mathcal{A}}$ from Definition A.2 and then identify $\mathcal{B}$'s choices in these situations. Finally, we define the terminal situations that follow from $\mathcal{B}$'s interpretive actions based on $act_{\mathcal{B}}$. This collection of situations forms a tree under the subset ordering.

Definition A.3.

$$T_0 = \{s_{1^y} \mid \sigma_1^y \in m_u(\varphi_1)\}$$

$$C_{\mathcal{A}} : T_0 \longrightarrow \mathcal{P}(\mathcal{L})$$
$$C_{\mathcal{A}}(s_{1^y}) = \{\varphi_1\}$$

$$T_1 = \{act_{\mathcal{A}}(s_{1^y}, \varphi_1) \mid s_{1^y} \in T_0\}$$

$$C_{\mathcal{B}} : T_1 \longrightarrow \mathcal{P}(\mathcal{I})$$
$$C_{\mathcal{B}}(t) = m_u(\varphi_1)$$

$$T_2 = \{act_{\mathcal{B}}(t, \tau) \mid t \in T_1,\ \tau \in C_{\mathcal{B}}(t)\}$$

$$T = \bigcup_i T_i$$

Each $C_{\mathcal{A}}(s)$ and $C_{\mathcal{B}}(t)$ are the same and are $\mathcal{A}$'s and $\mathcal{B}$'s choice sets in the relevant situations.

Proposition A.1. $\langle T, \subset \rangle$ is a "tree" (more accurately, forest).

Since the more basic building blocks of situations and actions have been strung together piece by piece, the fact that $\langle T, \subset \rangle$ is a tree is something that can now be *proved* rather than simply assumed.

Definition A.4. $\forall t, t' \in T_0,\ t \equiv_{\mathcal{A}} t'$ iff $t = t'$
$\forall t, t' \in T_1,\ t \equiv_{\mathcal{B}} t'$

Proposition A.2. $\equiv_{\mathcal{A}}, \equiv_{\mathcal{B}}$ are equivalence relations.

These equivalence relations capture the relevant epistemic properties of the two agents as follows.

Proposition A.3.

- For all $t, t' \in T_1$, $t \equiv_{\mathcal{B}} t'$ implies $C_{\mathcal{B}}(t) = C_{\mathcal{B}}(t')$.
- For all $t, t' \in T_1$, $t \equiv_{\mathcal{B}} t'$ implies $t \not\subset t'$.
- For all $t, t' \in T_1$, $t \equiv_{\mathcal{B}} t'$ implies $\eta(t) = \eta(t')$.

The first statement says that $\mathcal{B}$ has the same choices at each of various equivalent situations. This is important because if the choices weren't the same the agent could use that information to distinguish between epistemically equivalent situations, a contradiction. The second statement says that of two equivalent situations one cannot precede the other. This again makes intuitive sense because if such precedence were possible, the agent would know it, and it would make the situations epistemically distinguishable. The last statement requires Definition A.8 made below and says simply that the same agent has to act in all equivalent situations. The corresponding properties for $\mathcal{A}$ are even more trivially true. The key thing to note is that all these properties can now be *proved* from more basic assumptions.

A.4.2 Actions

The third element of the game tuple above, the set of actions in the game, is nothing but the union of all the choice sets in Definition A.3. This gathering of all the actions into a single set is just to maintain conformity with Kreps and Wilson's tuple, so that the game is rendered in a familiar form.

Definition A.5. $ACT = [\bigcup_{s \in T_0} C_{\mathcal{A}}(s)] \cup [\bigcup_{t \in T_1} C_{\mathcal{B}}(t)] = \{\varphi_1\} \cup m_u(\varphi_1)$

The map *act* assigns an appropriate set of actions in ACT to each noninitial situation in T.

Definition A.6. $act : T_1 \cup T_2 \longrightarrow ACT$

$$
\begin{aligned}
act(t) &= 2^{nd}[act_{\mathcal{A}}^{-1}(t)] = \varphi_1 \quad \text{if } t \in T_1 \\
&= 2^{nd}[act_{\mathcal{B}}^{-1}(t)] = \sigma_1^y \quad \text{if } t \in T_2 \text{ for some appropriate } y
\end{aligned}
$$

act maps a situation into the action that brings it about. The reason for labeling situations or nodes in the game tree with actions in this manner is because $\langle T, \subset \rangle$ is a tree and this means that each noninitial situation has a single action that generates it.

A.4.3 Agents

Definition A.7. $N = \{\mathcal{A}, \mathcal{B}\}$ is the set of agents.

The function η below determines whose turn it is to act.

Definition A.8. $\eta : T_0 \cup T_1 \longrightarrow N$

$$
\begin{aligned}
\eta(t) &= \mathcal{A} \quad \text{if } t \in T_0 \\
&= \mathcal{B} \quad \text{if } t \in T_1
\end{aligned}
$$

This is just a formal way of saying that $\mathcal{A}$ is the speaker and $\mathcal{B}$ the addressee.

A.4.4 Prior probabilities

The next item in the tuple is the prior probabilities. As we have seen, they are the most complex part of the game-theoretic structure because it is through them that each local game is connected with all the other local games, both locutionary and illocutionary, that materialize when a sentence is uttered. To enable the definition below, assume that the whole sentence uttered is $\varphi = \varphi_1\varphi_2 \ldots \varphi_n$ for some n, as was done in Section 8.1. In addition, remember that x_{-1} is a variable that stands for any of the vectors formed by taking all the combinations of the possible semantic contents of locutions other than φ_1. As I am abstracting from syntax here, we do not consider the vector of possible syntactic contents y.

Definition A.9. $P : m_u(\varphi_1) \times m_u(\varphi_2) \times \ldots \times m_u(\varphi_n) \times \{u\} \longrightarrow [0,1]$ such that $\sum_y P(\sigma_1^y \,|\, x_{-1}; u) = 1$ with $\sigma_1^y \in m_u(\varphi_1)$ and $x_{-1} \in m_u(\varphi_2) \times \ldots \times m_u(\varphi_n)$.

This is a key reason why games of partial information differ from games of incomplete information. Incomplete information games have a single pre-given probability distribution; as was first explained in Section 7.3, partial information games have multiple probability distributions for each local game. These form a *third* set of *strategic* choice variables along with the choice of utterance and interpretation from each $C_{\mathcal{A}}(s)$ and $C_{\mathcal{B}}(t)$. This strategic characteristic is made possible by the presence of the conditioning variables x_{-1} because each different instantiation of these variables creates a different probability distribution $P(\sigma_1^y \mid x_{-1}; u)$ and the agents have to *choose* one distribution from among many based on whether all such choices are in *global* equilibrium or not, the latter just being the equilibrium of the global game.

A.4.5 Information sets

Definition A.4 and Propositions A.2 and A.3 allow us to specify the information sets of each agent. They are *information* sets precisely because they have been shown to have the appropriate epistemic properties.

Definition A.10.

$$h_{\mathcal{A}} : T_0 \longrightarrow \mathcal{P}(T)$$
$$h_{\mathcal{A}}(t) = \{t' \in T_0 \mid t' \equiv_{\mathcal{A}} t\} = \{t\}$$
$$H_{\mathcal{A}} = \{h_{\mathcal{A}}(t) \mid t \in T_0\}$$

$$h_{\mathcal{B}} : T_1 \longrightarrow \mathcal{P}(T)$$
$$h_{\mathcal{B}}(t) = \{t' \in T_1 \mid t' \equiv_{\mathcal{B}} t\} = T_1$$
$$H_{\mathcal{B}} = \{h_{\mathcal{B}}(t) \mid t \in T_1\}$$

$$H = H_{\mathcal{A}} \cup H_{\mathcal{B}}$$

Again, I observe that the situation-theoretic construction makes it possible to build information sets from the simpler objects $\equiv_{\mathcal{A}}$ and $\equiv_{\mathcal{B}}$ rather than just defining and imposing them outright. It provides an explanation of why they arise from the more basic epistemic properties of the game.

A.4.6 Payoffs

At one level, the payoffs are the easiest elements to define as all that is needed is two real-valued functions defined on the terminal situations in T_2, with the understanding that each function encodes the same underlying preferences up to positive affine transformations. This would be the most general case. However, it is desirable to constrain these functions to respect the inequalities introduced in Section 7.3.

$$a_{\mathcal{A}} > c_{\mathcal{A}}; \quad a'_{\mathcal{A}} > c'_{\mathcal{A}}; \quad a_{\mathcal{B}} > c_{\mathcal{B}}; \quad a'_{\mathcal{B}} > c'_{\mathcal{B}}$$

These inequalities pertain to games where there are only two initial situations. For games with more initial situations, note that the additional terminal situations that result all belong with instances where $\mathcal{B}$ has erred in her interpretation. All such situations are mapped into the lowest level of payoffs $c_{\mathcal{A}}$ or its cognates.

These inequalities were derived from the underlying preferences agents have for successful communication on the one hand and for minimizing effort or cost on the other. This is why it makes sense to constrain the payoff functions we define to obey these inequalities.

A little manipulation of prior constructs is required to express the condition that a terminal situation represents a correct interpretation or an incorrect one. This is not difficult to do but I rehearse it a bit to make it more readable.

To identify terminal situations where the interpretation is correct, we require that the situation $t \in T_2$ satisfies:

$$1^{st}[act_{\mathcal{A}}^{-1}(1^{st}[act_{\mathcal{B}}^{-1}(t)])] = s_{1^y} \in T_0$$

$$2^{nd}[act_{\mathcal{B}}^{-1}(t)] = \sigma_1^y$$

Both these conditions flow from Definitions A.1, A.1, and A.6. The first condition simply traces the path back from the terminal situation t to the initial situation s_{1^y} in T_0 by Definition A.2 and the second condition requires that the interpretation leading to the terminal situation be σ_1^y, which is the content intended in s_{1^y} by Definition A.1, thereby ensuring that the interpretation is the right one. These two conditions jointly capture the requirements for the highest payoffs like $a_{\mathcal{A}}$ and its cognates. Similar conditions with suitable modifications are required to identify payoffs like $c_{\mathcal{A}}$ and its cognates.

Definition A.11. $v_{\mathcal{A}} : T_2 \longrightarrow \mathbb{R}$
$v_{\mathcal{B}} : T_2 \longrightarrow \mathbb{R}$

such that $\forall t, t' \in T_2$

if

$1^{st}[act_{\mathcal{A}}^{-1}(1^{st}[act_{\mathcal{B}}^{-1}(t)])] = s_{1^y} \in T_0$,
$2^{nd}[act_{\mathcal{B}}^{-1}(t)] = \sigma_1^y$,

and

$1^{st}[act_{\mathcal{A}}^{-1}(1^{st}[act_{\mathcal{B}}^{-1}(t')])] = s_{1^y} \in T_0$,
$2^{nd}[act_{\mathcal{B}}^{-1}(t')] \neq \sigma_1^y$

then

$v_{\mathcal{A}}(t) > v_{\mathcal{A}}(t')$ and $v_{\mathcal{B}}(t) > v_{\mathcal{B}}(t')$

$v = (v_{\mathcal{A}}, v_{\mathcal{B}})$

The terminal situations t' are those where the interpretation is incorrect. In general, there will be many terminal situations t' when the number of initial situations in the game is greater than two.

Both $v_{\mathcal{A}}$ and $v_{\mathcal{B}}$ obey the same constraints although the actual payoff numbers may be different for the two agents. v collects both these functions in an ordered pair.

A.4.7 The game tuple

All the elements required for a local game are now at hand.

Definition A.12. $g_u(\varphi_1) = \langle T, \subset; ACT, act; N, \eta; P; H; v \rangle$

The full local game is of course $\langle g_u(\varphi_1), \mathcal{I}_g \rangle$, where $\mathcal{I}_g$ represents common knowledge between $\mathcal{A}$ and $\mathcal{B}$ of $g_u(\varphi_1)$. In general, $\mathcal{I}_g$ represents the *information structure* of the *strategic interaction* and it may range from no shared information to full common knowledge.[1] This is part of the advantage gained by the situation-theoretic construction because it permits a natural extension of games where common knowledge obtains to more general strategic interactions where common knowledge may not obtain.

Earlier, in Section A.1, I chose to avoid undue generality by fixing the utterance to be φ_1. It is now time to relax this constraint by letting the utterance vary freely over $\mathcal{L}$.[2]

Definition A.13. $g_u(\alpha)$ where $\alpha \in \mathcal{L}$ is defined analogously by replacing φ_1 with α and making all other corresponding changes in each component of the game tuple above. $g_u(e) = g_e$ and $g_u(0) = g_0$.[3] $\mathcal{G}$ is defined to be just the collection of all the tuples so obtained. This gives us the map $g_u : \mathcal{L} \longrightarrow \mathcal{G}$.[4]

[1] See Parikh (2010: Section 3.3.4) for further discussion.

[2] Remember that $\mathcal{L}$ is the free monoid generated from the vocabulary $\mathcal{W}$ by the special concatenation operation $\circ_G$ together with the empty string e and the zero element 0 as mentioned in Section 4.1. Needless to say, $m_u(e) = \{\mathbf{1}\}$, $m_u(0) = \{\mathbf{0}\}$, and $m_u(\alpha) = \{\mathbf{0}\}$ when $\alpha \in \mathcal{L}$ is a grammatical but meaningless expression.

[3] See Section 10.1.

[4] I have already specified in Section A.2 that we have an anchor f such that $f(\dot{\varphi}) = \alpha$ in order to secure that α is uttered in u. It is necessary to make the game map a total function. We can do this as follows:

$$g_{u[f]}(\alpha) = \begin{cases} g & \text{if } f(\dot{\varphi}) = \alpha \\ g_\alpha & \text{otherwise} \end{cases}$$

This completes the construction of the local game map g_u, the primary goal of the Appendix.

Theorem A.4. $g_u(\alpha)$ where $\alpha \in \mathcal{L}$ is a local game of partial information.

As observed above, instead of simply *defining* or legislating that the tuple above is a game, the situation-theoretic construction makes this fact a *consequence* derived from more basic assumptions. It is possible to *prove* that the tuple is a local game of partial information.

Games of partial information are in many ways similar to but slightly more general than games of incomplete information. The particular subclass of partial information games we are concerned with – those that apply to communication – are in fact similar to but again slightly more general than the subclass of incomplete information games known in economics as signaling games.

We can now introduce syntactic contents into these games and define syntactic local games by using a similar function m'_u that maps elements of $\mathcal{L}$ into their possible parse trees based on the Syntactic Constraint.

A.5 The product of two games

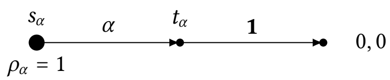
Now that games of partial information have been built and their collection $\mathcal{G}$ identified, it is possible to define their product $\otimes$. What has to be done is to start with two tuples, one for each multiplicand, and then construct the tuple for their product, building each component one by one from the components of the multiplicands. Along the way, it is necessary to prove that each component does in fact have all the required properties. For example, the payoff inequalities have to be preserved in the product when they are defined to be the sum of the corresponding payoffs in the multiplicands.

Once this is done, it becomes easy to show that $(\mathcal{G}, \otimes)$ is a monoid and that $g_u(\alpha_1 \circ \alpha_2) = g_u(\alpha_1) \otimes g_u(\alpha_2)$, facts that are required in Section 10.1 to prove that $g_u : \mathcal{L} \longrightarrow \mathcal{G}$ is an isomorphism in order to establish the universality of games of partial information in semantics.

I leave this construction and the proofs of the corresponding facts as an exercise for the reader.

---

where g is the game defined above when α is uttered in u and where g_α is the game below:



A.6 The compact form

A compact form for the locutionary global game can now be defined as discussed in Section 8.1. This is straightforward to do as all its components have already been constructed.

A.7 Solution concepts

The next step would be to define the notion of a strategy for such games and examine various solution concepts. Because all the local games generated by an utterance are interconnected through the prior probabilities, it becomes necessary to extend the standard definitions of a strategy and the corresponding equilibria. This involves *two* things: one is to provide for the strategic nature of the prior probability distributions and the other is to build a definition of *global* equilibrium based on the solution concept one starts with, presumably Pareto-Nash equilibrium. I have shown how to do this informally in Section 8.1 and the definitions do not pose any special obstacles.

Whatever solution concepts we employ, they make possible the construction of the second basic map of Equilibrium Semantics $f_u : \mathcal{G} \longrightarrow \mathcal{I}$, whose existence and uniqueness were established by the universality result of Section 10.1. Once we have both maps g_u and f_u, we have essentially derived our fundamental global fixed point equations: Equations 6 and 7 from Theorems 8.5 and 8.6.

This completes a mathematical rendering of the core framework of Equilibrium Semantics from first principles. It can be extended to all the other games mentioned at the start of the Appendix in a natural way.

References

Adajian, Thomas. 2012. The definition of art. In Edward N. Zalta (ed.), *The Stanford encyclopedia of philosophy*, Winter 2012. http://plato.stanford.edu/archives/win2012/entries/art-definition/.

Adler, Jonathan E. & Lance J. Rips (eds.). 2008. *Reasoning: Studies of human inference and its foundations*. Cambridge, UK: Cambridge University Press.

Allais, Maurice. 1953. Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école Americaine. *Econometrica* 21. 503–546.

Allen, James. 1995. *Natural language understanding*. Menlo Park, CA: Benjamin Cummings.

Ariely, Dan. 2008. *Predictably irrational: The hidden forces that shape our decisions*. New York: HarperCollins.

Arrow, Kenneth J. 1982. Risk perception in psychology and economics. *Economic Inquiry* 20(1).

Arrow, Kenneth J. & Gerard Debreu. 1954. Existence of a competitive equilibrium for a competitive economy. *Econometrica* 22(3). 265–290.

Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation*. Cambridge, UK: Cambridge University Press.

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

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

Austin, J. L. 1961. *Philosophical papers*. London: Oxford University Press. Edited by J. O. Urmson and G. J. Warnock.

Austin, J. L. 1975. *How to do things with words*. 2nd edn. Cambridge, MA: Harvard University Press. Edited by J. O. Urmson and Marina Sbisa.

Austin, J. L. 1961/1979a. How to talk: Some simple ways. In *Philosophical papers*, 3rd edn., 134–153. Oxford: Oxford University Press. Edited by J. O. Urmson and G. J. Warnock.

Austin, J. L. 1961/1979b. Performative utterances. In *Philosophical papers*, 3rd edn. Oxford: Oxford University Press. Edited by J. O. Urmson and G. J. Warnock.

Austin, J. L. 1961/1979c. Truth. In *Philosophical papers*, 3rd edn., 117–133. Oxford: Oxford University Press. Edited by J. O. Urmson and G. J. Warnock.

Bach, Kent. 1987a. On communicative intentions: A reply to Recanati. *Mind and Language* 2. 141–154.

Bach, Kent. 1987b. *Thought and reference*. Oxford: Clarendon Press.

Bach, Kent. 2001. You don't say. *Synthese* 128. 15–44.

Bach, Kent. 2012a. Meaning and communication. In D. Fara & G. Russell (eds.), *The Routledge companion to the philosophy of language*, 79–90. London: Routledge.

Bach, Kent. 2012b. Saying, meaning, and implicating. In K. Allan & K. Jaszczolt (eds.), *The Cambridge handbook of pragmatics*, 47–68. Cambridge, UK: Cambridge University Press.

Bach, Kent & R. M. Harnish. 1979. *Linguistic communication and speech acts*. Cambridge, MA: The MIT Press.

Barsalou, L. W. 1983. Ad hoc categories. *Memory and Cognition* 11. 211–227.

Barsalou, L. W. 1987. The instability of graded structure: Implications for the nature of concepts. In U. Neisser (ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge, UK: Cambridge University Press.

Barthes, Roland. 1957. *Mythologies*. France: Editions de Seuil.

Barthes, Roland. 1977. The death of the author. In *Image-music-text*, 142–148. New York: Hill & Wang.

Barwise, Jon. 1988. Three views of common knowledge. In Moshe Vardi (ed.), *Theoretical aspects of reasoning about knowledge II*. Los Altos, CA: Morgan Kaufmann.

Barwise, Jon. 1989a. On the model theory of common knowledge. In *The situation in logic*. Stanford: CSLI Publications.

Barwise, Jon. 1989b. Situations, facts, and true propositions. In *The situation in logic*. Stanford: CSLI Publications.

Barwise, Jon. 1989c. Situations, sets, and the axiom of foundation. In *The situation in logic*. Stanford: CSLI Publications.

Barwise, Jon. 1989d. *The situation in logic*. Stanford: CSLI Publications.

Barwise, Jon & John Etchemendy. 1986. *The liar*. London: Oxford University Press.

Barwise, Jon & John Etchemendy. 1991. Information, infons, and inference. In Robin Cooper, Kuniaki Mukai & John Perry (eds.), *Situation theory and its applications*, vol. 1, 33–78. Stanford: CSLI Publications.

Barwise, Jon & John Perry. 1983. *Situations and attitudes*. Cambridge, MA: The MIT Press.

Benz, Anton. 2010. Implicatures and the recognition of speaker intentions in a game-theoretic model. In Luis A. Pérez Miranda & Aitor Izagirre Madariaga (eds.), *Advances in cognitive science: Learning, evolution, and social action*, 15–

34. Bilbao: University of the Basque Country Press. Proceedings of the ILCLI international workshop on cognitive science, IWCogSc-10.
Benz, Anton. 2012. Errors in pragmatics. *Journal of Logic, Language, and Information* 21. 97–116.
Benz, Anton, Christian Ebert, Gerhard Jäger & Robert van Rooij (eds.). 2011. *Language, games, and evolution: Trends in current research on language and game theory*. Berlin: Springer.
Benz, Anton, Gerhard Jäger & Robert van Rooij (eds.). 2006. *Game theory and pragmatics*. U.K.: Palgrave Studies in Pragmatics, Language & Cognition, Palgrave Macmillan.
Bernoulli, D. 1738/1954. Exposition of a new theory of the measurement of risk. *Econometrica* 22. 23–36.
Bianchi, C. (ed.). 2004. *The semantics/pragmatics distinction*. Stanford University: CSLI Publications.
Blume, Andreas & Oliver Board. 2013. Language barriers. *Econometrica* 81(2). 781–812.
Blume, Andreas & Oliver Board. 2014. Intentional vagueness. *Erkenntnis* 79. 855–899.
Bordwell, David. 1989. *Making meaning: Inference and rhetoric in the interpretation of cinema*. Cambridge, MA: Harvard University Press.
Briggs, Rick. 1985. Knowledge representation in Sanskrit and artificial intelligence. *The AI Magazine* 6(1). 32–39.
Brown, P. & Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. New York: Cambridge University Press.
Burge, Tyler. 1973. Reference and proper names. *Journal of Philosophy* 73.
Burge, Tyler. 1975. On knowledge and convention. *Philosophical Review* 84(2). 249–255.
Camerer, Colin F. 2003. *Behavioral game theory*. New Jersey: Princeton University Press.
Cappelen, Herman & Ernie Lepore. 2005a. *Insensitive semantics: A defense of semantics minimalism and speech act pluralism*. Oxford: Blackwell.
Cappelen, Herman & Ernie Lepore. 2005b. Radical and moderate pragmatics: Does meaning determine truth conditions? In Zoltán G. Szabó (ed.), *Semantics versus pragmatics*, 45–71. Oxford: Clarendon Press.
Carruthers, Peter. 2004. On being simple minded. *American Philosophical Quarterly* 41(3). 205–220.
Carston, Robyn. 2004. Truth-conditional content and conversational implicature. In C. Bianchi (ed.), *The semantics/pragmatics distinction*, 65–100. Stanford University: CSLI Publications.

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

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cialdini, Robert B. 2006. *Influence: The psychology of persuasion*. Revised edition. New York: Harper Business.

Clark, Herbert H. & Catherine R. Marshall. 1981. Definite reference and mutual knowledge. In Joshi, Webber & Sag (eds.), *Elements of discourse understanding*. Cambridge University Press.

Clark, Robin. 2011. *Meaningful games*. Cambridge, MA: MIT Press.

Clark, Robin & Prashant Parikh. 2007. Game theory and discourse anaphora. *Journal of Logic, Language, and Information* 16. 265–282.

Cohen, L. Jonathan. 1985. A problem about ambiguity in truth-conditional semantics. *Analysis* 45. 129–135.

Cohen, L. Jonathan. 1986. How is conceptual innovation possible? *Erkenntnis* 25. 221–238.

Connolly, William E. 1993. *The terms of political discourse*. 3rd edn. New Jersey: Princeton University Press.

Crawford, V. P. & J. Sobel. 1982. Strategic information transmission. *Econometrica* 50. 1431–1451.

Crowley, Terry & Claire Bowern. 2010. *An introduction to historical linguistics*. 4th edn. Oxford: Oxford University Press.

de Jaegher, Kris. 2003. A game-theoretic rationale for vagueness. *Linguistics and Philosophy* 26. 637–659.

de Saussure, Ferdinand. 1960. *Course of general linguistics*. Trans. by W. Baskin. London: Peter Owen.

de Saussure, Ferdinand. 1916/1972. *Cours de linguistique générale*. Paris: Éditions Payot.

Deacon, Terrence. 1997. *The symbolic species: The co-evolution of language and the brain*. New York: W. W. Norton.

Deutsch, Eliot & Ron Bontekoe (eds.). 2007. *A companion to world philosophies*. Oxford: Blackwell Publishing.

Devitt, Michael. 2013a. Three methodological flaws of linguistic pragmatism. In Carlo Penco & Filippo Domaneschi (eds.), *What is said and what is not: The semantics/pragmatics interface*, 285–300. Stanford: CSLI Publications.

Devitt, Michael. 2013b. What makes a property "semantic"? In A. Capone, F. Lo Piparo & M. Carapezza (eds.), *Perspectives on pragmatics and philosophy*, 87–112. Berlin: Springer.

Devlin, Keith. 1991. *Logic and information.* Cambridge, UK: Cambridge University Press.

Dewatripont, Mathias & Jean Tirole. 2005. Modes of communication. *Journal of Political Economy* 113(6). 1217–1238.

Dickinson, Emily. 1993. *Dickinson: Poems.* New York: Everyman's Library Pocket Poets, Alfred A. Knopf, Inc.

Dretske, Fred I. 1981. *Knowledge and the flow of information.* Cambridge, MA: The MIT Press.

Dretske, Fred I. 2003. Experience as representation. *Philosophical Issues, A supplement to Noûs* 13. 67–82.

Dretske, Fred I. 2009. Information-theoretic semantics. In Brian McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *The Oxford handbook of philosophy of mind.* Oxford University Press.

Dummett, Michael. 1993. *Frege: Philosophy of language.* 2nd edn. Cambridge, MA: Harvard University Press.

Dummett, Michael. 1996. *Origins of analytical philosophy.* Cambridge, MA: Harvard University Press.

Eagleton, Terry. 2008. *Literary theory: An introduction.* Anniversary edition. Minneapolis: University of Minnesota Press.

Égré, Paul & Nathan Klinedinst. 2011. *Vagueness and language use.* Houndsmills: Palgrave Macmillan.

Elbourne, P. 2008. The argument from binding. *Philosophical Perspectives* 22. 89–110.

Fagin, Ronald, Joseph Y. Halpern, Yoram Moses & Moshe Y. Vardi. 2003. *Reasoning about knowledge.* Cambridge, MA: The MIT Press.

Farrell, Joseph. 1987. Cheap talk, coordination, and entry. *RAND Journal of Economics* 18. 34–39.

Farrell, Joseph. 1993. Meaning and credibility in cheap-talk games. *Games and Economic Behavior* 5. 514–531.

Farrell, Joseph & Matthew Rabin. 1996. Cheap talk. *Journal of Economic Perspectives* 10(3). 103–118.

Fernández, Eva M. & Helen Smith Cairns. 2011. *Fundamentals of psycholinguistics.* Oxford: Wiley-Blackwell.

Franke, Michael. 2009. *Signal to act: Game theory in pragmatics.* University of Amsterdam dissertation.

Frege, Gottlob. 1892/1980. On sense and meaning. In Peter Geach & Max Black (eds.), *Translations from the philosophical writings of Gottlob Frege*, 56–78. Basil Blackwell Ltd.

Frisson, Steven. 2009. Semantic underspecification in language processing. *Language and Linguistics Compass* 3. 111–127.

Gallie, W. B. 1956. Essentially contested concepts. *Proceedings of the Aristotelian Society, New Series* 56. 167–198.

Gärdenfors, Peter. 2000. *Conceptual spaces: The geometry of thought*. Cambridge, MA: The MIT Press.

Garfield, Jay L. (ed.). 1995. *The fundamental wisdom of the middle way: Nāgārjuna's Mūlamadhyamakakārikā*. Oxford: Oxford University Press.

Geanakoplos, John, David Pearce & Ennio Stacchetti. 1989. Psychological games and sequential rationality. *Games and Economic Behavior* 1. 60–79.

Geertz, Clifford. 1973. *The interpretation of cultures: Selected essays*. New York: Basic Books.

Gilbert, Margaret. 1981. Game theory and "convention". *Synthese* 46(1). 41–93.

Gintis, Herbert. 2000. *Game theory evolving*. New Jersey: Princeton University Press.

Glimcher, Paul. 2004. *Decisions, uncertainty, and the brain: The science of neuroeconomics*. Cambridge, MA: The MIT Press.

Grice, Paul. 1957. Meaning. *Philosophical Review* 66. 377–388.

Grice, Paul. 1968. Utterer's meaning, sentence-meaning and word-meaning. *Foundations of Language* 4. 1–18.

Grice, Paul. 1969. Utterer's meaning and intentions. *Philosophical Review* 78. 147–177.

Grice, Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, vol. 3, 41–58. New York: Academic Press.

Grice, Paul. 1989a. Logic and conversation. In *Studies in the way of words*, 1–143. Cambridge, MA: Harvard University Press.

Grice, Paul. 1989b. Meaning revisited. In *Studies in the way of words*, 283–303. Cambridge, MA: Harvard University Press.

Grice, Paul. 1989c. *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Gronendijk, J. & M. Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14. 39–100.

Habermas, Jürgen. 1976/1998. What is universal pragmatics? In *On the pragmatics of communication*, 21–103. Cambridge, MA: The MIT Press. Edited by Maeve Cooke.

Habermas, Jürgen. 1988/1998. Toward a critique of the theory of meaning. In *On the pragmatics of communication*, 277–306. Cambridge, MA: The MIT Press. Edited by Maeve Cooke.

Halpern, Joseph Y. & Leandro Rego. 2006. Extensive games with possibly unaware players. In *Proceedings of the fifth international joint conference on autonomous agents and multiagent systems*, 744–751.

Halpern, Joseph Y. & Leandro Rego. 2007. Generalized solution concepts in games with possibly unaware players. In *Proceedings of the eleventh conference on theoretical aspects of rationality and knowledge*.

Hampton, James A. 2007. Typicality, graded membership, and vagueness. *Cognitive Science* 31(3). 355–384.

Hampton, James A., B. Aina, J. M. Andersson, H. Mirza & S. Parmar. 2012. The Rumsfeld Effect: The unknown unknown. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38. 340–355.

Hampton, James A., Z. Estes & C. L. Simons. 2005. Comparison and contrast in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31. 1459–1476.

Hampton, James A. & Martin L. Jönsson. 2012. Typicality and compositionality: The logic of combining vague concepts. In M. Werning, W. Hintzen & E. Machery (eds.), *Oxford handbook of compositionality*, 385–402. Oxford: Oxford University Press.

Harman, Gilbert. 1974. Review of Stephen Schiffer's "Meaning". *Journal of Philosophy* 71(7). 224–229.

Harman, Gilbert. 1986. *Change in view*. Cambridge, MA: The MIT Press.

Harsanyi, J. C. 1967. Games with incomplete information played by Bayesian players. *Management Science* 14. 159–182, 320–334, 486–502.

Harsanyi, J. C. & R. Selten. 1988. *A general theory of equilibrium selection in games*. Cambridge, MA: The MIT Press.

Heidegger, Martin. 1927/1962. *Being and time*. New York: Harper & Row. Translated by John Macquarrie and Edward Robinson.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. University of Massachusetts, Amherst dissertation.

Hendricks, Vincent & John Symons. 2015. Epistemic logic. In Edward N. Zalta (ed.), *The Stanford encyclopedia of philosophy*, Fall 2015. https://plato.stanford.edu/archives/fall2015/entries/logic-epistemic.

Hirsch, E. D. 1967. *Validity in interpretation*. New Haven: Yale University Press.

Hirsch, E. D. 1978. *The aims of interpretation*. Chicago: University of Chicago Press.

Hobbs, Jerry R. 2004. Abduction in natural language understanding. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 724–741. Oxford: Blackwell Publishing.

Horn, Laurence R. 1972. On the semantic properties of the logical operators in English. *Indiana University Linguistics Club.*

Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Deborah Schiffrin (ed.), *Meaning, form, and use in context (GURT '84).* Washington: Georgetown University Press.

Horn, Laurence R. 2014. Trivial pursuits: On being orderly. Presented at CUNY pragmatics workshop: Relevance, game theory, and communication.

Hugly, Philip & Charles Sayward. 1979. A problem about conversational implicature. *Linguistics and Philosophy* 3. 19–25.

Husserl, Edmund. 1913/1967. The thesis of the natural standpoint and its suspension. In Joseph J. Kockelmans (ed.), *Phenomenology: The philosophy of Edmund Husserl and its interpretation*, 68–80. New York: Anchor Books. From Edmund Husserl, "Ideas: General introduction to phenomenology," Volume I, sections 27–32, translated by W. R. Boyce Gibson (George Allen and Unwin), 1931.

Huttegger, Simon M. 2007. Evolution and the explanation of meaning. *Philosophy of Science* 74. 1–27.

Iseminger, Gary (ed.). 1992. *Intention and interpretation.* Philadelphia: Temple University Press.

Israel, David & John Perry. 1989. What is information? In P. Hanson (ed.), *Information, language, and cognition: Vancouver studies in cognitive science*, vol. 1. University of British Columbia Press.

Jäger, Gerhard. 2012. Game theory in semantics and pragmatics. In C. Maienborn, P. Portner & K. von Heusinger (eds.), *Semantics: An international handbook of natural language meaning*, vol. 3, 2487–2516. Berlin: de Gruyter.

Jäger, Gerhard, Lars Koch-Metzger & Frank Riedel. 2011. Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals. *Games and Economic Behavior.*

Joshi, A. K. 1985. Tree-adjoining grammars: How much context sensitivity is required to provide reasonable structural descriptions? In D. Dowty, L. Karttunen & A. Zwicky (eds.), *Natural language parsing*, 206–250. Cambridge, UK: Cambridge University Press.

Joshi, A. K. & Y. Schabes. 1997. Tree-adjoining grammars. In G. Rosenberg & A. Salomaa (eds.), *Handbook of formal languages*, 69–123. Berlin: Springer.

Jurafsky, Daniel. 2004. Pragmatics and computational linguistics. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 578–604. Oxford: Blackwell Publishing.

Jurafsky, Daniel & James H. Martin. 2000. *Speech and language processing*. New Jersey: Prentice Hall.

Jurafsky, Daniel & James H. Martin. 2009. *Speech and language processing*. 2nd edn. New Jersey: Prentice Hall.

Kahneman, Daniel, Paul Slovic & Amos Tversky (eds.). 1982. *Judgement under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kahneman, Daniel & Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2). 263–291.

Kamp, Hans. 1981. A theory of truth and semantic representation. In J. A. G. Groenendijk, T. M. V. Janssen & M. B. J. Stokhof (eds.), *Formal methods in the study of language*, 277–322. Amsterdam: Mathematical Centre Tracts 135.

Kartik, Navin. 2009. Strategic communication with lying costs. *Review of Economic Studies* 76. 1359–1395.

Kawamoto, Alan H. 1993. Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language* 32. 474–516.

Keefe, R. & P. Smith. 1996. *Vagueness: A reader*. Cambridge, MA: The MIT Press.

Klir, George J., Ute St. Clair & Bo Yuan. 1997. *Fuzzy set theory: Foundations and applications*. New Jersey: Prentice Hall.

Kneale, W. C. 1962. Universality and necessity. *British Journal for the Philosophy of Science* 12.

Korta, Kepa & John Perry. 2011. *Critical pragmatics*. Cambridge, UK: Cambridge University Press.

Kreps, David. 1986. Out of equilibrium beliefs and out of equilibrium behaviour. Working Paper, Graduate School of Business, Stanford University.

Kreps, David & Robert Wilson. 1982. Sequential equilibrium. *Econometrica* 50. 863–894.

Kripke, Saul A. 1977. Speaker's reference and semantic reference. In Peter A. French, Theodore E. Uehling Jr. & Howard K. Wettstein (eds.), *Contemporary perspectives in the philosophy of language*, 6–27. University of Minnesota Press.

Kripke, Saul A. 1980. *Naming and necessity*. Cambridge, MA: Harvard University Press.

Kripke, Saul A. 1982. *Wittgenstein on rules and private language: An elementary exposition*. Cambridge, MA: Harvard University Press.

Kuhn, H. W. 1953. Extensive games and the problem of information. In H. W. Kuhn & A. W. Tucker (eds.), *Contributions to the theory of games*, vol. 2, 193–216. Princeton: Princeton University Press.

Labov, William. 2001. *Principles of linguistic change: Social factors*. Vol. 2. Oxford: Blackwell Publishers.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge, UK: Cambridge University Press.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University Press.

Lewis, David. 1972. General semantics. In Donald Davidson & Gilbert Harman (eds.), *Semantics of natural language*, 169–218. Dordrecht: Reidel.

Lewis, David. 1975. Language and languages. In Keith Gunderson (ed.), *Language, mind, and knowledge*, 3–35. University of Minnesota Press.

Lewis, David. 1981. Scorekeeping in a language game. In *Philosophical papers*, vol. 1, 233–249. Oxford University Press.

Lipman, Barton. 2009. Why is language vague? *Boston University Working Paper*.

Manning, Christopher & Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press.

Mante, Valerio, David Sussillo, Krishna V. Shenoy & William T. Newsome. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503. 78–84.

Matilal, B. K. & P. K. Sen. 1988. The context principle and some Indian controversies over meaning. *Mind* 97. 73–97.

McCready, Eric. 2012. Emotive equilibria. *Linguistics and Philosophy* 35. 243–283.

McGuinness, Brian (ed.). 1982. *Wittgenstein and his times*. Oxford: Blackwell.

Medin, D. L. & M. M. Schaffer. 1978. Context theory of classification learning. *Psychological Review* 85. 207–238.

Mohanty, J. N. 2000. *Classical indian philosophy*. Maryland: Rowman & Littlefield Publishers.

Mohanty, J. N. 2007. A history of Indian philosophy. In Eliot Deutsch & Ron Bontekoe (eds.), *A companion to world philosophies*, 24–48. Oxford: Blackwell Publishing.

Morris, Charles. 1938. Foundations of the theory of signs. In *International encyclopedia of unified science*, vol. 1. Chicago, IL: University of Chicago Press.

Murphy, Gregory L. 2004. *The big book of concepts*. Cambridge, MA: The MIT Press.

Murphy, M. Lynne. 2010. *Lexical meaning*. Cambridge, UK: Cambridge University Press.

Myerson, Roger B. 1989. Credible negotiation statements and coherent plans. *Journal of Economic Theory* 48. 264–303.

Myerson, Roger B. 1991. *Game theory: Analysis of conflict*. Boston: Harvard University Press.

Nash, J. 1951. Non-cooperative games. *Annals of Mathematics* 54. 286–295.

Neale, Stephen. 1992. Paul Grice and the philosophy of language. *Linguistics and Philosophy* 15. 509–559.

Nosofsky, R. M. 1992. Exemplars, prototypes, and similarity rules. In A. Healy, S. Kosslyn & R. Shiffrin (eds.), *From learning theory to connectionist theory: Essays in honor of W. K. Estes*, vol. 1, 149–168. Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. & T. J. Palmeri. 1997. An exemplar-based random walk model of speeded categorization. *Psychological Review* 104. 266–300.

Onifer, William & David A. Swinney. 1981. Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition* 9(3). 225–236.

Parikh, Prashant. 1987a. Conventional meaning and equilibrium. Unpublished.

Parikh, Prashant. 1987b. *Language and strategic inference*. Stanford University dissertation.

Parikh, Prashant. 1990. Situations, games, and ambiguity. In Robin Cooper, Kuniaki Mukai & John Perry (eds.), *Situation theory and its applications*, vol. 1, 449–470. Stanford: CSLI Publications.

Parikh, Prashant. 1992. A game-theoretic account of implicature. In Yoram Moses (ed.), *Theoretical aspects of reasoning about knowledge*, 85–94. California: Morgan Kaufmann Publishers, Inc.

Parikh, Prashant. 2000. Communication, meaning, and interpretation. *Linguistics and Philosophy* 23. 185–212.

Parikh, Prashant. 2001. *The use of language*. Stanford University: CSLI Publications.

Parikh, Prashant. 2006a. Pragmatics and games of partial information. In Anton Benz, Gerhard Jäger & Robert van Rooij (eds.), *Game theory and pragmatics*, 101–122. New York: Palgrave Macmillan.

Parikh, Prashant. 2006b. Radical semantics: A new theory of meaning. *Journal of Philosophical Logic* 35. 349–391.

Parikh, Prashant. 2010. *Language and equilibrium*. Cambridge, MA: The MIT Press.

Peirce, Charles Sanders. 1867–1913/1955. Abduction and induction. In Justus Buchler (ed.), *Philosophical writings of Peirce*, 150–156. New York: Dover Books.

Peirce, Charles Sanders. 1867–1913/1991. *Writings on semiotic*. North Carolina: University of North Carolina Press. Edited by James Hoopes.

Pietarinen, Ahti-Veikko (ed.). 2007. *Game theory and linguistic meaning*. Amsterdam: Elsevier Publications.

Pinker, Steven, Martin A. Nowak & James J. Lee. 2008. The logic of indirect speech. *Proceedings of the National Academy of Sciences* 105(3). 833–838.

Popper, Karl. 1935/2002. *The logic of scientific discovery*. London: Routledge Classics.

Portner, Paul. 2004. The semantics of imperatives within a theory of clause types. In R. Young (ed.), *SALT XIV*, 235–252. Ithaca, NY: Cornell University.

Puglisi, Andrea, Andrea Baronchelli & Vittorio Loreto. 2008. Cultural route to the emergence of linguistic categories. *Proceedings of the National Academy of Sciences* 105(23). 7936–7940.

Putnam, Hilary. 1975. The meaning of "meaning". In *Mind, language, and reality: Philosophical papers*, vol. 2. Cambridge, UK: Cambridge University Press.

Pylkkänen, L., R. Llinás & G. L. Murphy. 2006. The representation of polysemy: Meg evidence. *Journal of Cognitive Neuroscience* 18. 97–109.

Quine, W. V. O. 1960. *Word and object*. Cambridge, MA: The MIT Press.

Quine, W. V. O. 1969a. Ontological relativity. In *Ontological relativity and other essays*. New York: Columbia University Press.

Quine, W. V. O. 1969b. Speaking of objects. In *Ontological relativity and other essays*. New York: Columbia University Press.

Rabiner, Lawrence R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2). 257–286.

Raiffa, Howard. 1970. *Decision analysis: Introductory lectures on choices under uncertainty*. Reading, Massachusetts: Addison-Wesley.

Recanati, François. 1986. On defining communicative intentions. *Mind and Language* 1. 213–242.

Recanati, François. 2004. *Literal meaning*. Cambridge, UK: Cambridge University Press.

Recanati, François. 2010. *Truth-conditional pragmatics*. Oxford: Oxford University Press.

Rizzolatti, G. & C. Sinigaglia. 2008. *Mirrors in the brain: How we share our actions and emotions*. Oxford: Oxford University Press.

Roberts, Craige. 1996. Information structure: Towards an integrated formal theory of pragmatics. In Jae Hak Yoon & Andreas Kathol (eds.), *OSUWPL volume 49: Papers in semantics*, 91–136. Ohio: The Ohio State University Department of Linguistics.

Rodd, J., G. Gaskell & W. Marslen-Wilson. 2002. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language* 46. 245–266.

Rosch, E. 1978. Principles of categorization. In E. Rosch & B. B. Lloyd (eds.), *Cognition and categorization*, 27–48. Hillsdale, NJ: Erlbaum.

Rosch, E. & C. B. Mervis. 1975. Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology* 7. 573–605.

Ross, Ian. 2006. *Games interlocutors play: The unification of implicature*. University of Pennsylvania dissertation.

Ross, J. F. 1981. *Portraying analogy*. Cambridge, UK: Cambridge University Press.

Rubinstein, Ariel. 1998. *Modeling bounded rationality*. Cambridge, MA: The MIT Press.

Rubinstein, Ariel. 2000. *Economics and language*. Cambridge, UK: Cambridge University Press.

Russell, Bertrand. 1919. Descriptions. In *Introducton to mathematical philosophy*, 167–180. George Allen & Unwin Publishers Ltd.

Schelling, T. C. 1960. *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Schiffer, Stephen. 1972. *Meaning*. Oxford: Oxford University Press.

Schiffer, Stephen. 2003. *The things we mean*. Oxford: Oxford University Press.

Schiffer, Stephen. 2010. Vague properties. In Richard Dietz & Sebastiano Moruzzi (eds.), *Cuts and clouds, vagueness, its nature, and its logic*, 109–130. Oxford: Oxford University Press.

Searle, John R. 1969. *Speech acts*. Cambridge, UK: Cambridge University Press.

Searle, John R. 1978. Literal meaning. *Erkenntnis* 13. 207–224.

Searle, John R. 1979. Metaphor. In A. Ortony (ed.), *Metaphor and thought*, 92–123. Cambridge, UK: Cambridge University Press.

Sen, Amartya. 2003. Sraffa, Wittgenstein, and Gramsci. *Journal of Economic Literature* 41(4). 1240–1255.

Shannon, Claude E. & Warren Weaver. 1949. *The mathematical theory of communication*. Urbana & Chicago: University of Illinois Press.

Shepard, R. N. 1987. Toward a universal law of generalization for psychological science. *Science* 237. 1317–1323.

Shieber, Stuart. 1986. *An introduction to unification-based approaches to grammar*. Stanford University: CSLI Publications.

Shoham, Yoav & Kevin Leyton-Brown. 2009. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. New York: Cambridge University Press.

Simon, Herbert A. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69. 99–118.

Simon, Herbert A. 1956. Rational choice and the structure of the environment. *Psychological Review* 63. 129–138.

Skyrms, Brian. 2010. *Signals: Evolution, learning, and information*. Oxford: Oxford University Press.

Smith, E. E. & D. L. Medin. 1981. *Categories and concepts*. Cambridge, MA: Harvard University Press.

Spence, Michael. 1973. Job market signaling. *Quarterly Journal of Economics* 87. 355–374.

Spence, Michael. 2001. Signaling in retrospect and the informational structure of markets. Nobel prize lecture at http://nobelprize.org/nobel-prizes/economics/laureates/2001/spence-lecture.pdf.

Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and cognition.* Oxford: Blackwell Publishing.

Sperber, Dan & Deirdre Wilson. 2002. Pragmatics, modularity and mind-reading. *Mind & Language* 17. 3–23.

Sperber, Dan & Deirdre Wilson. 2008. A deflationary account of metaphors. In R. Gibbs (ed.), *The Cambridge handbook of metaphor and thought*, 84–105. Cambridge, UK: Cambridge University Press.

Spiegelberg, Herbert. 1981. *The phenomenological movement: A historical introduction.* Berlin: Springer.

Stalnaker, Robert C. 1984. *Inquiry*. Cambridge, MA: The MIT Press.

Stalnaker, Robert C. 1991. The problem of logical omniscience, I. *Synthese* 89. 425–440. Reprinted in Stalnaker 1999.

Stalnaker, Robert C. 1998. On the representation of context. *Journal of Logic, Language, and Information* 7. Reprinted in Stalnaker 1999.

Stalnaker, Robert C. 1999. *Context and content.* Oxford: Oxford University Press.

Stalnaker, Robert C. 2002. Common ground. *Linguistics and Philosophy* 25. 701–721.

Stalnaker, Robert C. 2006. Saying and meaning, cheap talk and credibility. In Anton Benz, Gerhard Jäger & Robert van Rooij (eds.), *Game theory and pragmatics*, 83–100. New York: Palgrave Macmillan.

Stanley, Jason. 2007. *Language in context.* Oxford: Oxford University Press.

Steedman, Mark. 2001. *The syntactic process (Language, speech, and communication).* Cambridge, MA: The MIT Press–A Bradford Book.

Strawson, P. F. 1956. On referring. In Antony Flew (ed.), *Essays in conceptual analysis*, 21–52. London: MacMillan & Co. Ltd.

Strawson, P. F. 1964. Intention and convention in speech acts. *Philosophical Review* 73. 439–460.

Sutton, Charles & Andrew McCallum. 2006. An introduction to conditional random fields for relational learning. In Lise Getoor & Ben Taskar (eds.), *Introduction to statistical relational learning*. The MIT Press.

Swinney, David A. 1979. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18. 645–659.

Szabó, Zoltán G. (ed.). 2005. *Semantics versus pragmatics.* Oxford: Clarendon Press.

Taylor, Charles. 2013. Retrieving realism. In Jospeh K. Schear (ed.), *Mind, reason, and being-in-the-world*, 61–90. London: Routledge.

Taylor, Charles. 1980/1999. Theories of meaning. In *Human agency and language: Philosophical papers*, vol. 1, 248–292. Cambridge, UK: Cambridge University Press.

Taylor, Charles. 1981/1999. The concept of a person. In *Human agency and language: Philosophical papers*, vol. 1, 97–114. Cambridge, UK: Cambridge University Press.

Taylor, Charles. 1983/1999. Hegel's philosophy of mind. In *Human agency and language: Philosophical papers*, vol. 1, 77–96. Cambridge, UK: Cambridge University Press.

Taylor, Charles. 1985/1999. *Human agency and language: Philosophical papers*. Vol. 1. Cambridge, UK: Cambridge University Press.

Taylor, Charles. 1977/1999a. Self-interpreting animals. In *Human agency and language: Philosophical papers*, vol. 1, 45–76. Cambridge, UK: Cambridge University Press.

Taylor, Charles. 1977/1999b. What is human agency? In *Human agency and language: Philosophical papers*, vol. 1, 15–44. Cambridge, UK: Cambridge University Press.

Taylor, Charles. 1978/1999. Language and human nature. In *Human agency and language: Philosophical papers*, vol. 1, 215–247. Cambridge, UK: Cambridge University Press.

Thomason, Richmond H. 1990. Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 325–364. Cambridge, MA: The MIT Press.

Tomasello, Michael. 2010. *Origins of human communication*. Cambridge, MA: The MIT Press.

Traugott, Elizabeth C. & Richard B. Dasher. 2007. *Regularity in semantic change*. Cambridge, UK: Cambridge University Press.

Travis, Charles. 1996. Meaning's role in truth. *Mind* 105. 451–466.

van Bekkum, Wout, Jan Houben, Ineke Sluiter & Kees Versteegh. 1997. *The emergence of semantics in four linguistic traditions: Hebrew, Sanskrit, Greek, Arabic*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

van Rooij, Robert. 2004. Signaling games select Horn strategies. *Linguistics and Philosophy* 27. 493–527.

Vlaev, Ivo, Nick Chater, Neil Stewart & Gordon D. A. Brown. 2011. Does the brain calculate value? *Trends in Cognitive Sciences* 15(11). 546–554.

von Neumann, John & Oskar Morgenstern. 1944/1947. *Theory of games and economic behavior*. 2nd edn. Princeton: Princeton University Press.

Wall, Robert. 1972. *Introduction to mathematical linguistics*. New Jersey: Prentice Hall.

Warglien, Massimo & Peter Gärdenfors. 2013. Semantics, conceptual spaces, and the meeting of minds. *Synthese* 190. 2165–2193.

Watson, Joel. 2002. *Strategy: An introduction to game theory*. New York: W. W. Norton & Company.

Whewell, William. 1840. *The philosophy of the inductive sciences, founded upon their history*. London: John W. Parker. In two volumes.

Williamson, Timothy. 1994. *Vagueness*. London: Routledge.

Wilson, Deirdre. 2011. Parallels and differences in the treatment of metaphor in relevance theory and cognitive linguistics. *Intercultural Pragmatics* 8(2). 177–196.

Wilson, Deirdre & Robyn Carston. 2007. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In Noel Burton-Roberts (ed.), *Pragmatics*, 230–259. London: Palgrave.

Wilson, Deirdre & Dan Sperber. 1986. On defining relevance. In Richard Grandy & Richard Warner (eds.), *Philosophical grounds of rationality*, 243–258. Oxford: Clarendon Press.

Wilson, Deirdre & Dan Sperber. 2004. Relevance theory. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 607–632. Oxford: Blackwell Publishing.

Wimsatt, William K. & Monroe C. Beardsley. 1946. The intentional fallacy. *Sewanee Review* 54. 468–488.

Wittgenstein, Ludwig. 1953/1968. *Philosophical investigations*. Trans. by G. E. M. Anscombe. 3rd edn. New York: Macmillan Publishing Company, Inc.

Wittgenstein, Ludwig. 1921/1961. *Tractatus logico-philosophicus*. Trans. by D. F. Pears & B. F. McGuinness. London: Routledge & Kegan Paul.

Yang, Charles. 2009. Population structure and language change. Unpublished.

Zaefferer, Dietmar. 1977. Understanding misunderstandings: A proposal for an explanation of reading choices. *Journal of Pragmatics* 1. 329–346.

Zalta, Edward N. 1983. *Abstract objects: An introduction to axiomatic metaphysics* (Synthese Library 160). Dordrecht: D. Reidel.

Zalta, Edward N. 1988. *Intensional logic and the metaphysics of intensionality*. Cambridge, MA: The MIT Press.

Zermelo, Ernst. 1913. Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In *Proceedings of the fifth international congress of mathematicians*, 501–504. Cambridge University Press.

Name index

Subject index

Did you like this book?

This book was brought to you for free

Communication and content

Communication and content presents a comprehensive and foundational account of meaning based on new versions of situation theory and game theory. The literal and implied meanings of an utterance are derived from first principles assuming little more than the partial rationality of interacting agents. New analyses of a number of diverse phenomena – a wide notion of ambiguity and content encompassing phonetics, syntax, semantics, pragmatics, and beyond, vagueness, convention and conventional meaning, indeterminacy, universality, the role of truth in communication, semantic change, translation, Frege's puzzle of informative identities – are developed. Communication, speaker meaning, and reference are defined. Frege's context and compositional principles are generalized and reconciled in a fixed-point principle, and a detailed critique of Grice, several aspects of Lewis, and some aspects of the Romantic conception of meaning are offered. Connections with other branches of linguistics, especially psycholinguistics, sociolinguistics, historical linguistics, and natural language processing, are explored.

The book will be of interest to scholars in philosophy, linguistics, artificial intelligence, and cognitive science. It should also interest readers in related fields like literary and cultural theory and the social sciences.

ISBN 978-3-96110-198-6

9 783961 101986