ELSEVIER

*Current Research in the Semantics/
Pragmatics Interface:*

Volume 18

GAME THEORY AND LINGUISTIC MEANING

Editor: Ahti-Veikko Pietarinen

CRiS/PI

Game Theory and Linguistic Meaning

Current Research in the Semantics/Pragmatics Interface

Series Editors:
K.M. Jaszczolt, University of Cambridge, UK
K. Turner, University of Brighton, UK

New releases:

MARIA ALONI, ALASTAIR BUTLER AND PAUL DEKKER (Eds.)
Questions in Dynamic Semantics
"Building on Karttunen's and Hamblin's earlier work, Groenendijk and Stokhof in the 1980's established a framework for the semantics of questions and the pragmatics of answerhood that led to a burst of advances. This collection, centered on one of the topics that have made the "Amsterdam school" justly famous, displays the great strengths of their approach and illustrates recent theoretical advances in dynamic formal semantics and pragmatics. It will immediately become essential reading for scholars and students in this field."
Barbara H. Partee, University of Massachusetts, USA

Other titles in this series:

TURNER (ed.)	The Semantics/Pragmatics Interface from Different Points of View
JASZCZOLT	Discourse, Beliefs and Intentions: Semantic Defaults and Propositional Attitude Ascription
GEURTS	Presuppositions and Pronouns
JASZCZOLT (ed.)	The Pragmatics of Propositional Attitude Reports
PEETERS (ed.)	The Lexicon-Encyclopedia Interface
PAPAFRAGOU	Modality: Issues in the Semantics-Pragmatics Interface
LEEZENBERG	Contexts of Metaphor
NÉMETH & BIBOK (eds.)	Pragmatics and the Flexibility of Word Meaning
BRAS & VIEU (eds.)	Semantic and Pragmatic Issues in Discourse and Dialogue
GUTIÉRREZ-REXACH (ed.)	From Words to Discourse: Trends in Spanish Semantics and Pragmatics
KAMP & PARTEE (eds.)	Context-Dependence in the Analysis of Linguistic Meaning
PEREGRIN (ed.)	Meaning: The Dynamic Turn
DOBROVOL'SKIJ & PIIRAINEN	Figurative Language: Cross-cultural and Cross-linguistic Perspectives
WEDGWOOD	Shifting the Focus: From Static Structures to the Dynamics of Interpretation
BULTINCK	Numerous Meanings: The Meaning of English Cardinals and the Legacy of Paul Grice
VON HEUSINGER & TURNER (eds.)	Where Semantics Meets Pragmatics
ALONI *et al.* (eds.)	Questions in Dynamic Semantics

For more information visit: www.elsevier.com/locate/series/crispi

Game Theory and Linguistic Meaning

EDITED BY

Ahti-Veikko Pietarinen
University of Helsinki, Finland

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

First edition 2007

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-08-044715-5
Series ISSN: 1472-7870

Printed and bound in The United Kingdom

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

Current Research in the Semantics/Pragmatics Interface (CRiSPI)

The aim of this series is to focus upon the relationship between semantic and pragmatic theories for a variety of natural language constructions. The boundary between semantics and pragmatics can be drawn in many various ways; the relative benefits of each gave rise to a vivid theoretical dispute in the literature in the last two decades. As a side effect, this variety has given rise to a certain amount of confusion and lack of purpose in the extant publications on the topic. This series provides a forum where the confusion within existing literature can be removed and the issues raised by different positions can be discussed with a renewed sense of purpose. The editors intend the contributions to this series to take further strides towards clarity and cautious consensus.

This page intentionally left blank

Contents

Contributors

Ángel Alonso-Cortés, *Universidad Complutense, Spain*
Cecilia Di Chio, *University of Essex, UK*
Paolo Di Chio, *University of L'Aquila, Italy*
Robin Clark, *University of Pennsylvania, USA*
Pelle Guldborg Hansen, *Roskilde University, Denmark*
Gerhard Jäger, *University of Bielefeld, Germany*
David M. Levy, *George Mason University, USA*
Jun Miyoshi, *Kanto-Gakuin University, Japan*
Prashant Parikh, *University of Pennsylvania, USA*
Sandra J. Peart, *Baldwin-Wallace College, USA*
Ahthi-Veikko Pietarinen, *University of Helsinki, Finland*
Ian Ross, *University of Pennsylvania, USA*
Gabriel Sandu, *Institut d'Histoire et de Philosophie des Sciences et des Techniques, France*
Tatjana Scheffler, *University of Pennsylvania, USA*
John F. Sowa, *VivoMind Intelligence, Inc., USA*

This page intentionally left blank

Acknowledgments

During 2005, financial support for the work on this book has been provided by the Academy of Finland, (Grant 1103130, *Logic and Game Theory*, Principal Investigator Ahti-Veikko Pietarinen), and during 2006, by the University of Helsinki Excellence in Research Funds and by the Jenny and Antti Wihuri Foundation. I gratefully acknowledge their support.

AHTI-VEIKKO PIETARINEN
University of Helsinki

This page intentionally left blank

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 1

AN INVITATION TO LANGUAGE AND GAMES

Ahti-Veikko Pietarinen
University of Helsinki

1 INTRODUCTION

Language—a Game? That language may be compared to games, and hence to strategic and rational interactions, is an ancient idea. As a metaphysical thought, the opposition of *chôra* and *kosmos* in Plato's philosophy is the contest and play between the distracted and the ordered, the changing and the permanent. As a metaphor for argumentation, Aristotle's *Topics* and its later incarnations, such as the scholastic *Ars Obligatoria*, are set up as dialogical duels.

The last century was marked by a linguistic turn in philosophy. Those seeking to understand the expression of natural language sometimes chose to focus on games. Ferdinand de Saussure (1857–1913), a pioneer of structural linguistics, considered chess the man-made counterpart of the natural processes of language in *Course in General Linguistics* (1916), in which he compared language and chess. They both involve dynamics, their rules are conventional, and their strategies positional, Saussure argued. The difference lies in deliberation: while in chess the player intends various moves, in language moves are spontaneous and fortuitous. Saussure's comparison does not hold water: if we interpret the difference as that between what is strategic and what is non-strategic, the difference that Saussure ended up advocating erases practically all he wanted to see as chess-like in language.

Earlier, Charles Peirce (1839–1914) had expressed a much better thought-out allegory in which thought is mediated by expression, just as pawns and knights mediate the purposes and intentions of those playing the game of chess:

> Thinking always proceeds in the form of a dialogue—a dialogue between different phases of the *ego*—so that, being dialogical, it is essentially composed of signs, as its Matter, in the sense in which a game of chess has the chessmen for its matter. (Peirce 1967, MS 298, 1905, *Phaneroscopy*)

Peirce remarked that it was "a sop to Cerberus" to explain the meaning of signs in terms of strategic dialogues that refer to actual persons uttering and interpreting those signs.[1] Indeed, the

[1] See Chapter "The semantics/pragmatics distinction from the game-theoretic point of view" in this volume.

concept of strategy integral to game theories of rational decision was not available to him. In place of strategy, Peirce used the notion of a *habit of acting* in certain ways in certain kinds of circumstances.

The game metaphor has retained its strength in linguistics and philosophy alongside logic, contemporary mathematics and theories of computation (Pietarinen 2003, 2007a,b). But is the notion of a game worthy of serious linguistic theorising? What is its relevance? What has game theory brought to the table of theoretical linguists and philosophers of language?

2 Taking Games Seriously

Game theory, as a theory of strategic interaction, has arisen as a noteworthy tool for linguistic analysis, and has been used to expose the multiplicity of issues to do with linguistic meaning, its origins, and its change.

The Emergence of Game Theory The first mathematical result concerning games was suggested by Zermelo (1913) of certain finite, strictly competitive two-player games of perfect information, such as chess. He showed that a player can only avoid losing for a finite number of moves (if the opponent plays correctly), if and only if the opponent is able to force a win. The modern version of the theorem states that every such game is determined: either player 1 or player 2 has a winning strategy.

The notion of strategy was formalised during the 1920s by Émil Borel, John von Neumann, László Kalmár and Dénes König. The theory of games was established in John von Neumann and Oskar Morgenstern's 1944 Locus Classicus, *The Theory of Games and Economic Behavior*.

By the late 1930s, the relevance of game theory to other fields of science, and to economics in particular, was not yet fully acknowledged. John von Neumann, in a letter to Abraham Flexner (25 May 1934), confessed: "I have the impression that [economics] is not yet ripe... not yet fully enough understood... to be reduced to a small number of fundamental postulates—like geometry or physics" (quoted in Leonard 1995, p. 730). The influence of game theory grew slowly, and happened, to a considerable degree, via Morgenstern's attention, the co-author of *The Theory of Games*, though a full axiomatisation was never reached.

In lieu of axiomatisation, manifold applications of game theory have proved its scientific worth. "By their fruits ye shall know them," pronounced both Charles Peirce and David Hilbert in their independent and contemporaneous discoveries in logic and mathematics around the turn of the century, frequently applying the game metaphor to a variety of tasks. Economics, statistics, logic, mathematics, the social and political sciences, ethics, physics and biology have later all resorted to game theory in clarifying some of their most difficult and fundamental theoretical constructions.

Our focus in this book is on linguistics and the philosophy of language. Games are models of human actions, and language, speech and communication exemplify those actions. But we must distinguish two levels: games as a theoretical framework for studying the nature and the origins of linguistic meaning and games as models of large classes of rational human behaviour in actual communicative situations. Predominantly, this book is concerned with games in their former role, and it was this role that Wittgenstein, one of the first philosophers to systematically argue for the usefulness of games in the philosophy of language, thought underlies linguistic meaning. His insights later resurfaced in theories such as evolutionary and semantic games.

Language Games The first pages of Ludwig Wittgenstein's *Philosophical Investigations* (1953) introduce the idea of a language game in order to show that the words of a text, or a complete primitive language, derive their meaning from the role that they have in certain non-linguistic activities he calls 'games'. For Wittgenstein, the foundational purpose is not something that can be found in language but is external to them. Games are conceptually prior to symbolic codes. They are the activities and practices from which language derives its meaning. "[Y]ou can learn that the word has meaning by the particular use we make of it. We are like people who think that pieces of wood shaped more or less like chessmen or draughtstones standing on a chessboard constitute a game even if nothing has been said as to how they are to be used" (Wittgenstein, 2000–, 147, 39v), he writes in the thirties (ibid., 149: 18), soon followed by the comment: "For what we call the meaning of the word lies in the game we play with it." Likewise for sentences: "In which case do we say that a sentence has [a] point? That comes [close] to asking in which case do we call something a language game. I can only answer. Look at the family of language games that will show you whatever can be shown about the matter" (ibid., 148: 36v).

The purpose of the players in Wittgenstein's language games is to "show or tell what one sees". What the players try to achieve is to bring to the fore what they see to be the case in the context of an assertion: "It is true that the game of 'showing or telling what one sees' is one of the most fundamental language games, which means that what we in ordinary life call using language mostly presupposes this game" (ibid., 149: 1).

To show or to say that something is the case is to communicate those findings. In some cases that might involve the naming of objects, but that would not be the whole story. To name something is not yet effectual. It does not, Wittgenstein remarks, constitute a genuine move in a language game:

> Within naming something we haven't yet made a move in the language game,—any more that [sic] you have made a move in chess by putting a piece on the board. We may say: by giving a thing a name *nothing* [has] yet been done. It *hasn't* a name,—except in the game. This is what Frege meant by saying that a word has meaning only in its connection with [the context of] a sentence. (ibid., 226: 36)

Seeing and telling what is the case and naming something are not on the same dimension. It might suffice to give something a name and to rest content with that, but that reveals nothing about the meaning of expressions. Language games have to be actively played for a meaning of expressions to emerge, which in turn is a prerequisite for conversational meaning and speech acts, the second major arena for games and quite distinct from the first.

Logic of Conversation Game-theoretic approaches to theories of communication have recently enjoyed success. Paul Grice's (1989) writings have been instrumental in providing normative theories of communication that follow rationality and cooperation. The most attention has been paid to Grice's maxims of conversation—especially on the maxim of relation in theories of relevance (Sperber & Wilson 1995, Pietarinen 2004a), which he took to be imminent outcomes of these postulates. Less has been laid on the overall ethical project to elicit different maxims from the assumption that dialogue partners are rational and aim to increase the *summum bonum*—the 'ultimate good' as scholastics had it—by one kind of cooperative practice or another.

Implementations of the Gricean project have suffered from similar difficulties as the overall theory of games. It is not obvious that agents maximising expected utilities in fact aim at increasing the *summum bonum* since that may be a mere by-product of their activities. Nor is it clear that language users invariably act according to principles of rationality. This does not speak against using game theory in theories of communication and conversation—or in the "logic of conversation", as Grice calls it, since it no longer need subscribe to full rationality or complete information about payoffs. Rather, what we are given is a preliminary argument to the effect that there is something intrinsically congenial in theories of communication and theories of games.

Many have endeavoured to spell out this congeniality. For instance, Hintikka (1986) argued that Grice's programme ought to be operationalised by assigning payoffs to full communicative strategies, not to individual moves that interlocutors make. Only when communication terminates do we have enough evidence and reason to assess the value of the path taken by the speaker or the hearer through a multiplicity of possible conversational situations.

A further argument in favour of a strategic and dialogical outlook on communication is that language use and understanding is reciprocal, and the responsibilities are equally and mutually distributed between the speaker and the hearer. Originally, Grice's maxims pertained to singular utterances. Even today, the theory of relevance, which advocates strategic reasoning in terms of the maximisation of linguistic information and the minimisation of the cognitive processing effort required in gleaning relevant information, overlooks the interpreter's deliberations on what he or she interprets as relevant in the utterance.

Even if attempts to join game-theoretic assets with pragmatic elements of language use were to have their basis in theories of conversation, and even if Grice's cooperative principle were the main principle preserved in conversation, it is not exactly right to simply equate Grice's principle with cooperation in the game-theoretic sense. Grice's technical definition of cooperation (according to which the speaker's contribution ought to be such that is required by the accepted purpose of the exchange) is speaker oriented and says little about the actual and quite complex process of interpretation. In game theory, on the other hand (according to which players' roles are, in normal cases, symmetric in the sense that no one player cooperates less than the others), Grice's definition does not lead to interactive cooperation.

In fact, cooperation in the game-theoretic sense differs in one fundamental sense from cooperation in the sense in which Grice defined it. Cooperative approaches in economics do not attempt to model the precise manner in which agents communicate with each other because agents are assumed to endorse joint action, and any communication is relegated to pre-play situations. Such games are typically coalitional. From Grice's perspective, agents increase the *summum bonum* of some subsets of linguistic communities, not humankind at large.[2]

Formal Pragmatics Jürgen Habermas has also sought to drive a wedge between what is communicative and what is strategic in linguistic action, but with very different music. He considers the two modifiers to be fundamentally different in his own account of communicative practices (Habermas, 1995). His motivation in drawing this division is not to show that the attitudes of the participants may differ in the two settings—the communicative and the strategic—but because he sees some transparent structural differences in them. In communication, the structure of how language is used is "superimposed" on goal-driven action (Habermas, 1998, p. 205). Communication is replete, Habermas holds, with notions such as presuppositions, performatives, and other less objective constraints than strategic action. Its essence lies in the idea of interaction, but in

[2] Curiously, Grice was a keen admirer of Peirce's logic and philosophy, and that influence shows up in a number of junctions in his account of the "logic of conversation" (Pietarinen, 2004).

Habermas's view, not in interaction that involves strategic considerations. On the contrary, he maintains, strategic action is parasitic on communicative interaction.

But is it not misguided to try to draw these distinctions in terms of what is strategic and what is non-strategic? We know that strategic action may be performed cooperatively or non-cooperatively. Notwithstanding the performative contradictions that Habermas thinks ensue from manipulating the listener to give an answer that the speaker desires, communication may well be strategic in the sheer task of understanding and interpreting utterances. We may well play variable-sum games in which the outcomes assigned to total strategies mark varying degrees of understanding. What Habermas seeks to explain is that, if the crux of the strategic interaction falls within the principle of utility maximisation and hence self-interest, then it is incompatible with reciprocal understanding. However, the principle of utility maximisation, operationalised in game theory by solution concepts, well satisfies Habermas's desiderata of reaching understanding and agreement, having coordination and having cooperation. One just needs to shift the focus to cooperation, negotiation or bargaining, the essence of which is in variable-sum payoffs. This does not diminish the scope of communication; on the contrary, one inherits more precise tools and methods for tackling the structure of communication and discourse.

The goal of Habermas seems rather to be a reconstruction of the meaning of linguistic competence and an awareness of its rules. He assumes that language has an in-built notion of validity which assertions make use of. The force that acts of meaning something, such as illocutions, have is based on the assumption that such assertions can be checked for validity. Obliging acts convince the hearers:

> With their illocutionary acts, speaker and hearer raise validity claims and demand that they be recognized. But this recognition need not follow irrationally, since the validity claims have a cognitive character and can be tested. I would like, therefore, to defend the following thesis: *In the final analysis, the speaker can illocutionarily influence the hearer, and vice versa, because speech act-typical obligations are connected with cognitively testable validity claims*—that is, because the reciprocal binding and bonding relationship has a rational basis. The speaker who commits herself normally connects the specific sense in which she would like to take up an interpersonal relationship with a thematically stressed validity claim and thereby chooses a specific mode of communication. (Habermas, 2001, p. 85)

The term "formal" is to be taken in the sense of "rational reconstruction". But this does not pardon Habermas for confounding the cooperative and the strategic. Had his investigation moved on a more detailed level of rational reconstruction in the sense of game theory, the mix-up would have been exposed earlier.

To put the point in simple terms, what Habermas is after is communicative action that aims at reaching understanding, whereas strategic action exerts influence on others. But the former is not devoid of purpose. In explaining what people do, we need goal-driven action structures for both. This need has been recognised ever since the early phases of the formation of the theory of games:

> Even if the theory of noncooperative games were in a completely satisfactory state, there appear to be difficulties in connection with the reduction of cooperative games to noncooperative games. It is extremely difficult in practice to introduce into the cooperative games the moves corresponding to negotiations in a way which will reflect all the infinite variety permissible in the cooperative game, and to do this without giving one player an artificial advantage (because of his having the first chance to make an offer, let us say). (McKinsey, 1954, p. 359)

The infinite variety permissible in cooperative games is precisely the problem we encounter in contemporary studies in formal pragmatics, which deals not only with conventions but also with interpretations of context and environment. It is difficult to reduce all these to some formal framework of contingent but observable behaviour. Even more candidly, Shubik (1985, p. 293) reiterated this point thirty years later, suggesting that:

> [I]n much of actual bargaining and negotiation, communication about contingent behaviour is in words or gestures, sometimes with and sometimes without contracts and binding agreements. A major difficulty in applying game theory to the study of bargaining or negotiation is that the theory is not designed to deal with words and gestures—especially when they are deliberately ambiguous—as moves. Verbal sallies pose two unresolved problems in game-theoretic modelling: (1) how to code words, (2) how to describe the degree of commitment.

In the attempts to overcome these challenges, game theorists, linguists, philosophers and logicians have unearthed a rich frontier for strategic interaction in the evolution of meaning and language use, which has only recently started to be probed in full generality. Game theory is to help these people to understand what is at issue in a variety of linguistic contexts.

3 Consequences for Language Theory

What the allegory between games and language teaches is a serious and deep-seated philosophical problem concerning the relationship between thought, language and reality. All told, we have a concept that marries philosophical thinking with scientific methodology, such as the theory of games and rational decisions, linguistic pragmatics, logic, evolutionary biology and countless others. Aside from the pre-eminence of the intellectual history of the idea, the focus of most of the chapters in this book is on one of the two major methodologies: evolutionary game theory or game-theoretic semantics.[3]

Evolutionary Game Theory Most of the current theories on the market on language evolution are structural and functional rather than *strategic* in nature, and are built upon the presupposition that it is possible to model our 'innate linguistic endowment' and then correlate these models with some neo-Darwinian evolutionary theory.

The problem with neo-Darwinism is that what it perceives to be responsible for the preservation of favourable variations as well as for destroying unfavourable ones is the capacity for *adaptation* of these variations. But adaptation refers to structural and functional processes which are not the only, and maybe not even the most plausible, paradigms for theories of linguistic meaning or its change. Instead, complex meaning relations between assertions and the world may emerge through strategic interactions.

Evolutionary game theory studies the dynamics of strategic interactions of populations playing repeated games on a given resource (Maynard Smith, 1982). At the core of this theory is continuous change in strategies that the players use from one period of the play of the game to another, reflecting the variability of fitness derived from the level of success of previous rounds. Unlike in traditional game theory, no particular theory of rationality concerning language or players is presupposed, let alone common knowledge of rationality prone to game-theoretic paradoxes such as using backwards induction as our real-life solution concept.

[3] Pietarinen (2005) is a study that brings evolutionary game theory to bear on the theory of semantic games.

Of late, evolutionary game theory has been applied to the analysis of language from a variety of perspectives. A considerable portion of studies focuses on empirical issues. Studies exist on computer simulations (Nowak et al. 1999, Oliphant & Batali 1996), the iterated learning of grammars (Kirby, 2000), and the naming games (Steels & McIntyre, 1999), to mention just a few. These approaches take the communicability function to determine the payoffs of expressions (Skyrms, 1996). In addition to the paradigms appealing to communicability or performatory factors, crucial parameters in the evolution of linguistic meaning are found in the realms of semantics and pragmatics.

Several chapters in this book take up the issue of applying evolutionary game theory to the contemporary questions of semantics, pragmatics and their interconnections.

Game-Theoretic Semantics While the evolution of the semantic component of language has received less attention than the evolution of phonology, syntax, grammar acquisition or learning, game-theoretic methods have recently played a noteworthy role in formal semantics and pragmatics. This is another major thread in the present book.

The traditional game-theoretic approach to semantics has two players, Myself and Nature, who assume the roles of the verifier and the falsifier of the expressions presented to them (Hintikka, 1973). The two players play a non-cooperative game on a shared commodity, which is a non-empty (possibly infinite) domain of the model and from which they choose individuals. The model is an interpreted structure in which the formulas of the underlying logic or natural language are true or false.

Any sentence of English defines a game between Myself and Nature. The game rules for the quantificational expressions such as *some*, *a(n)*, *every* and *any* prompt a player to choose an individual from the relevant domain (a choice set) I, labelling the individual with a name if it does not already have one. The game continues with respect to an output sentence defined by the game rules.

Moves may also refer to a wide variety of lexical and morphological categories such as modals, intensional verbs, tense operators, pronouns, definite descriptions, possessives, genitives, prepositional phrases, eventualities, adverbs of quantification, aspectual particles and polarity items (Hintikka & Kulas 1983, 1985; Hintikka & Sandu 1991, Pietarinen 2001b).

A play of the game terminates when such components are reached in which further applications of game rules are no longer permitted. Their truth in a given interpretation determines whether Myself or Nature wins the play.

An example of the game rule for *some* is as follows.

(G.some): If the game has reached a sentence of the form

$\mathsf{X} - \textit{some}\ \mathsf{Y}\ \textit{who}\ \mathsf{Z} - \mathsf{W},$

then Myself chooses an individual from I, say b. The game continues with the sentence

$\mathsf{X} - \mathsf{b} - \mathsf{W}, \mathsf{b}\ \textit{is a}\ \mathsf{Y}, \textit{and}\ \mathsf{b}\ \mathsf{Z}.$

Here *who* Z (or *where* Z, *when* Z etc.) is the entire relative clause, and the main verb phrase W and the head noun in Y are in the singular. For simplicity, the relative clause markers are often omitted. X marks free linguistic context.

For *any* the game rule is dual to (G.some):

(G.any): If the game has reached the sentence

X – *any* Y *who* Z – W,

then Nature may choose an individual, say b. The game is continued with the sentence

X – b – W, b *is a* Y, *and* b.

Another example is for negation:

(G.not): If the game has reached a sentence of the form *neg*(A), the players exchange roles, and the winning conventions will also change. The game continues with respect to A.

The operation *neg*(A) is a functor forming sentential negations of A.

Legitimate moves are thus determined by the constituent expressions under evaluation.

The semantic games that are being played here may be thought of either as in their *normal* or in *extensive form*. In normal-form games, the strategy profiles are arrays of Skolem functions. In extensive-form games, the root is the complete expression, and the terminal histories are labelled with atomic formulas or simple linguistic items. The strategy profiles are constituted by the move-by-move responses to other players' moves in any of the information sets in which a player makes his move. Extensive-form games show the evolved interaction with the details of individual choices, responses, positions and information the players have. Specifying the strategy profiles completely determines the outcome of the game.

The payoffs are assigned to strategies in normal-form games and to terminal histories in extensive-form games. A strategy that invariably leads the player whose role at the beginning of the game is the verifying (resp. falsifying) one to the winning terminal position interpreted as True (resp. False) will be his or her winning strategy. The existence of such strategies shows when a compound formula, a sentence of natural language or a segment of discourse is true and when it is false in a given model.

Game-theoretic semantics carries with it a couple of further assumptions. Just as in the traditional theory of games (von Neumann & Morgenstern, 1944), the players are hyper-rational optimisers. Furthermore, semantic games are static one-off games. They have a finite horizon as the input strings are finite in length. Strategies are pure, and in terms of the conditions for the truth of the expressions, Nature's function remains stationary. In addition, the game and its equilibrium are common knowledge.

A Test Case: Anaphora Interpreting anaphora is a good test case for the efficiency of game-theoretic semantics. It concerns the relative accessibility of the information about either the choices of individuals or the use of strategies. The accessibility of information can pertain to earlier parts of the same semantic game or to the plays of games in the periodical past.

Extensive-form games contain information that may be lost in representing the games in their normal form. An example is resolving anaphoric relations. Consider, for instance, the discourse *A knight sees a dragon. He escapes.* What the suitable or intended value of the anaphoric pronoun *he* is, is often found in the past discourse referring to some earlier derivational history of that discourse (*a knight*), and there is no way of recovering that information merely from the normal form of the game. In interpreting anaphora, extensive games provide the needed diachronic models of interaction.

However, in interpreting anaphora more is needed than just the record of derivational histories and the ensuing access to earlier choices made in the game. Players also need access to information concerning strategies that they themselves or their adversaries have entertained in earlier parts of the game. For that to be possible, extensive forms can be extended to *hyperextensive-form* games (for details, see Sandu & Janasik 2003, Pietarinen & Sandu 2004). These games code the information concerning the strategies that players have used earlier in the game into the local states of the players at each non-terminal history in which a player is to move. Hyperextensive games account for functional anaphora in sentences such as *Every man carried a gun. Most of them used it.* Here, the values for *the gun* and *it* are given by a function from men to guns, producing for each man, a particular gun.

Furthermore, as soon as pragmatic change is at issue, an even greater modicum of diachronism than just access to earlier actions or strategies made in the extensive or hyperextensive game is needed. Players need to have access to actions and strategies emanating from earlier plays of the game, too. This transfer of information from earlier plays to future ones is at issue in pragmatic change: as each game is played within a changing environment, there is bound to be variation in the strategies from one play of the game to another. "A language game does change with time," noted Wittgenstein in *On Certainty* (256).

Linguistic meaning thus has two levels of diachronism: the more semantically oriented question of game-internal references to past actions and the 'trans-structural' question of the amount of information concerning actions transmitted from earlier periods to future ones. Pragmatic features per se are not subordinate to the latter constraint and may involve shorter intervals taking place within single runs of the games. However, changes in such features are parasitic on aspects of evolutionary dynamics in sequences of periods comprising total extensive games, because that is the only way games are able to be played in changing and mutating environments and contexts that nurture the meaning.

Interpreting anaphora thus becomes a matter of the relative accessibility of the information concerning either the choices of individuals or the use of strategies. In other words, the accessibility of information pertains either to earlier parts of the same semantic game (intra-structural pragmatics), or to the histories of earlier plays of games (trans-structural pragmatics).

4 OVERVIEW

The purpose of this book is to introduce the reader to the variety of linguistic contexts in which the notion of a game proves its worth. As attested in the following chapters, games need not be left as an apt allegory for understanding communicative meaning or issues in linguistic interaction, but can be reworked into a theory which provides an efficient tool for the analysis of linguistic meaning in all senses of that inquiring term.

What are the fundamental issues raised by Wittgenstein's language games to theories of semantics and logic? JOHN F. SOWA's chapter "Language Games, A Foundation for Semantics and Ontology" takes as a starting point the idea that the multiplicity of language games accounts for the flexibility of natural languages, while staying within one game makes expressions more precise. Peirce noted that "one and the same proposition may be affirmed, denied, judged, doubted, inwardly inquired into, put as a question, wished, asked for, effectively commanded, taught, or merely expressed, and does not thereby become a different proposition" (Peirce, 1976, p. 248). Wittgenstein's examples of different language games included giving orders and obeying them, describing the appearance of an object, giving its measurements, drawing, reporting an event, speculating about an event, forming and testing a hypothesis, presenting the results of an experiment in tables and diagrams, making up a story and reading it, play-acting, singing catches, guessing riddles, making a joke, telling it, solving a problem in practical arithmetic, translating from one language into another, asking, thanking, cursing, greeting and praying (Wittgenstein, 1953, §23). To accommodate these shifting interpretations of expressions common in communication, Sowa rejects an assumption of a single logic and calls for a major paradigm shift in formal semantics, which adopts a dynamic framework of logics and ontologies, characterising the variety of language use as a variability of the application of different language games.

In "Counterfeiting Truth: Statistical Reporting on the Basis of Trust" DAVID M. LEVY and SANDRA J. PEART trace the ideas of the strategic constitution of linguistic meaning and rational interaction back to Adam Smith's (1723–1790) account of trade phenomena in monetary institutions. Smith observed that language and economics both attempt to explain similar regularities in human interaction and in the design of social systems. Rules and protocols governing language arise by mutual consent, and language as a rule-based system develops and is revised in accordance with usage over periods of time. Smith made a major theoretical point in emphasising the priority of use over rule governing. He referred to the affinity between trade and language in *The Wealth of Nations* (1776) as two sides of the same underlying process: "[The division of labour] is the necessary, though very slow and gradual, consequence of a certain propensity in human nature which has in view no such extensive utility; the propensity to truck, barter, and exchange one thing for another. ... [It is] the necessary consequence of the faculties of reason and speech" (Smith, 1991/1776, p. 19). In particular, Smith's insight was that institutions rely on trust and trust is carried by language. Statistical deceit is an instance of that larger insight. Levy and Peart take statistical estimators to be conventions, which were famously introduced into the study of language by David Lewis (1969) in his game-theoretic examination of conventions. Statistical estimators, Levy and Peart argue, are at work in contributing to the semantic framework of language, and they demonstrate that, in contrast with Lewis's game-theoretic account of conventions, a convention of statistical estimators is conducive to conflict rather than coordination.

In his *Dissertation on the Origin of Languages* (1767), in which Adam Smith concentrates largely on lexical semantics, the focus is on the emergence of simple words such as proper names, adjectives, prepositions, comparatives, demonstrative pronouns and verbs (Smith, 1970/1767). He distinguishes between words that emerge from the desire to express qualities of objects and those that emerge from the desire to express the relationships in which objects stand to each other.

Accordingly, Smith uncovered further insights into the workings of language, including the time-honoured questions concerning its origins. ÁNGEL ALONSO-CORTÉS in "From Signals to Symbols: Grounding Language Origins in Communication Games" brings together the fields of economics and linguistics on the topic of the origins of language. The key concern is how linguistic signs or symbols have inherited design features present in communication. Alonso-Cortés demonstrates how some features of language can be adequately understood as a result of coordination games. Developing upon Smith's insights, Alonso-Cortés argues that modern language originated as a consequence of trade or exchange relationships and the division of labour among early humans. As an economic activity, both trade relationships and the division of labour call for coordination, and the outcome is a causal relationship between game-theoretic activities and general properties of the linguistic symbol.

After Lewis published his 1969 work on conventions, which laid the foundation for a game-theoretic approach to social conventions, a vast number of studies concerning the role and the nature of evolutionary arguments in linguistics have sprung up. This evolutionary turn, PELLE

GULDBORG HANSEN observes in "Evolutionary Games and Social Conventions", has marked a transition from the classical assumptions of perfect rationality and common knowledge to assumptions about agents as members of evolutionally constrained populations. He provides an extensive review of the relations between social conventions on the one hand and phenomena such as Pareto efficiency, risk, discrimination, self-interest and cooperation on the other, and goes on to make a general argument in support of the evolutionary turn in the theory of convention by a progressive exposition of its successful application to a variety of paradigmatic games.

The state of the art in evolutionary models and computer simulations concerning the evolutionary emergence of language is the topic of CECILIA DI CHIO and PAOLO DI CHIO in their chapter on evolutionary language games, a discipline which can be seen to have emerged from the union of evolutionary game theory and the philosophical ideas revolving around language games. They review some of the key works on evolutionary language games and propose simulation models for the evolution of language in view of verifying some previous results and demonstrating how the presence of a topological structure influences the communication among individuals.

What is the relationship between semantics and pragmatics? GERHARD JÄGER argues in "Game Dynamics Connects Semantics and Pragmatics" that the best response dynamics lends itself to an epistemic interpretation and that this provides a suitable game-theoretic foundation for pragmatic reasoning in the Gricean tradition. Continuing the themes of the previous chapters, Jäger provides a wide-ranging overview of evolutionary interpretations of game theory, compares two versions of it, replicator dynamics and best response dynamics, and explores the ensuing notions of evolutionary stability.

Moving towards actual models of communication and conversation, JUN MIYOSHI builds some game-theoretic models of conversations and explores links between game theory and linguistic conversation. Miyoshi analyses a simple example of conversation and discusses the question of how it could be formulated into a game format. Miyoshi then presents a family of games with perfect and complete information as a general model of conversations, and applies some theorems of game theory to it, followed by an argument for games of incomplete information as a more realistic model of conversations. Miyoshi also suggests how game-programming techniques are applicable to the modelling of conversations.

It is commonplace in pragmatics to invent new methods based on either empirical data or theoretical assumptions to tackle problems that are thought to be inadequately handled by the apparatus of formal semantics. "Situations and Solution Concepts in Game-Theoretic Approaches to Pragmatics" by IAN ROSS explores one such recent method in theoretical linguistics, bidirectional optimality theory, and argues that, if restricted to operating over lexical items or simple clauses, it is unable to account for certain scalar implicatures determined by contextual factors. According to Ross, in order to accommodate such cases in this framework, the relevant units of optimisation must be multiclause sentences. If this step is taken, the predictions of bidirectional optimality theory and games of partial information—the latter proposed by Prashant Parikh (2001) to formally address Gricean issues around the "logic" of conversation—converge, although they remain distinct in the general case. Such robust context-dependent examples of scalar implicature show, Ross argues, that adequate models cannot reduce such phenomena to a lexical account.

Pushing the ideas of games of partial information further and taking situation theory to contribute to the fundamental philosophical assumptions concerning semantics and pragmatics, PRASHANT PARIKH and ROBIN CLARK present in their chapter "An Introduction to Equilibrium Semantics for Natural Language" a case for equilibrium semantics, a new framework for the study of meaning that combines semantics and pragmatics into a single discipline. At the heart of equilibrium semantics is the idea that the referential and the communicative are no longer separated, and they accomplish this by the semantics that build use into reference at the level of situation-theoretic "grounding situations" so that the need for a separate discipline of pragmatics vanishes.

The chapter by Parikh and Clark predicts a merger not only between semantics and pragmatics but also between the theoretical mechanisms of game-theoretic semantics in the sense of games of partial information and game-theoretic semantics in the sense of Jaakko Hintikka's *œuvre*. By taking a look at some central aspects of the game-theoretic semantics of the latter kind, TATJANA SCHEFFLER criticises the ordering principles in operation in the theory which guide the application of game rules. She argues that sentences with quantifier scope ambiguities demonstrate that ordering principles cannot impose a fixed hierarchy on game rules. Instead, Scheffler proposes that the principles allow game rules to be played in different orders, which yields two or more different games for some input sentences, distinct games corresponding to distinct semantic interpretations. Based on data involving complex quantifier scope ambiguities, including inverse linking examples, she suggests a new ordering principle for quantifiers, argues that a hierarchy is needed that determines the relative precedence of ordering principles, and tests the approach with respect to coordination and quantifier scope.

The scope issue is also pertinent to the very functionality of game-theoretic semantics. In fact, it has turned out that the notion of scope is very ambiguous in linguistic semantics. Can we have a feasible semantic interpretation that draws a logical distinction between central notions of scope commonplace in language and which also satisfies some vital quantificational constraints? To answer the challenge, GABRIEL SANDU in his chapter "Two Notions of Scope" offers a dynamic version of game-theoretic semantics to account for the all-important distinction between logical and binding notions of scope.

What else is game-theoretic semantics good for? One thing that can be done is to extend its applicability. In "Semantic Games and Generalised Quantifiers" AHTI-VEIKKO PIETARINEN proposes to marry generalised quantifiers with game-theoretic semantics. To accomplish this, several semantic game rules for various types of generalised quantifiers are formulated in that chapter. Moreover, Pietarinen argues that game-theoretic semantics surpasses relational semantics in that it provides a generic method of dealing with context-dependent quantifiers in terms of strategic content and also in that it offers a general semantics for branching generalised quantifiers.

There are further uses for game-theoretic semantics. In "Games, Quantifiers and Pronouns" ROBIN CLARK argues that reference tracking, the ability to successfully assign referents to discourse anaphors, is an example of how linguistic agents can strategically manage a resource and is therefore amenable to a game-theoretic analysis. Clark continues to fuse Hintikka-style game-theoretic semantics for natural language with games of partial information. The basic idea is that once discourse entities have been introduced, they can be treated as a resource available as public knowledge to the participants of the discourse. The participants can then treat the problem of associating referents with discourse anaphors as a game that can be solved rationally. Clark demonstrates how the referents for discourse anaphors can be found by solving for the Pareto-Nash equilibrium of the game. The idea is that both the speaker and the hearer are involved in a strategic interaction and that the basic structure of the problem is a matter of public knowledge. Because of this common mutual knowledge, the participants in the conversation are able to formulate coherent strategies dealing with reference tracking, the ability to correctly assign discourse referents to pronouns. Clark develops a technique which relies on the management of a data structure, a game board, and shows how speakers can strategically use this resource during the course of a conversation. He also demonstrates how new quantifier rules of game-theoretic semantics allow for quantified expressions to establish discourse entities, and addresses the problem of scope.

What does all this imply for the overall theories of semantics and pragmatics of language and their relationship? From the game-theoretic point of view, AHTI-VEIKKO PIETARINEN argues in the concluding chapter, what is semantic and what is pragmatic in language cannot be distinguished by the rule-governed and structural features of game theory. From the perspective of game theory, the sole difference is whether players entertain epistemic relationships with respect to the solution concepts and strategy profiles in the game-theoretic analysis of linguistic meaning. This implies that the distinction is in a very real and concrete sense illusory. Epistemologically, however, the distinction makes a world of difference. The field of interactive epistemology, for instance, marrying game theory, epistemic logic and the theory of rational action, is emerging as a significant contribution to strategic meaning in linguistics. Such an integration is likely to significantly improve our theoretical understanding of a substantial number of phenomena connected with linguistic meaning which frequently arise in the individual chapters.

ACKNOWLEDGMENTS

Supported by the Academy of Finland (Grant 1103130, *Logic and Game Theory*, Principal Investigator Ahti-Veikko Pietarinen), the University of Helsinki Excellence in Research Funds and the Jenny and Antti Wihuri Foundation.

REFERENCES

Borel, E. (1921). La thèorie du jeu et les equations intègrales à noyau symètrique. *Comptes Rendus Hebdomadaires des Sèances de l'Acadèmie des Sciences*, **173**, 1304-1308. (Translated L. J. Savage: 1953. The theory of play and integral equations with skew symmetric kernels. *Econometrica*, **21**, 97-100.)

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

Habermas, J. (1995). Peirce and communication. In: *Peirce and Contemporary Thought* (K. Ketner, ed.), pp. 243-266. Fordham University Press, New York.

Habermas, J. (1998). *On the Pragmatics of Communication* (M. Cooke, ed.). MIT Press, Cambridge, Mass.

Habermas, J. (2001). *On the Pragmatics of Social Interaction: Preliminary Studies in the Theory of Communicative Action* (B. Fultner, trans.). MIT Press, Cambridge, Mass.

Hintikka, J. (1973). *Logic, Language-Games and Information*. Oxford University Press, Oxford.

Hintikka, J. (1986). Logic of conversation as a logic of dialogue. In: *Philosophical Grounds of Rationality* (R. E. Grandy and R. Warner, eds.), pp. 259-276. Clarendon Press, Oxford.

Hintikka, J. and J. Kulas (1983). *The Game of Language: Studies in Game-Theoretical Semantics and its Applications*. Reidel, Dordrecht.

Hintikka, J. and J. Kulas (1985). *Anaphora and Definite Descriptions*. Reidel, Dordrecht.

Hintikka, J. and G. Sandu (1991). *On the Methodology of Linguistics*. Blackwell, Oxford.

Janasik, T., A.-V. Pietarinen and G. Sandu (2003). Anaphora and extensive games. In: *Chicago Linguistic Society 38: The Main Session* (M. Andronis, E. Debenport, A. Pycha and K. Yoshimura, eds.), pp. 285-295. Chicago Linguistic Society, Chicago.

Japaridze, G. (2006). In the beginning was game semantics.... In: *Logic and Games: Foundational Perspectives* (O. Majer, A.-V. Pietarinen and T. Tulenheimo, eds.).

Kirby, S. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In: *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (C. Knight, M. Studdert-Kennedy and J. Hurford, eds.), pp. 303-323. Cambridge University Press, Cambridge.

Leonard, R. J. (1995). From parlor games to social science: von Neumann, Morgenstern, and the creation of game theory 1928–1944. *Journal of Economic Literature*, **23**, 730-761.

Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press, Cambridge, Mass.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.

McKinsey, J. C. C. (1954). *Introduction to the Theory of Games*. McGraw-Hill, New York.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. John Wiley, New York.

Nowak, M. A., J. B. Plotkin and D. C. Krakauer (1999). The evolutionary language game. *Journal of Theoretical Biology*, **200**, 147-162.

Oliphant, M. and J. Batali (1996). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter* 11(1).

Parikh, P. (2001). *The Use of Language*. CSLI Publications, Stanford.

Peirce, C. S. (1967). Manuscripts in the Houghton Library of Harvard University, as identified by Richard Robin. *Annotated Catalogue of the Papers of Charles S. Peirce* (Amherst: University of Massachusetts Press, 1967), and in The Peirce Papers: A supplementary catalogue. *Transactions of the C. S. Peirce Society*, **7**, 1971, 37-57.

Peirce, C. S. (1976). *The New Elements of Mathematics* (C. Eisele, ed., Vol. 4). Mouton, Berlin.

Pietarinen, A.-V. (2003). Games as formal tools versus games as explanations in logic and science. *Foundations of Science*, **8**, 317-364.

Pietarinen, A.-V. (2004). Grice in the wake of Peirce. *Pragmatics & Cognition*, **12**, 295-315.

Pietarinen, A.-V. (2005). Evolutionary game-theoretic semantics and its foundational status. In: *Evolutionary Epistemology, Language and Culture: A Nonadaptationist Systems-Theoretical Approach* (N. Gontier, J. P. Van Bendegem and D. Aerts, eds.), pp. 429-452. Springer, Dordrecht.

Pietarinen, A.-V. (2006). *Signs of Logic: Peircean Themes on the Philosophy of Language, Games, and Communication* (Synthese Library **329**). Springer, Dordrecht.

Pietarinen, A.-V. (2007a). Who plays games in philosophy? In: *Chess and Philosophy* (B. Hale, ed.). Open Court.

Pietarinen, A.-V. (2007b). Towards intellectual history of logic and games. In: *Logic and Games: Foundational Perspectives* (Majer, O., A.-V. Pietarinen and T. Tulenheimo, eds.).

Pietarinen, A.-V. and G. Sandu (2004). IF logic, game-theoretical semantics, and philosophy of science. In: *Logic, Epistemology and the Unity of Science* (S. Rahman, J. Symons, D. Gabbay and J. P. Van Bendegem, eds.), pp. 105-138. Kluwer, Dordrecht.

Saussure, F. de (1916/1983). *Course in General Linguistics* (R. Harris, trans.), Duckworth, London.

Shubik, M. (1985). *Game Theory in the Social Sciences*. MIT Press, Cambridge, Mass.

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

Smith, A. (1970/1767). *A Dissertation on the Origin of Languages, or Considerations Concerning the First Formation of Languages and the Different Genius of Original and Compounded Languages*. Gunter Narr, Tübingen.

Smith, A. (1991/1776). *The Wealth of Nations*. Prometheus Books, New York.

Sperber, D. and D. Wilson (1995). *Relevance Theory: Communication and Cognition* (2nd edition). Blackwell, Oxford.

Steels, L. and A. McIntyre (1999). Spatially distributed naming games. *Advances in Complex Systems*, **1**, 301-323.

Wittgenstein, L. (1953). *Philosophical Investigations* (3rd edition 1967). Blackwell, Oxford.

Wittgenstein, L. (2000). *Wittgenstein's Nachlass, The Bergen Electronic Edition* (The Wittgenstein Trustees). Oxford University Press, Oxford.

Zermelo, E. (1913). Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In: *Proceedings of the Fifth International Congress of Mathematicians 2* (E. W. Hobson and A. E. H. Love, eds.), pp. 501-504. Cambridge University Press, Cambridge. (Translation: On an application of set theory to the theory of the game of chess, Schwalbe, U. and P. Walker (2001). Zermelo and the early history of game theory, *Games and Economic Behaviour*, **34**, 123-137.)

This page intentionally left blank

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 2

LANGUAGE GAMES, A FOUNDATION FOR SEMANTICS AND ONTOLOGY

John F. Sowa
VivoMind Intelligence, Inc.

The issues raised by Wittgenstein's language games are fundamental to any theory of semantics, formal or informal. Montague's view of natural language as a version of formal logic is at best an approximation to a single language game or a family of closely related games. But it is not unusual for a short phrase or sentence to introduce, comment on, or combine aspects of multiple language games. The option of dynamically switching from one game to another enables natural languages to adapt to any possible subject from any perspective for any humanly conceivable purpose. But the option of staying within one precisely defined game enables natural languages to attain the kind of precision that is achieved in a mathematical formalism. To support the flexibility of natural languages and the precision of formal languages within a common framework, this article drops the assumption of a fixed logic. Instead, it proposes a dynamic framework of logics and ontologies that can accommodate the shifting points of view and methods of argumentation and negotiation that are common during discourse. Such a system is necessary to characterize the open-ended variety of language use in different applications at different stages of life—everything from an infant learning a first language to the most sophisticated adult language in science and engineering.

1 The Infinite Flexibility of Natural Languages

Natural languages are easy to learn by infants, they can express any thought that any adult might ever conceive, and they are adapted to the limitations of human breathing rates and short-term memory. The first property implies a finite vocabulary, the second implies infinite extensibility, and the third implies a small upper bound on the length of phrases. Together, they imply that most words in a natural language will have an open-ended number of senses—ambiguity is inevitable. Charles Sanders Peirce and Ludwig Wittgenstein are two philosophers who understood that vagueness and ambiguity are not defects in language, but essential properties that enable it to express anything and everything that people need to say. This article takes these insights as inspiration for a system of metalevel reasoning, which relates the variable meanings of a finite set of words to a potentially infinite set of concept and relation types, which are used and reused in dynamically evolving lattices of theories, which may be expressed in an open-ended variety of logics.

At the beginning of his career, Wittgenstein, like many of the early researchers in artificial intelligence, thought he had found the key to solving the problems of understanding language and reasoning. In his first book, the *Tractatus Logico-Philosophicus*, he presented an elegant view of semantics that directly or indirectly inspired the theories of formal semantics and knowledge representation that were developed in the 20th century: an elementary proposition expresses an atomic fact about a state of affairs (*Sachverhalt*), which consists of a configuration of objects (*Verbindung von Gegenständen*); a compound proposition is a Boolean combination of elementary propositions; everything in the world can be described by some proposition, elementary or compound; and everything that can be said can be clearly expressed by some proposition about such configurations. His conclusion was the famous one-sentence Chapter 7, which conveniently dismissed all exceptions: "Whereof one cannot speak, thereof one must be silent."

The *Tractatus* inspired Rudolf Carnap's version of logical positivism and Alfred Tarski's model-theoretic semantics. One of Tarski's students, Richard Montague, extended model theory to intensional verbs, such as *believe*, *want*, or *seek*. Montague's grammar (1970) mapped a *fragment* of English to models with an elaborate construction of multiple worlds instead of Wittgenstein's single world. Around the same time, Woods (1968, 1972) and Winograd (1972) implemented model-theoretic systems for talking about moon rocks and the blocks world. Winograd's thesis adviser, Marvin Minsky, was also a technical adviser for the movie *2001*, which featured the HAL 9000, a computer that not only spoke and understood English, but could also read lips, interpret human intentions, and conceive plans to thwart them. When the movie appeared in 1968, Minsky claimed it was a conservative prediction about AI technology in 2001.

Although Wittgenstein and Winograd had a strong influence on later developments, both of them became disillusioned about a decade after their early successes. After Wittgenstein published his first book, which he believed had solved all the solvable problems of philosophy, he went to teach school in an Austrian mountain village. Unfortunately, his pupils did not think or speak the way his theory predicted. It was impossible to find any truly atomic facts that could not be further analyzed or viewed from an open-ended number of different perspectives. Winograd also became discouraged by the difficulty of generalizing and extending his early system, and he later published a harsh critique of his own and other methods for translating natural language to logic (Winograd & Flores 1986). Today, no AI system has any ability that can remotely compare to the HAL 9000, and textbooks based on Montague's approach are illustrated with toy examples that more closely resemble Montague's fragment than the English that anybody actually reads, writes, or speaks.

The precision of logic is valuable, but what logic expresses so precisely may have no relationship to what was intended or required. A formal specification that satisfies the person who wrote it might not satisfy the users' requirements. Engineers summarize the problem in a pithy slogan: "Customers never know what they want until they see what they get." More generally, the precision and clarity that are so admirable in the final specification of a successful design are the result of a lengthy process of trial, error, and revision. In most cases, the process of revision never ends until the system is obsolete.

Unlike formal languages, which can only express the finished result of a lengthy analysis, natural languages can express every step from an initially vague idea to the final specification. During his career as an experimental physicist and a practicing engineer, Peirce learned the difficulty of stating any general principle with absolute precision:

> It is easy to speak with precision upon a general theme. Only, one must commonly surrender all ambition to be certain. It is equally easy to be certain. One has only to be sufficiently vague. It is not so difficult to be pretty precise and fairly certain at once about a very narrow subject. (CP 4.237)

This quotation summarizes the futility of any attempt to develop a precisely defined ontology of everything, but it offers two useful alternatives: an informal classification, such as a thesaurus or terminology, and an open-ended collection of formal theories about narrowly delimited subjects. It also raises the questions of how and whether these resources might be used as a bridge between informal natural language and formally defined logics and programming languages.

Even if an ideal semantic representation were found, it would not answer the question of how any system, human or machine, could learn and use the representation. Children rapidly learn to associate words with the things and actions they see and do without analyzing them into atomic facts or evaluating Montague's functions from possible worlds to truth values. The following sentence was spoken by Laura Limber at age 34 months and recorded by her father, the psychologist John Limber (1973):

> *When I was a little girl, I could go "geek, geek" like that;*
> *but now I can go "This is a chair."*

In this short passage, Laura combined subordinate and coordinate clauses, past tense contrasted with present, the modal auxiliaries *can* and *could*, the quotations "geek, geek" and "This is a chair", metalanguage about her own linguistic abilities, and parallel stylistic structure. The difficulty of simulating such ability led Alan Perlis to remark "A year spent in artificial intelligence is enough to make one believe in God" (1982).

2 WITTGENSTEIN'S ALTERNATIVE

Although Wittgenstein criticized his earlier theory of semantics and related theories by Frege and Russell, he did not reject everything in the *Tractatus*. He continued to have a high regard for logic and mathematics, and he taught a course on the foundations of mathematics, which turned into a debate between himself and Alan Turing. He also retained the picture theory of the *Tractatus*, which considered the relationships among words in a sentence as a *picture* (*Bild*) of relationships in the world. What he abandoned, however, was the claim that there exists a unique decomposition of the world into atomic facts and a privileged vantage point for taking pictures of those facts. A chair, for example, is a simple object for someone who wants to sit down; but for a cabinet maker, it has many parts that must be carefully fit together. For a chemist developing a new paint or glue, even the wood is a complex mixture of chemical compounds, and those compounds are made up of *atoms*, which are not really atomic after all. Every one of those views is a valid picture of a chair for some purpose.

In the *Philosophical Investigations*, Wittgenstein showed that ordinary words like *game* have few, if any common properties that characterize all their uses. Competition is present in ball games, but absent in solitaire or ring around the rosy. Organized sport follows strict rules, but not spontaneous play. And serious games of life or war lack the aspects of leisure and enjoyment. Instead of unique defining properties, games share a sort of *family resemblance*: baseball and chess are games because they resemble the family of activities that people call games. Except for technical terms in mathematics, Wittgenstein maintained that most words are defined by family resemblances. Even in mathematics, the meaning of a symbol is its use, as specified by a set of rules or axioms. A word or other symbol is like a chess piece, which is not defined by its shape or physical composition, but by the rules for using the piece in the game of chess. As he said,

> There are *countless*—countless different kinds of use of what we call 'symbols,' 'words,' 'sentences.' And this multiplicity is not something fixed, given once and for all; but new types of language, new language games, as we may say, come into existence, and others become obsolete and get forgotten. (§23)

As examples of language games, he cited activities in which the linguistic component is unintelligible outside a framework in which the nonlinguistic components are the focus. A child or a nonnative speaker who understood the purpose of the following games could be an active participant in most of them with just a rudimentary understanding of the syntax and vocabulary:

> Giving orders, and obeying them; describing the appearance of an object, or giving its measurements; constructing an object from a description (a drawing); reporting an event; speculating about an event; forming and testing a hypothesis; presenting the results of an experiment in tables and diagrams; making up a story, and reading it; play acting; singing catches; guessing riddles; making a joke, telling it; solving a problem in practical arithmetic; translating from one language into another; asking, thanking, cursing, greeting, praying. (§23)

Only the game of describing an object could be explained in the framework of the *Tractatus*. Wittgenstein admitted that it could not explain its own language game: "My propositions are elucidatory in this way: he who understands me finally recognizes them as senseless..." (6.54). The theory of language games, however, is capable of explaining the language game of writing a book about anything, including language games.

In his later work, Wittgenstein faced the full complexity of language as it is used in science and everyday life. Instead of the fixed boundaries defined by necessary and sufficient conditions, he used the term *family resemblances* for the "complicated network of overlapping and criss-crossing similarities" (1953, §66) in which vagueness is not a defect:

> One might say that the concept 'game' is a concept with blurred edges.—"But is a blurred concept a concept at all?"—Is an indistinct photograph a picture of a person at all? Is it even always an advantage to replace an indistinct picture with a sharp one? Isn't the indistinct one often exactly what we need?
>
> Frege compares a concept to an area and says that an area with vague boundaries cannot be called an area at all. This presumably means that we cannot do anything with it.—But is it senseless to say: "Stand roughly (ungefähr) there"? (§71)

Frege's view is incompatible with natural languages and with every branch of empirical science and engineering. With their background in engineering, Peirce and Wittgenstein recognized that all measurements have a margin of error or granularity, which must be taken into account at every step from design to implementation. The option of vagueness enables language to accommodate the inevitable vagueness in observations and the plans that are based on them.

After a detailed analysis of the *Tractatus*, Simons (1992) admitted that Wittgenstein's later criticisms are valid: "We might say that not everything *we* say *can* be said clearly" (p. 357). But he was not ready to adopt language games as the solution: Wittgenstein "became a confirmed—some, including myself, would say *too* confirmed—believer in the messiness of things." Yet things really are messy. As Eugene Wigner (1960) observed, "the unreasonable effectiveness" of mathematics for representing the fundamental principles of physics is truly surprising. The basic equations, such as $F = ma$, are deceptively simple; even their relativistic or quantum mechanical extensions can be written on one line. The messiness results from the application of the simple equations to the enormous number of atoms and molecules in just a tiny speck of matter. When applied to the simplest living things, such as a bacterium, even the fastest supercomputers are incapable of solving the equations. In any practical calculation, such as predicting the weather, designing a bridge, or determining the effects of a drug, drastic approximations are necessary. Those approximations are always tailored to domain-dependent special cases, each of which resembles a mathematical variant of what Wittgenstein called a language game. In fact, he said "We can get a rough picture of [the language games] from the changes in mathematics" (§23).

Although Wittgenstein's ideas are highly suggestive, his definitions are not sufficiently precise to enable logicians to formalize them. Some confusion is caused by the English term *language game*, which suggests a kind of competition that is not as obvious in the original German *Sprachspiel.* Perhaps a better translation might be *language play* or, as Wittgenstein said, the language used with a specific type of activity in a specific form of life (Lebensform). Hattiangadi (1987) suggested that the meaning of a word is the set of all possible *theories* in which it may be used; each theory would characterize one type of activity and the semantics of the accompanying language game. The term *sublanguage,* which linguists define as a semantically restricted dialect (Kittredge & Lehrberger 1982), may be applied to a family of closely related language games and the theories that determine their semantics. The crucial problem is to determine how the members of such families are related to one another, to the members of other families, and to the growing and changing activities of the people—children and adults—who learn them, use them, and modify them.

3 Models of Language

Any theory of language should be simple enough to explain how infants can learn language and powerful enough to support sophisticated discourse in the most advanced fields of science, business, and the arts. Some formal theories have the power, and some statistical theories have the simplicity. But an adequate theory must explain both and show how a child can grow from a simple stage to a more sophisticated stage without relearning everything from scratch: each stage from infancy to adulthood adds new skills by extending, refining, and building on the earlier representations and operations.

During the second half of the 20th century, various models of language understanding were proposed and implemented in computer programs. All of them have been useful for processing some aspects of language, but none of them have been adequate for all aspects of language or even for full coverage of just a single aspect:

- **Statistical.** In the 1950s, Shannon's information theory and other statistical methods were popular in both linguistics and psychology, but the speed and storage capacity of the early computers were not adequate to process the volumes of data required. By the end of the century, the vastly increased computer power made them competitive with other methods for many purposes. Their strength is in pattern-discovery methods, but their weakness is in the lack of a semantic interpretation that can be mapped to the real world or to other computational methods.

- **Syntactic.** Chomsky's transformational grammar and related methods dominated linguistic studies in the second half of the 20th century, they stimulated a great deal of theoretical and computational research, and the resulting syntactic structures can be adapted to other paradigms, including those that compete with Chomsky and his colleagues. But today, Chomsky's contention that syntax is best studied independently of semantics is at best unproven and at worst a distraction from a more integrated approach to language.

- **Logical.** By the 1970s, the philosophical studies from Carnap and Tarski to Kripke and Montague led to formal logics with better semantic foundations and reasoning methods than any competing approach. Unfortunately, those methods can only interpret sentences that have been deliberately written in a notation that looks like a natural language, but is actually a syntactic variant of the underlying logic. None of them can generate logical formulas from the language that people speak or write for the purpose of communicating with other people.

- **Lexical.** Instead of forcing language into the mold of formal logic, lexical semanticists study all features of syntax, vocabulary, and context that can cause sentences to differ in meaning. The strength of lexical semantics is a greater descriptive adequacy and a sensitivity to more aspects of meaning than other methods. Its weakness is a lack of an agreed definition of the meaning of 'meaning' that can be related to the world and to computer systems.

- **Neural.** Many people believe that neurophysiology may someday contribute to better theories of how people generate and interpret language. That may be true, but the little that is currently known about how the brain works can hardly contribute anything to linguistic theory. Systems called *neural networks* are statistical methods that have the same strengths and weaknesses as other statistical methods, but they have little resemblance to the way actual neurons work.

Each of these approaches is based on a particular technology: mathematical statistics, grammar rules, dictionary formats, or networks of neurons. Each of them ignores those aspects of language for which the technology is ill adapted. For people, however, language is seamlessly integrated with every aspect of life, and they do not stumble over boundaries between different technologies. Wittgenstein's language games do not compartmentalize language by the kinds of technology that produce it, but by subject matter and mode of use. That approach seems more natural, but it raises the question of how a computer could recognize which game is being played, especially when aspects of multiple games are combined in the same paragraph or even the same sentence.

The greatest strength of natural language is its flexibility and power to express any sublanguage ranging from cooking recipes to stock-market reports and mathematical formulas. A flexible syntactic theory, which is also psychologically realistic, is *Radical Construction Grammar* (RCG) by Croft (2001). Unlike theories that draw a sharp boundary between grammatical and ungrammatical sentences, RCG can accept any kind of construction that speakers of a language actually use, including different choices of constructions for different sublanguages:

> Constructions, not categories or relations, are the basic, primitive units of syntactic representation. ... The grammatical knowledge of a speaker is knowledge of constructions (as

> form-meaning pairings), words (also as form-meaning pairings), and the mappings between words and the constructions they fit in. (p. 46)

RCG makes it easy to borrow a word from another language, such as *connoisseur* from French or H_2SO_4 from chemistry, or to borrow an entire construction, such as *sine qua non* from Latin or $x^2 + y^2 = z^2$ from algebra. In the sublanguage of chemistry, the same meaning that is paired with H_2SO_4 can be paired with *sulfuric acid*, and the constructions of chemical notation can be freely intermixed with the more common constructions of English syntax.

A novel version of lexical semantics, influenced by Wittgenstein's language games and related developments in cognitive science, is the theory of *dynamic construal of meaning* (DCM) proposed by Cruse (2000) and developed further by Croft and Cruse (2004). The fundamental assumption of DCM is that the most stable aspect of a word is its spoken or written sign; its meaning is unstable and dynamically evolving as it is construed in each context in which it is used. Cruse coined the term *microsense* for each subtle variation in meaning as a word is used in different language games. That is an independent rediscovery of Peirce's view: the spelling or shape of a sign tends to be stable, but each interpretation of a sign token depends on its context in a pattern of other signs, the physical environment, and the interpreter's memory of previous patterns. Croft and Cruse showed how the DCM view of semantics could be integrated with a version of RCG, but a more detailed specification is required for a computer implementation.

In surveying the difficulties of language translation, Steiner (1975) observed that the most amazing fact about languages is the multiplicity of radically different means for expressing idiosyncratic views of the world:

> No two historical epochs, no two social classes, no two localities use words and syntax to signify exactly the same things, to send identical signals of valuation and inference. Neither do two human beings. Each living person draws, deliberately or in immediate habit, on two sources of linguistic supply: the current vulgate corresponding to his level of literacy, and a private thesaurus. The latter is inextricably a part of his subconscious, of his memories, so far as they may be verbalized, and of the singular, irreducibly specific ensemble of his somatic and psychological identity. Part of the answer as to whether there can be 'private language' is that aspects of every language act are unique and individual. They form what linguists call an 'idiolect'. Each communicatory gesture has a private residue. The 'personal lexicon' in every one of us inevitably qualifies the definitions, connotations, semantic moves current in public discourse. The concept of a normal or standard idiom is a statistically-based fiction (though it may, as we shall see, have real existence in machine translation). The language of a community, however uniform its social contour, is an inexhaustibly multiple aggregate of speech-atoms, of finally irreducible personal meanings. ... Thus a human being performs an act of translation, in the full sense of the word, when receiving a speech-message from any other human being. (pp. 47–48)

The multiplicity of unique language forms, which makes translation difficult even for the best human translators, is an even greater challenge for machine translation. Steiner's remark about a "private thesaurus" for each person's idiolect and a "statistically-based fiction" for MT is intriguing. It suggests the possibility of supporting artificial idiolects by compiling a thesaurus classified according to the language games the machine is designed to play.

4 Semantic Representations

The hypothesis of a prelinguistic semantic representation is as old as Aristotle:

> Spoken words are symbols of experiences (*pathêmata*) in the psyche; written words are symbols of the spoken. As writing, so is speech not the same for all peoples. But the experiences themselves, of which these words are primarily signs, are the same for everyone, and so are the objects of which those experiences are likenesses. (*On Interpretation* 16a4)

Whether that representation is called experience in the psyche, conceptual structure, language of thought, or natural logic is less important than its expressive power, its topological structure, and the kinds of operations that can be performed with it and on it.

Some representations are designed to support Steiner's informal "aggregates of speech atoms" or "irreducible personal meanings", but others force language into a rigid, logic-based framework. From his work as a lexicographer, Peirce realized that symbols have different meanings for different people or for the same person on different occasions:

> For every symbol is a living thing, in a very strict sense that is no mere figure of speech. The body of the symbol changes slowly, but the meaning inevitably grows, incorporates new elements and throws off old ones. (CP 2.222)

But as a mathematician and logician, he also recognized the importance of discipline and fixed definitions: "Reasoning is essentially thought that is under self-control" (CP 1.606). Yet self-control is always exercised for a specific purpose. As the purpose changes, the language game changes, and the symbols acquire new meanings.

Although there is no direct way of observing the internal representations, many of their properties can be inferred from the features of natural languages and the kinds of reasoning people express in languages, both natural and artificial. Any adequate theory must directly or indirectly support the following features:

1. Every natural language has a discrete set of meaningful units (words or morphemes), which are combined in systematic ways to form longer phrases and sentences.

2. The basic constructions for combining those units express relational patterns with two or three arguments (e.g., a subject, an optional direct object, and an optional indirect object). Additional arguments are usually marked by prepositions or postpositions.

3. The logical operators of conjunction, negation, and existence are universally present in all languages. Other operators (e.g., disjunction, implication, and universal quantification) are more problematical.

4. Proper names, simple pronouns, and indexicals that point to something in the text or the environment are universal, but some languages have more complex systems of anaphora than others.

5. Metalanguage occurs in every natural language, and it appears even in Laura Limber's remark at age three. It supports the introduction of new words, new syntax, and the mapping from new features to older features and to extralinguistic referents.

6. Simple metalanguage requires at least one level of nested structure. Most major languages support multiple levels of nested clauses and phrases, any of which could contain metalevel comments.

Points #1 and #2 indicate that the semantic representation must support graph-like structures (of which strings and trees are special cases). With the addition of points #3 and #4, it supports a subset of first-order logic. Full FOL would require a flexible syntax that can support nested or embedded constructions, which English and other major languages provide. Points #5 and #6, combined with a flexible syntax, can support highly expressive logical constructions.

As this summary shows, natural languages can express complex logic, but it does not imply that complex logic is a prerequisite for language. Infants successfully use language to satisfy their needs as soon as they begin to utter single words and short phrases. Preschool children learn and use complex language long before they learn any kind of mathematics or formal logic. Although all known natural languages have complex syntax, some rare languages, such as Pirahã (Everett 2005), seem to lack the levels of nesting needed to express full FOL. Everett noted that the Pirahã people have no word for *all* or *every* or even a logically equivalent paraphrase. That limitation would make it hard for them to invent mathematics and formal logic. In fact, their ability to count is limited to the range *one, two, many.*

An adequate semantic representation must be able to cover the full range of language used by people in every culture at every stage of life. In modern science, educated adults create and talk about abstruse systems of logic and mathematics. But the Pirahã show that entire societies can live successfully with at best a rudimentary logic and mathematics. As Peirce observed, logical reasoning is a disciplined method of thought, not a prerequisite for thought—or the language that expresses it.

5 A Wittgensteinian Approach to Language

A semantic approach inspired by Wittgenstein's language games was developed by Margaret Masterman, one of six students in his course of 1933-1934 whose notes were compiled as *The Blue Book* (Wittgenstein 1958). In the late 1950s, Masterman founded the Cambridge Language Research Unit (CLRU) as a discussion group, which became one of the pioneering centers of research in computational linguistics. Her collected papers (Masterman 2005) present a computable version with similarities to Cruse's DCM:

- A focus on semantics, not syntax, as the foundation for language: "I want to pick up the relevant basic-situation-referring habits of a language in preference to its grammar" (p. 200).

- A context-dependent classification scheme with three kinds of structures: a thesaurus with multiple groups of words organized by areas of use, a *fan* radiating from each word type to each area of the thesaurus in which it occurs, and dynamically generated combinations of fans for interpreting the word tokens of a text.

- Emphasis on images as a language-independent foundation for meaning with a small number (about 50 to 100) of combining elements represented by ideographs or monosyllables, such as IN, UP, MUCH, THING, STUFF, MAN, BEAST, PLANT, DO.

- Recognition that analogy and metaphor are fundamental to the creation of novel uses of language, especially in the most advanced areas of science. In electromagnetism, for example, Maxwell's elegant mathematics is the culmination of a lengthy process that began with Faraday's vague analogies about lines of force.

Figure 1 shows a fan for the word *bank* with links to each area of Roget's *Thesaurus* in which the word occurs (p. 288). The numbers and labels identify areas in the thesaurus, which, Masterman claimed, correspond to "Neo-Wittgensteinian families".

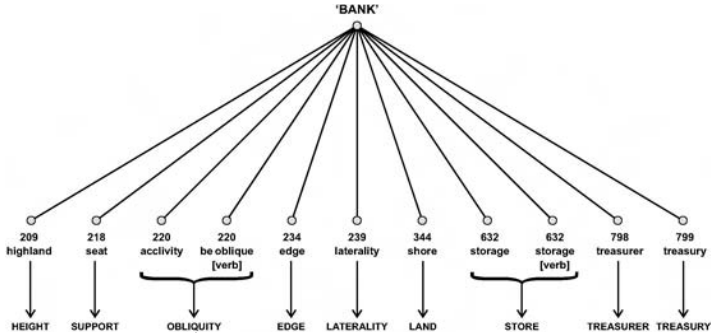


Figure 1: A word fan for *bank*

To illustrate the use of word fans, Masterman analyzed the phrases *up the steep bank* and *in the savings bank.* All the words except *the* would have similar fans, and her algorithm would "pare down" the ambiguities "by retaining only the spokes that retain ideas which occur in each." For this example, it would retain "OBLIQUITY 220 in 'steep' and 'bank'; whereas it retains as common between 'savings' and 'bank' both of the two areas STORE 632 and TREASURY 799." She went on to discuss methods of handling various exceptions and complications, but all the algorithms use only words and families of words that actually occur in English. They never use abstract or artificial markers, features, or categories. That approach suggests a plausible cognitive theory: From an infant's first words to an adult's level of competence, language learning is a continuous process of building and refining the stock of words, families of words grouped by use in the same contexts, and patterns of connections among the words and families.

Wittgenstein's language games and the related proposals by Cruse, Croft, and Masterman are more realistic models of natural language than the rigid theories of formal semantics. Yet scientists, engineers, and computer programmers routinely produce highly precise language-like structures by disciplined extensions of the methods used for ordinary language. Furthermore, the level of precision needed to write computer programs can be acquired by school children without formal training. A complete theory of language must be able to explain every level of competence from the initial vague stages to the most highly disciplined representations and reasoning methods of science. Different language games may require attention to different details with different granularity, but there is no evidence for a discontinuity in the methods of language generation and understanding.

6 Language Games as a Basis for Semantics

To handle both formal and informal language, Masterman's approach must be extended with links to logic, but in a way that permits arbitrary revisions, changes of perspective, and levels of granularity. Figure 2 illustrates the issues in relating logic, models, and the world. At the right is a theory expressed in the Peirce-Peano notation for logic. In the middle is a formal model shown as a graph in which nodes represent objects and arcs represent relations among those objects. With varying degrees of formality, logicians from Aristotle and the medieval Scholastics to Bolzano, Peirce, Wittgenstein, and Tarski reached a consensus on how to evaluate the denotation of a proposition in terms of a model. But on the left of Figure 2, the mapping of models to the world is an approximation that raises the most contentious issues. As the engineer and statistician George Box (2005) said, "All models are wrong; some are useful."

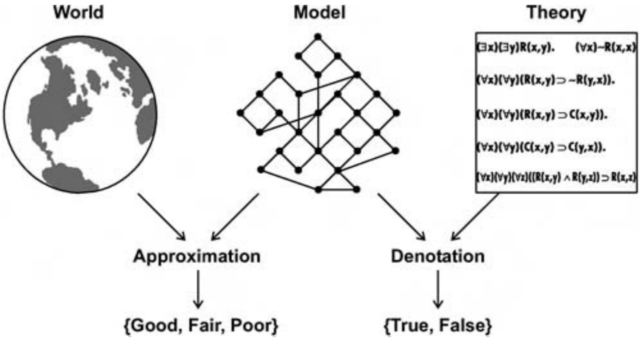


Figure 2: The world, a model, and a theory

The approximate mapping of models to the world is the source of the vagueness that must be addressed in every theory of epistemology, ontology, phenomenology, and philosophy of science. In the *Tractatus*, Wittgenstein assumed an exact mapping from language to logic to models to the world. As he said,

> "The totality of true thoughts is a picture of the world" (3.01). "The picture is a model of reality" (2.12). "The proposition is a picture of reality, for I know the state of affairs presented by it, if I understand the proposition" (4.021). "Reality is determined by the truth or falsity of the proposition; it must therefore be completely described by the proposition" (4.023).

Tarski (1933) was more cautious. He avoided the complexities of natural language and the world by limiting his claims to the relationship between a formal language and a model. Carnap, Kripke, and Montague extended Tarski's approach to modal logic by assuming a multiplicity of models, one of which represents the real world and the others represent possible worlds. Barwise and Perry (1983) avoided a giant model of everything by assuming finite chunks of the world called *situations*. Yet as Devlin (1991) observed, nobody could state the criteria for selecting significant chunks: "Situations are just that: situations. They are abstract objects introduced so that we can handle issues of context, background, and so on." In short, situations determine meaning, but there are no criteria for distinguishing a meaningful situation from an arbitrary chunk of space-time.

In his later philosophy, Wittgenstein shifted the focus from abstract mappings between language and the world to the human activities that give meaning to chunks of the world and the language about them. To accommodate language games in a framework that can represent any theory about any model for any purpose, Sowa (2000) proposed an infinite lattice of all possible theories expressible in a given logic. Each theory would represent the rules or axioms of one language game or a family of closely related games. The lattice is a generalization hierarchy, in which the most general theory at the top is true for every possible model; the bottom is the inconsistent theory that is false for every model. Every theory in between is true for a subset of the models of the theories above it and a superset of the models of the theories below it. Figure 3 shows the four basic operators for navigating the lattice: *contraction, expansion, revision,* and *analogy.*

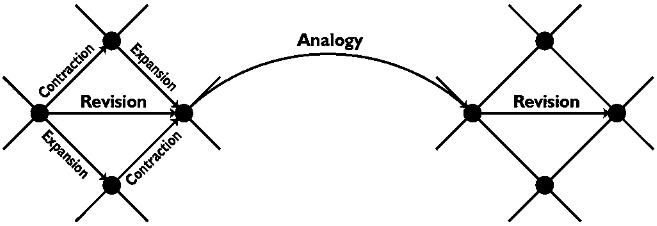


Figure 3: Four operators for navigating the lattice of theories

The operators of contraction and expansion follow the arcs of the lattice, revision makes short hops sideways, and analogy makes long-distance jumps. The first three operators, which delete and add axioms, correspond to the AGM operators for theory revision (Alchourrón et al. 1985). The analogy operator makes longer jumps through the lattice by systematically relabeling the names of types and relations. All methods of nonmonotonic reasoning can be viewed as strategies for walking or jumping through the lattice in order to find a theory that is a suitable approximation to some aspect of the world for some purpose:

1. *Induction* is an expansion strategy for increasing the number of provable statements (theorems) while reducing the number of assumptions (axioms).

2. *Abduction* is another expansion strategy, which often uses analogy to "guess" or hypothesize a likely theory, whose predictions by deduction are tested against further observations.

3. *Default logics* can be considered shorthand descriptions for families of closely related theories. The *supremum* or most specific common generalization of all theories in a family is the classical theory obtained by ignoring all defaults. Other theories in the family are obtained by expanding the supremum with one or more of the defaults.

4. *Negation by failure* is a variant of default logic. The supremum is a theory defined by the

conjunction of a given set of axioms. Each failure to prove some proposition p expands the current theory with the negation ~p.

5. Reasoning methods that use *certainty factors* or *fuzzy values* can be viewed as variants of a default logic in which each proposition has a metalevel measure of its approximation to some aspect of the world. The result of fuzzy reasoning is a theory whose propositions exceed some minimum level of approximation.

These reasoning methods have a common goal: the discovery or construction of an appropriate theory somewhere in the lattice. Combinations of various methods may be applied iteratively to derive theories whose models are better and better approximations to the world.

Figure 4 illustrates a word fan that maps the words of a language to concept types to canonical graphs and to a lattice of theories. The fan on the left of Figure 4 links each word to an open-ended list of *concept types*, each of which corresponds to some area of a thesaurus, as in Masterman's system. The word *bank*, for example, could be linked to types with labels such as Bank799 or Bank_Treasury.

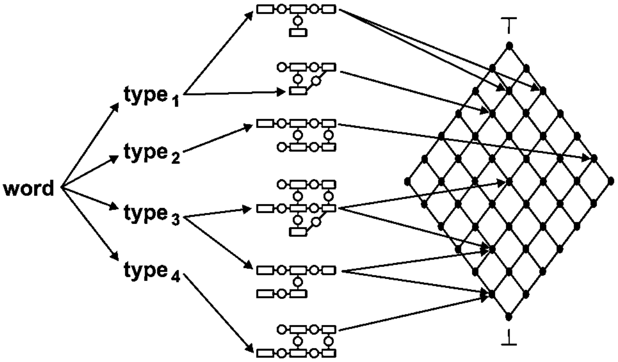


Figure 4: Words → types → canonical graphs → lattice of theories

In various language games, those types could be further specialized in subtypes, which would correspond to Cruse's microsenses. When precision is necessary, the lattice enables any theory to be specialized, revised, or refined in order to tighten the constraints or add any amount of detail. In a formal logic, vagueness is not possible, but vagueness in natural language can be represented in two ways: first, the types and theories at the upper levels of the lattice may be underspecified to include a broad range of more specialized language games at lower levels; second, some canonical graphs may lead to more than one theory, and further information may be needed to determine which one is intended.

For this article, canonical graphs are represented by *conceptual graphs* (CGs), a formally defined version of logic that uses the model-theoretic foundation of Common Logic (ISO/IEC 2006). Equivalent operations may be performed with other notations, but graphs support highly structured operations that are computationally more efficient and cognitively more realistic than the rules of inference of predicate calculus (Sowa and Majumdar 2003). Figure 5 illustrates three canonical graphs for the types `Give`, `Easy`, and `Eager`.

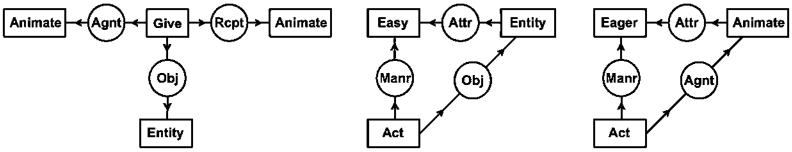


Figure 5: Canonical graphs for the types **`Give`**, **`Easy`**, and **`Eager`**

A canonical graph for a type is a conceptual graph that specifies one of the patterns characteristic of that type. On the left, the canonical graph for `Give` represents the same constraints as a typical *case frame* for a verb. It states that the agent `(Agnt)` must be `Animate`, the recipient `(Rcpt)` must be `Animate`, and the object `(Obj)` may be any `Entity`. The canonical graphs for `Easy` and `Eager`, however, illustrate the advantage of graphs over frames: a graph permits cycles, and the arcs can distinguish the directionality of the relations. Consider the following two sentences:

Bob is easy to please.   *Bob is eager to please.*

For both sentences, the concept `[Person: Bob]` would be linked via the attribute relation `(Attr)` to the concept `[Easy]` or `[Eager]`, and the act `[Please]` would be linked via the manner relation `(Manr)` to the same concept. But the canonical graph for `Easy` would make Bob the object of `Please`, and the graph for `Eager` would make Bob the agent. The first sentence below is acceptable because the object may be any entity, but the constraint that the agent of an act must be animate would make the second unacceptable:

The book is easy to read.   * *The book is eager to read.*

Chomsky (1965) used the easy/eager example to argue for different syntactic transformations associated with the two adjectives. But the canonical graphs state semantic constraints that cover a wider range of linguistic phenomena with simpler syntactic rules. A child learning a first language or an adult reading a foreign language can use semantic constraints to interpret sentences with unknown or even ungrammatical syntax. Under Chomsky's hypothesis that syntax is a prerequisite for semantics, such learning is inexplicable.

Canonical graphs with a few concept nodes are adequate to discriminate the general senses of most words, but the canonical graphs for detailed microsenses can become much more complex. The microsenses for the adjective *easy* occur in very different patterns for a book that is easy to read, a person that is easy to please, or a car that is easy to drive. For the verb *give*, a large dictionary lists dozens of senses, and the number of microsenses is enormous. The prototypical act of giving is to hand something to someone, but a large object can be given just by pointing to it and saying "It's yours." When the gift is an action, as in giving a kiss, a kick, or a bath, the canonical graph used to parse the sentence has a few more nodes. But the graphs required to understand the implications of each type of action are far more complex, and they are related to the graphs for taking a bath or stealing a kiss.

The canonical graph for *buy* typically has two acts of giving: money from the buyer to the seller, and goods from the seller to the buyer. But the graphs for specialized microsenses may have far more detail about the buyers, the sellers, the goods sold, and other people, places, and things involved. Buying a computer, for example, can be done by clicking some boxes on a screen and typing the billing and shipping information. That process may trigger a series of international transactions, which can be viewed by going to the UPS web site to check when the computer was airmailed from Hong Kong and delivered to New York. In talking or reasoning about a successful purchase, most of the detail can be ignored, but it may become important if something goes wrong.

7 Language, Logic, and Lebensform

The role of logic in natural language semantics is a controversial issue. Although Montague rejected Chomsky's emphasis on syntax, he adopted Chomsky's distinction between competence and performance, but with semantics at the focus. Instead of an idealized syntax that characterizes the ultimate human competence, Montague (1970) assumed "a theory of truth, of a formal language that I believe may be reasonably regarded as a fragment of ordinary English." But a cognitively realistic theory must also address the question of how that competence is acquired. At age three, Laura Limber correctly used the words *can* and *could* to contrast her own linguistic abilities at different points in time. Presumably that implies a competence for conceiving different contexts, comparing what is possible in each, and expressing her conclusions in English. Yet it seems unlikely that a three-year-old child would have the full logical machinery of Montague's possible worlds.

Linguists and logicians working in Montague's tradition have refined, extended, and restricted his logic in various ways. Fox and Lappin (2005), for example, developed Property Theory with Curry Typing (PTCT) as "a first-order representation language that provides fine-grained intensionality, limited expressive power, and a richly expressive type system." Any such proposal for an ideal formal logic raises some serious issues:

1. Is that formal logic innate? Or does a child learn it in successive stages? As Laura's speech indicates, the semantics for some version of metalanguage and modal logic is acquired very early. But how expressive are those early stages, and how are they learned?

2. Languages such as Pirahã show that an entire community can live successfully without having any native speaker who has achieved the logical sophistication assumed by systems such as Montague's or PTCT. Does that imply that different languages have different kinds of semantic competence? Or that some do not reach the ultimate level of human competence? Or that semantics can be revised and extended indefinitely with no fixed limit?

3. Scientists often invent radically new theories whose mathematical foundations are quite different from any version of formal semantics. When two mathematicians talk about their theories on the telephone, they use the linguistic forms of their native language without the aid of other notations. Does that imply that the formal logic that characterizes their speech must incorporate the semantics of the mathematics they conceived? Does there exist any fixed logic that can characterize everything that is humanly conceivable? Or does Gödel's undecidability theorem rule out that possibility?

4. Did human semantic competence evolve from a more primitive stage around the time of *Homo habilis*, about two million years ago? Or did it spring full-blown into the psyche of Adam and Eve, perhaps 60 thousand years ago? If it did not evolve, why did the human vocal tract and brain size take a few million years to attain their current forms? If it did evolve, what kinds of intermediate stages could there be?

In the *Tractatus*, Wittgenstein proposed a first-order semantic theory that was far more restricted than Montague's or PTCT. It could not characterize the speech of Laura Limber or his pupils in the Austrian village. In the *Philosophical Investigations*, he said that "to imagine a language (eine Sprache vorstellen) is to imagine a form of life (Lebensform)" (§19). Every form of life determines one or more language games, which impose requirements on the expressive power of the associated logic. The various forms of life would include the activities of hunting and gathering by the Pirahã or the so-called "civilized" activities of shopping in a supermarket, reporting a medical diagnosis, and directing traffic around a construction site. Each activity involves constraints imposed by the culture and the environment, which determine the vocabulary, the semantic patterns, and the conventional moves in the corresponding language game.

These considerations suggest that the goal of a fixed formal semantics for all of language is as unrealistic as Hilbert's goal of a fixed foundation for all of mathematics. For many language games, the semantics could be logically simpler than anything required for a general theory of everything. But when new circumstances require changes in the old games or the invention of a totally new game, more complex logical features may be required. The quoted sentence by Laura, which is considerably more complex than most of her utterances at that age, illustrates an important principle: even though most sentences express a rather simple logic, the logical and syntactic complexity increases when someone compares different language games, suggests an innovation in an old game, or proposes a totally new one.

The questions of how language and logic are learned are fundamental to understanding the role of logic in semantics. Frege and Russell, for example, adopted the universal quantifier, negation, and material implication as their three primitives. But those three operators are among the most problematical—logically, linguistically, computationally, and pedagogically. Following is a brief summary of the issues:

- **Existential-conjunctive logic.** Conjunction and the existential quantifier are the two operators that are central to all uses of language. They are the only two that are necessary for observation statements, they are the two most frequently occurring operators in translations from language to logic, and they are needed to represent a child's earliest utterances.

- **Negation.** Words for negation also occur very early in a child's speech, but they raise an enormous number of questions. What aspects of the utterance or the environment do they negate? And do they represent the denial, rejection, absence, or prohibition of those aspects? Many languages use different words or syntax to distinguish different varieties of negation, which must be related to one another.

- **Other logical operators.** Conjunction, negation, and the existential quantifier are sufficient to define all the other operators of first-order logic, but all the problems of negation are inherited by every operator defined in terms of it. Words for many of those operators occur in most natural languages, but they are not the first to be learned, and their semantics is rarely identical to the usual definitions in classical FOL.

- **Commands, statements, and questions.** Imperatives, such as crying for food or attention, precede language, and many of an infant's earliest utterances are refinements of those cries. Without imperatives and interrogatives, declaratives can only paint a static picture of the world. Commands and questions animate that picture and integrate it with the activities that give it meaning.

- **Speech acts.** Peirce, Wittgenstein, Austin, and others studied the use of language, the purpose or intention of any particular statement, and its role in relation to the speaker, listener, discourse, environment, and accompanying activity. Without considering the use, it is impossible for anyone to understand an infant's utterances and often misleading to try to understand an adult's.

- **Context.** Most versions of logic are deliberately designed to have a context-free syntax, but almost all aspects of natural language are context sensitive. Although the word *context* has multiple senses, just the basic definition as a chunk of text is sufficient to raise the questions: How are the contents of one chunk of text related to other chunks, to the environment, to the participants in the discourse, and to the goals of the participants?

- **Metalanguage.** From infancy, children are surrounded by language about language, which they imitate successfully by their third year: praise, blame, corrections, prompts, explanations, definitions, and examples of how language maps to things, activities, and people, including themselves. All the tenses and modalities of verbs are metalinguistic commentary, which can be defined by language about language or logic about logic (Sowa 2003).

- **Propositions.** Some metalanguage is about syntax and vocabulary, but much of it is about language-independent propositions. Many logicians avoid the notion of proposition by talking only about sentences, but that approach ignores the fact that people find it easier to remember what was said than to remember how it was said. Other logicians identify a proposition with the set of possible worlds in which it is true, but that definition is much too coarse grained. It cannot distinguish 2+2=4 from Fermat's last theorem.

- **Fuzziness.** Hedges and "fuzzy" qualifiers such as *almost* or *nearly* have spawned a variety of fuzzy or multivalued logics with a range of truth values or certainty factors between 1 for true and 0 for false. But many logicians have pointed out the problems with interpreting those numbers as truth values. A more nuanced approach should observe the distinction in Figure 2: truth values are metalevel commentary about the mapping of a sentence to a model; fuzzy values estimate the adequacy of the model as an approximation to the world for the purpose of a given language game.

Conventional model theory, by itself, is insufficient to accommodate all these aspects of language in a cognitively realistic formalism. Although Wittgenstein contributed to that paradigm, he recognized its limitations and proposed language games as an alternative. The challenge is to formalize language games and integrate them with related research in cognitive science.

Some promising techniques published decades ago were ignored because they did not fit the popular paradigms. Among them are the *surface models* by Hintikka (1973). Like situations, surface models are finite. But unlike situations, which are considered chunks of the world, surface models are constructed as approximations to the world, as in Figure 2. Instead of trying to define criteria for meaningful situations, Hintikka proposed a method for constructing surface models that represent the individuals and relations explicitly mentioned in the discourse. In that same book, Dunn (1973) published an alternative semantics for modal logics based on sets of *laws* and *facts*. Each possible world *w* is replaced by a pair of sets (**M**,**L**), in which **M** consists of all facts that are true in *w* and **L** consists of the laws of *w*—the subset of facts that are necessarily true. Dunn showed that this construction is isomorphic to Kripke semantics, it avoids the dubious ontology of possible worlds, and it treats accessibility as a derived relation instead of a primitive. Sowa (2003) showed that Dunn's semantics simplifies the computational and the theoretical methods by treating multimodal reasoning as metalevel reasoning about the choice of laws. These techniques can be combined with the lattice of theories to formalize language games:

- For each type of language game *g*, define a set **L** of propositions as the laws, rules, or axioms of a theory that characterizes any game of type *g*.
- During the play of a game of type *g*, construct a surface model that is derived from the facts that are consistent with **L** and known or assumed to be true as a result of statements during the play.
- Specialized theories at lower levels of the lattice represent the axioms of possible games, and generalizations higher in the lattice represent the axioms common to a family of games.
- Since any game may be associated with extralinguistic activity, some observable facts about individuals, states, and events may be incorporated in the surface model without being explicitly mentioned in language.
- Any fact that is inconsistent with the current game triggers theory revision operations that move through the lattice to find a theory of a related game consistent with that fact.

This approach retains the power and precision of formal methods within a dynamically extensible or negotiable framework. The construction of a surface model need not be monotonically increasing, since various statements, observations, and objections may trigger revisions—either to the surface model or to the laws of the language game that governs its construction. The result of a successful dialog or negotiation is a surface model that is consistent with the axioms of some theory in the lattice and the facts agreed or observed during the discourse. But not all discourse reaches a settled conclusion. Some participants may refuse to accept some statements about the laws and facts, or they may take action to change them.

The most promising and most neglected work is Peirce's research on semiotics and its relationships to both logic and language (Sowa 2006). Many aspects that Peirce discovered, anticipated, or developed in detail are usually associated with other philosophers and logicians:

- Tarski: model theory and metalanguage.
- Davidson: event semantics.
- Austin: speech acts.
- Grice: conversational implicatures.
- Perry: the essential indexical.

- Kamp: nested contexts for discourse representation structures.
- Carnap, Kripke, Montague: possible worlds.

Some of the more recent developments have gone into much greater detail than Peirce had. But Peirce demonstrated that these and other aspects of language are part of a unified vision. Furthermore, Peirce's "left-handed brain", as he called it, often put old ideas in surprisingly new perspectives.

As an example, Peirce observed that a proposition corresponds to "an entire collection of equivalent propositions with their partial interpretants" (CP 5.569). To formalize that insight, a proposition may be defined as an equivalence class of sentences in some language **L** under some meaning-preserving translation (MPT) defined over the sentences of **L**. An MPT is then defined as any function *f* over sentences of **L** that satisfies four constraints: invertible, truth preserving, vocabulary preserving, and structure preserving. If *f* satisfies only the first two constraints, the equivalence classes are much too big: each would consist of all sentences that are true in a given set of possible worlds. Furthermore, that function is not efficiently computable because proving that 2+2=4 is in the same equivalence class as Fermat's last theorem took three centuries of research by the best mathematicians in the world.

If the constraints on vocabulary and structure are too strong, the MPT *f* becomes the identity function, which is trivially computable, but it leaves only one sentence type in each class. By imposing reasonable constraints on vocabulary and structure, Sowa (2000) defined several MPTs that are cognitively realistic and computable in polynomial or even linear time. These functions can be specified in just a few lines, the translations can be learned in pedagogically simple steps, and the method is sufficiently flexible to allow different options of MPTs for different language games. By contrast, the proposal of Curry typing (Fox and Lappin 2005) is a fixed, rigid system that takes 40 pages to specify and makes no provision for learnability.

In summary, language games can be formalized in an open-ended framework that can accommodate any use of language for any purpose. At one extreme are the versions of mathematics and logic with specialized ontologies designed for science and engineering. At the other extreme are the vague ideas and insights whose consequences are not well understood. In between are the discussions, negotiations, compromises, and analyses that are necessary to translate a vague idea to a precise plan or to revise the plan as circumstances change. To adopt this approach requires a major paradigm shift in formal semantics. It does not reject logic, but it applies logic to a broader range of problems with a greater sensitivity to the way language is actually used by people at every stage of life.

REFERENCES

Alchourrón, C., P. Gärdenfors and D. Makinson (1985). On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, **50**, 510-530.

Box, G. E. P., J. S. Hunter and W. G. Hunter (2005). *Statistics for Experimenters: Design, Innovation, and Discovery* (2nd edition). Wiley-Interscience, New York.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Mass.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press, Oxford.

Croft, W. and D. A. Cruse (2004). *Cognitive Linguistics.* Cambridge University Press, Cambridge.

Cruse, D. A. (2000). Aspects of the micro-structure of word meanings. In: *Polysemy: Theoretical and Computational Approaches* (Y. Ravin and C. Leacock, eds.), pp. 30-51. Oxford University Press, Oxford.

Devlin, K. (1991). Situations as mathematical abstractions. In: *Situation Theory and its Applications* (J. Barwise, J. M. Mark Gawron, G. Plotkin and S. Tutiya, eds.), pp. 25-39. CSLI Publications, Stanford.

Dunn, J. M. (1973). A truth value semantics for modal logic. In: *Truth, Syntax and Modality* (H. Leblanc, ed.), pp. 87-100. North-Holland, Amsterdam.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology*, **46**, 621-646.

Fox, C. and S. Lappin (2005). *Foundations of Intensional Semantics.* Blackwell, Oxford.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought.* MIT Press, Cambridge, Mass.

Hattiangadi, J. N. (1987). *How is Language Possible? Philosophical Reflections on the Evolution of Language and Knowledge.* Open Court, La Salle.

Hintikka, J. (1973). Surface semantics: definition and its motivation. In: *Truth, Syntax and Modality* (H. Leblanc, ed.), pp. 128-147. North-Holland, Amsterdam.

ISO/IEC (2006). *Common Logic: A Framework for a Family of Logic-Based Languages.* Final Committee Draft. http://cl.tamu.edu.

Kittredge, R. and J. Lehrberger (eds.) (1982). *Sublanguage: Studies of Language in Restricted Semantic Domains.* de Gruyter, New York.

Majumdar, A. K., J. F. Sowa and P. Tarau (2007). Graph-based algorithms for intelligent knowledge systems. In: *Handbook of Applied Algorithms* (A. Nayak and I. Stojmenovic, eds.), Wiley & Sons, New York.

Masterman, M. (2005). *Language, Cohesion and Form* (Y. Wilks, ed.). Cambridge University Press, Cambridge.

Montague, R. (1970). English as a formal language. In: *Formal Philosophy* (R. Montague, 1974), pp. 188-221. Yale University Press, New Haven.

Peirce, C. S. (1931-1958). (CP) *Collected Papers of Charles S. Peirce* (C. Hartshorne, P. Weiss and A. Burks, eds., 8 vols.). Harvard University Press, Cambridge, Mass.

Perlis, A. J. (1982). Epigrams in programming. *SIGPLAN Notices*, ACM. www.cs.yale.edu-/homes/perlis-alan/quotes.html

Simons, P. (1992). *Philosophy and Logic in Central Europe from Bolzano to Tarski.* Kluwer, Dordrecht.

Sowa, J. F. (1984). *Conceptual Structures: Information Processing in Mind and Machine.* Addison-Wesley, Reading, Mass.

Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations.* Brooks/Cole, Pacific Grove.

Sowa, J. F. (2003). Laws, facts, and contexts: Foundations for multimodal reasoning. In: *Knowledge Contributors* (V. F. Hendricks, K. F. Jørgensen and S. A. Pedersen, eds.), pp. 145-184. Kluwer, Dordrecht. http://www.jfsowa.com/pubs/laws.htm

Sowa, J. F. (2006). Peirce's contributions to the 21st Century. In: *Conceptual Structures: Inspiration and Application* (H. Schärfe, P. Hitzler and P. Øhrstrøm, eds.), pp. 54-69. Lecture Notes in Artificial Intelligence **4068**, Springer, Berlin.

Sowa, J. F. and A. K. Majumdar (2003). Analogical reasoning. In: *Conceptual Structures for Knowledge Creation and Communication* (A. de Moor, W. Lex and B. Ganter, eds.), pp. 16-36. Lecture Notes in Artificial Intelligence **2746**, Springer-Verlag, Berlin. http://www.jfsowa.com/pubs/analog.htm

Steiner, G. (1975). *After Babel: Aspects of Language and Translation* (3rd edition, 1998). Oxford University Press, Oxford.

Tarski, A. (1933). Der Wahrheitsbegriff in den formalisierten Sprachen (The concept of truth in formalized languages). In: *Logic, Semantics, Metamathematics* (2nd edition), pp. 152-278. Hackett Publishing, Indianapolis.

Wigner, E. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, **13**.

Winograd, T. (1972). *Understanding Natural Language.* Academic Press, New York.

Winograd, T. and F. Flores (1986). *Understanding Computers and Cognition.* Ablex, Norwood.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus.* Routledge & Kegan Paul, London.

Wittgenstein, L. (1953). *Philosophical Investigations.* Blackwell, Oxford.

Wittgenstein, L. (1958). *The Blue and Brown Books.* Blackwell, Oxford.

Woods, W. A. (1968). Procedural semantics for a question-answering machine. *AFIPS Conference Proc.*, pp. 457-471. FJCC.

Woods, W. A., R. M. Kaplan and B. L. Nash-Webber (1972). *The LUNAR Sciences Natural Language System.* Final Report, NTIS N72-28984.

This page intentionally left blank

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 3

COUNTERFEITING TRUTH: STATISTICAL REPORTING ON THE BASIS OF TRUST

David M. Levy
George Mason University

Sandra J. Peart
Baldwin-Wallace College

1 INTRODUCTION

Semantics and game theory offer modern approaches to very old problems.[1] David Lewis introduced game theoretic concepts into the study of language in his examination of conventions.[2] In this chapter we study the language of a specific sort of conventions: statistical estimators. Such estimators have the important property of being both well-defined mathematical objects and devices that form the basis of factual claims asserted and, perhaps, believed by rational agents.[3] The convention we analyze allows econometric reporting to proceed on the basis of trust.[4] In

[1] Carnap (1942, pp. v–vi): "Semantical concepts, especially the concept of truth, have been discussed by philosophers since ancient times. But a systematic development with the help of the exact instruments of modern logic has been undertaken only in recent years. ... On the basis of these preliminary analyses, Alfred Tarski (who is now in this country) laid the foundation of a systematical construction." Tarski's work is central to that of Carnap (1942, p. vi) and Quine (1940, p. 4), among others. Luschei (1962) is a full-length attempt that uses manuscript and memory to recover the contributions of Stanislaw Lesniewski.

[2] Barwise & Moss (1996, p. 4): "The philosopher David Lewis uncovered a deep source of circularity in human affairs, described in his famous study of convention (Lewis, 1969). All social institutions, from language to laws to customs about which side of the sidewalk to use, are based on conventions shared by the community in question. But what does it mean for a society to share a convention? Certainly, part of what it means is that those who accept some convention, say, C, behave in a given away. But Lewis also argues that another important part of what makes C a convention is that those who accept C also accept that C is a shared convention."

[3] Lewis (1969, p. 204): "One kind of semantics analyzes truth, analyticity, and the rest in relation to possible interpreted languages, in abstraction from any users thereof. This is the kind of semantics done by Frege, Tarski, and (most of the time) Carnap. ... The other kind of semantics analyzes truth, analyticity, and the rest, in relation to an agent or a population of agents. This is the kind of semantics done by the later Wittgenstein, Grice, Skinner, Quine, Morris, Ziff, and (sometimes) Carnap."

[4] Dewald et al. (1986) first publicly demonstrated how hard it was, even for journal editors, to obtain the data used to obtain published estimates. Without the data it is difficult to reproduce the published results. Are publishing

contrast with Lewis, we shall demonstrate that such a convention is conducive to conflict rather than co-ordination.

Long before game theory and semantics, indeed, long before economics itself, exchange conducted by means of money was linked to language. In the *Republic* (371c) Socrates talks about "money as a token for the purpose of exchange." Economists have long argued that, for money to function as a mechanism of exchange, there must be some assurance—carried by institutions and language—of its quality. Our argument is simple. Supposing money and language are interrelated the way that philosophers and economists often claim they are interrelated, if we do not take money solely on the basis of trust, why do we take claims regarding truth on the basis of trust?

There are two parts of our chapter. First, we review Adam Smith's argument that the evolution of monetary institutions is tied up in the problem of detecting deceitful metal offered in exchange. Smith points to no such comparable institution by which deceitful policy advocacy is detected and severely punished.[5] Yet his recommendation for caution in the evaluation of policy advocacy points to the caution that routinely prevailed in monetary matters before public safeguards evolved to make the metallic content of the medium of exchange transparent and to preserve its quality. Second, we turn to a different sort of deceit, in the reporting of statistical evidence. We apply Smith's insights regarding counterfeit money to the case of incentives for deceit in reporting statistical results. In the production of "truth", there is no evolved institution that compares to the Mint. We summarize our recent work regarding how another institution—competing expert witnesses—might deal with deceitful statistical arguments.

We juxtapose these two broad topics, money and truth telling, to emphasize the common structure they share, that of an institutional framework that relies (rightly or wrongly) on trust carried by language. It is important to emphasize, in addition, that these are part of our larger enterprise. Economists model ordinary people as seeking the private good of happiness. Yet we persist in thinking of *ourselves*, qua economists, as seeking the public good of truth. And we have failed to confront the inconsistency in such a modeling procedure (Peart & Levy, 2005).

2 ADAM SMITH ON DECEIT

As economists have only recently re-acquainted themselves with language as an object of study (Rubinstein, 2000), a passage from Smith's *Lectures on Jurisprudence* that links money and language might not come readily to mind:

> The offering of a shilling, which to us appears to have so plain and a simple a meaning, is in reality offering an argument to persuade one to do so and so as it is for his interest. Men always endeavour to persuade others to be of their opinion even when the matter is of no consequence to them... (1978, 352)

If offering money is a form of persuasion wrapped up in the semantic notions of meaning and truth, then what is the semantic counterpart of counterfeiting money?

incentives conducive to truth seeking? This is the subject of the issue of *Social Epistemology* for which Feigenbaum & Levy (1993) served as the jumping off point.

[5] In an age in which torture was routine state policy, the penalties inflicted upon the attacks on the monetary basis of the state were noticeable for their savagery. An attack on the sovereign's monetary authority was viewed in much the same light as an attack on the physical body of the sovereign (Kelly, 1981). The juxtaposition of functions of the United States Secret Service—protecting the President and combating counterfeit currency—is a surviving instance of such an identification.

In Chapter 4 of Book 1 of the *Wealth of Nations* Smith gives a social evolutionary account of the economic institution of money.[6] He explains how metallic commodities came to be used as money. This, however, created a set of problems. First, there is the matter of weight:

> The use of metals in this rude state was attended with two very considerable inconveniencies; first with the trouble of weighing; and, secondly, with that of assaying them. In the precious metals, where a small difference in the quantity makes a great difference in the value, even the business of weighing, with proper exactness, requires at least very accurate weights and scales. The weighing of gold in particular is an operation of some nicety. In the coarser metals, indeed, where a small error would be of little consequence, less accuracy would, no doubt, be necessary. Yet we should find it excessively troublesome, if every time a poor man had occasion either to buy or sell a farthing's worth of goods, he was obliged to weigh the farthing. (I. iv ¶7)

Then there is problem of assaying:

> The operation of assaying is still more difficult, still more tedious, and, unless a part of the metal is fairly melted in the crucible, with proper dissolvents, any conclusion that can be drawn from it, is extremely uncertain. Before the institution of coined money, however, unless they went through this tedious and difficult operation, people must always have been liable to the grossest frauds and impositions, and instead of a pound weight of pure silver, or pure copper, might receive in exchange for their goods, an adulterated composition of the coarsest and cheapest materials, which had, however, in their outward appearance, been made to resemble those metals. (I. iv ¶7)

For each problem, a set of solutions is offered:

> To prevent such abuses, to facilitate exchanges, and thereby to encourage all sorts of industry and commerce, it has been found necessary, in all countries that have made any considerable advances towards improvement, to affix a public stamp upon certain quantities of such particular metals, as were in those countries commonly made use of to purchase goods. Hence the origin of coined money, and of those public offices called mints; institutions exactly of the same nature with those of the aulnagers and stampmasters of woollen and linen cloth. All of them are equally meant to ascertain, by means of a public stamp, the quantity and uniform goodness of those different commodities when brought to market. (I. iv ¶7)

Smith then argues that history can be explained as following an evolutionary pathway:

> The first publick stamps of this kind that were affixed to the current metals, seem in many cases to have been intended to ascertain, what it was both most difficult and most important to ascertain, the goodness or fineness of the metal, and to have resembled the sterling mark which is at present affixed to plate and bars of silver, or the Spanish mark which is sometimes affixed to ingots of gold, and which being struck only upon one side of the piece, and not covering the whole surface, ascertains the fineness, but not the weight of the metal. (I. iv ¶8)
>
> The inconveniency and difficulty of weighing those metals with exactness gave occasion to the institution of coins, of which the stamp, covering entirely both sides of the piece and sometimes the edges too, was supposed to ascertain not only the fineness, but the weight of the metal. Such coins, therefore, were received by tale as at present, without the trouble of weighing. (I. iv ¶9)

[6] F. A. Hayek's defense of evolved institutions, which develops ideas in David Hume, suggests that all evolved conventions are equally useful. This claim, and the response to it, are studied in Peart & Levy (2006). Lewis's construction shares Hayek's Humean roots (1969, p. 3), but it does not make such a claim.

The passages we omitted above, and those which follow, suggest why it took Smith twenty years to complete the *Wealth of Nations*. He has surely forgotten more about the history of coinage than these two readers will ever know. When Smith describes the state policy of debasing coinage as a type of fraud, perhaps his readers recalled the proverbial question—who guards the guardians?

The problem of deceit is critical to what might be considered as Smith's public choice view of state policy. Needless to say, a policy of state-sponsored monopolies is the systematic target of the *Wealth of Nations*. Smith explains this policy is founded upon preventing deceit. This argument appears in the conclusion of Book 1 in which the interests of the different classes of society are contrasted. We start with the workers' employers:

> His employers constitute the third order, that of those who live by profit. It is the stock that is employed for the sake of profit, which puts into motion the greater part of the useful labour of every society. The plans and projects of the employers of stock regulate and direct all the most important operations of labour, and profit is the end proposed by all those plans and projects. But the rate of profit does not, like rent and wages, rise with the prosperity, and fall with the declension of the society. On the contrary, it is naturally low in rich, and high in poor countries, and it is always highest in the countries which are going fastest to ruin. The interest of this third order, therefore, has not the same connection with the general interest of the society as that of the other two. (I. xi ¶264)

Smith appeals to a learning by doing explanation for differential competence:

> Merchants and master manufacturers are, in this order, the two classes of people who commonly employ the largest capitals, and who by their wealth draw to themselves the greatest share of the public consideration. As during their whole lives they are engaged in plans and projects, they have frequently more acuteness of understanding than the greater part of country gentlemen. As their thoughts, however, are commonly exercised rather about the interest of their own particular branch of business, than about that of the society, their judgment, even when given with the greatest candour (which it has not been upon every occasion) is much more to be depended upon with regard to the former of those two objects, than with regard to the latter. Their superiority over the country gentleman is, not so much in their knowledge of the public interest, as in their having a better knowledge of their own interest than he has of his. (I. xi ¶264)

This competence has cash value:

> It is by this superior knowledge of their own interest that they have frequently imposed upon his generosity, and persuaded him to give up both his own interest and that of the public, from a very simple but honest conviction, that their interest, and not his, was the interest of the public. The interest of the dealers, however, in any particular branch of trade or manufactures, is always in some respects different from, and even opposite to, that of the public. To widen the market and to narrow the competition, is always the interest of the dealers. To widen the market may frequently be agreeable enough to the interest of the public; but to narrow the competition must always be against it, and can serve only to enable the dealers, by raising their profits above what they naturally would be, to levy, for their own benefit, an absurd tax upon the rest of their fellow-citizens. (I. xi ¶264)

All of this motivates Smith's advice to his readers. Lacking an institution that serves as the rhetorical equivalent of the public mint, each citizen must weigh and assay arguments made by policy makers, just as the quality of metals offered in exchange had been judged in barbarous times:[7]

> The proposal of any new law or regulation of commerce which comes from this order, ought always to be listened to with great precaution, and ought never to be adopted till after having been long and carefully examined, not only with the most scrupulous, but with the most suspicious attention. It comes from an order of men, whose interest is never exactly the same with that of the public, who have generally an interest to deceive and even to oppress the public, and who accordingly have, upon many occasions, both deceived and oppressed it. (I. xi ¶264)

The question to which we now turn is whether competition among deceivers is sufficient to solve Smith's problem of deceit in the arena of statistical reporting.

3 WHAT DOES THE ECONOMIST WANT?

To model a deceitful philosopher, we need to say what he wants.[8] We represent this issue in terms of our previous work on ethics and estimation (Levy & Peart, 2006). In Figure 1, we present competing preferences over estimates where we model the trade-off between bias and statistical efficiency. We depart from the textbook treatment of the goals of statistical research and allow bias in one direction to be a desired property of an estimate. A researcher may prefer to represent the world one way rather than another. The constraint we imagine follows the simple mechanics of specification search or data mining, where one makes many estimates and picks a favorite (Leamer 1983, Denton 1985). In particular, these constraints, the replication set, result from computing a number of unbiased estimates and mapping out the frontier combination of bias and efficiency (Feigenbaum & Levy, 1996).

We consider two sorts of preferences—one for a public-spirited statistician and one for someone with both public and private wants. The public-spirited statistician is interested only in statistical efficiency, a number without a sign. Either the statistician does not care about the value of the parameter to be estimated or, perhaps he does care, but he is unwilling to give up any amount of statistical efficiency to get a more pleasing estimate. In Figure 1, this possibility is described by indifference curve JJ. For such a statistician the rational estimate is j^*. When positive bias is a good, however, indifference curves take the shape marked by II. Thus the rational estimate, one in which some statistical efficiency is traded away for some gain in bias, is i^*.

The American legal system seems an ideal case to consider such rational choice estimation in a competitive context because the motivation for non-transparencies is all-too-obvious. In this context, the problem is that contending clients hire expert econometricians to press their case before a jury.

Structural equation estimation is a natural test ground for thinking about how the theorists' motivations are affected because the identifying restrictions flow from theoretical insight. It is perhaps not a coincidence that structural equation estimation is also fertile ground to study deceitful estimation because current conventions do not require the researcher to document the consequences of different selections of instrumental variables.

[7] This interpretation of Smith might save him from the wrath of George Stigler for having failed to apply the full-information self-interested model in political discussion (Stigler, 1971).

[8] This section is a largely a summary of the work reported in Levy & Peart (2006) in which we employ the motivational claim of a sympathetic statistician who is influenced by the wants of a client.

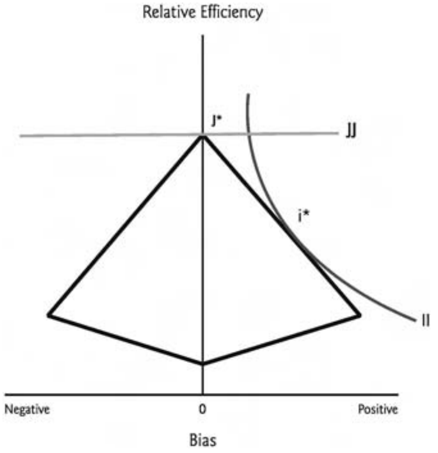


Figure 1: Competing rational estimates

This is the convention which we explore. The regression strategy need not be revealed. We need report only the equation system selected from the search.

Consider a demand and supply system (D & S) of the following structure:[9]

$$\text{Quantity } = \beta_1 + \beta_2 \text{ Price } + \beta_3 \text{ Income } + \eta \qquad (D)$$
$$\text{Price } = \alpha_1 + \alpha_2 \text{ Quantity } + \alpha_3 \text{ Cost } + \alpha_4 \text{ Weather } + \alpha_5 \text{ Politics } + \epsilon \qquad (S)$$

We suppose that the statistician has preferences over the estimated value of β_2. A researcher is required by convention to report only D, mentioning S casually. Thus, one can choose whether to include one, two or three exogenous variables from S. The rational choice estimate is the result of computing all possible combinations which identify a system and then picking. As above, we suppose the client and the sympathetic expert wants both bias and statistical efficiency. We measure the efficiency of estimator ι by the minimum mean square error [MSE*] of the estimates considered relative to the MSE of estimator ι; thus, MSE^*/MSE_ι.

A simulation is provided to give some idea of the ease with which biased estimates can be generated by such a selection procedure. There are several technical details. First, what is the distribution of the exogenous variables? If they are omitted not only do they change the error distribution but also the degree of over-identification, which changes dramatically the property of 2SLS estimates (Phillips, 1983). In the case considered, all exogenous variables are assumed to be a standard normal. Thus, omitting an exogenous variable in search of a pleasing outcome will not change the normality of the resulting errors.

[9] The alphas are all 1; β_1 is 10; β_2 is -1; β_3 is 3.

We consider two types of search. First, there is an unconstrained search for the maximum (minimum) value of the estimates of β_2. In the Tables below this is called "Max" and "Min." Second, there is a search which is constrained by the desire to have at least two exogenous variables in the supply curve. These are called "C-Max" and "C-Min." This will suggest how much the researcher might be willing to give up in efficiency to get bias. 100,000 experiments for N=25, 100, 400, 1600 are performed in *Shazam 8.0* (White, 1997).

All of the simultaneous estimates are replicable "two-stage least squares" estimates or "inefficient two-stage least squares" although only 2SLS and OLS are non-deceitful. The divergence between the "rational choice" estimate and the transparent 2SLS estimate can be thought of as transparency bias. Such bias persists through the case of N=1600.[10]

Table 1: Normal Exogenous Variables

100,000 Replications

	N=25		N=100		N=400		N=1600	
	Bias	Efficiency	Bias	Efficiency	Bias	Efficiency	Bias	Efficiency
OLS	0.40	0.35	0.40	0.08	0.40	0.02	0.40	0.02
2SLS	0.03	1.00	0.01	1.00	0.00	1.00	0.00	1.00
C-Min	-0.21	0.27	-0.09	0.48	-0.04	0.54	-0.02	0.54
C-Max	0.17	0.58	0.08	0.66	0.04	0.63	0.02	0.63
Min	-1.74	0.00	-0.22	0.14	-0.09	0.21	-0.04	0.21
Max	1.87	0.00	0.16	0.32	0.08	0.30	0.04	0.30

While the bias declines in absolute value as N increases, the reduction in bias from increasing N by a factor of four can be held in check by moving from the C-Max (C-Min) to Max (Min). This suggests that the problem of convergence will depend upon how the possible models increase as N increases. The simulation considered only exogenous variables which were truly included in the structure. We leave the problem of identifying the system by employing random numbers for future research. The problem of "pseudo-identification" raises theoretical questions that emerged at the dawn of simultaneous equation estimation and seem to have re-appeared in a new guise.[11]

The literature on the economics of expert witnesses has supposed that the jury decision will be made on the basis of an average of such biased estimates. This average is what the jury believes to be true. The conclusion of Froeb & Kobayashi (1996) for the case of biased experts before a jury, is that the average of their estimates will be unbiased.[12] And, it will be obvious

[10] Judging from 10,000 experiments the bias persists through N=6400. If the bias were measured in terms of the median of the estimates instead of the mean, it too would persist. The experiments were repeated with all exogenous variables following a uniform distribution between 0 and 1. Since it is not surprising that the amount of the bias is acutely sensitive to the distribution of the omitted exogenous variables, these results are not reported.

[11] We have benefitted from a conversation with Arthur Goldberger about the concerns of the Cowles Commission on pseudo-identification of structural equation estimates and with Adolf Buse on the modern discussion of weak-identification.

[12] In this, they are followed by Posner who contends that this property of a competitive procedure makes the idea of a court-appointed expert witness unwarranted: "The use of a court-appointed expert is problematic when (for example, in the damages phase of the case) the expert witness's bottom line is a number. For then, in the case of opposing witnesses, the trier of fact can 'split the difference,' after weighting each witness's estimate by its plausibility" (Posner, 1999, p. 1539).

from the tables above that, roughly speaking, the policy determined by the average of Min and Max or by the average of C-Min and C-Max will be unbiased.

However, this policy will have a higher variance than a policy determined by both using 2SLS. Thus, we create the familiar prisoner's dilemma in statistical context. While it is in the interest of each statistician considered separately to engage in selective under-reporting of results, it is in the interest of the statisticians considered as a group not to under-report. This is shown by the result that the diagonal element is roughly unbiased but the cell where both statisticians engage in "bias seeking" behavior has lower statistical efficiency than when they restrain themselves.

As an illustration of the point, a simulation of a quarter million replications was conducted to generate the statistician's dilemma using the case of normal exogenous variables with N=400. Here bias is computed in terms of deviation from the 2SLS estimate so as to represent the transparency bias. The efficiency is now the mean square error relative to the minimum where bias is measured in terms of deviation from the mean 2SLS estimate.

Table 2: Econometrician's Dilemma

Normal Exogenous Variables, N=400

250,000 Replications

	2SLS		C-Min		Min	
	Bias	Efficiency	Bias	Efficiency	Bias	Efficiency
2SLS	0.00	1.00	-0.02	0.81	-0.05	0.50
C-Max	0.02	0.88	0.00	0.97	-0.03	0.73
Max	0.04	0.64	0.02	0.92	-0.01	0.86

The optimistic conclusion of Froeb & Kobayashi (1996), followed by Posner (1999), depends upon their exclusive focus on the problem of bias. But if variance is also an issue, because one worries about the efficiency of the process, then their optimism about the unrestricted competitive process of expert witness seems more complicated than they suggest. A rule which constrains experts to report only 2SLS results would have a smaller variance than the competitive process modeled above.

4 Conclusion

Even under the idealized conditions described above, competition generates the obvious problem of a prisoners' dilemma. This results from a convention which, contrary to those modeled in Lewis (1969), forms the basis of conflict rather than co-ordination. The result suggests that it should be possible to propose a pareto superior convention. We offer one such, a computationally-intensive version of final-offer arbitration, in Levy & Peart (2006).

For the larger project at hand, game theory and semantics, we have presented a tiny model of an enormous problem. How does the ordinary person deal with advice, carried in language and reporting conventions, from motivated experts? Warts and all, competition provides one answer. Yet the harder problems emerge when the incentives of experts are so asymmetric that there is no viable competition at a level of statistical detail. Our study of the eugenic episode in statistics and economics (Peart & Levy, 2005) finds very little competitive opposition to this ghastly "scientific" development.

One promising approach to deal with the rational choice of statistical deceit comes out of biomedical research, in which clinical trials are quite literally matters of life and death (Berger et al., 2006). The authors suggest that experts, who are sympathetic to patients being victimized by the advice flowing from ill-designed clinical statistical procedures, might follow the thought experiment of John Rawls. So, medical experts would imagine themselves behind a veil of ignorance in which their private rational choice considerations are set aside.

> In the context of the research design, the "veil of ignorance" idea would require that researchers agree to construe as optimal only those design methods that all research would willingly assent antecedentially (i.e., before they had looked at a particular set of data.) (Berger et al., 2006)

Our suggestion of statistical arbitration might be one method that passes the deep test proposed by Rawls. If an expert will not pre-commit to a procedure, his clients might well have a good reason to ask why not.

REFERENCES

Barwise, J. and L. Moss (1996). *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena.* CSLI Publications, Stanford.

Berger, V. W., J. R. Matthews and E. N. Grosch (2006). On improving research methodology in medical studies. *National Cancer Institute Working Paper.*

Carnap, R. (1942). *Introduction to Semantics.* Harvard University Press, Cambridge, Mass.

Denton, F. (1985). Data mining as an industry. *Review of Economics and Statistics*, **57**, 124-127.

Dewald, W. G., J. G. Thursby and R. G. Anderson (1986). Replication in empirical economics: *The Journal of Money, Credit and Banking* Project. *American Economic Review*, **76**, 587-603.

Feigenbaum, S. and D. M. Levy (1993). The market for (ir)reproducible econometrics. *Social Epistemology*, **7**, 215-232.

Feigenbaum, S. and D. M. Levy (1996). The technological obsolescence of scientific fraud. *Rationality and Society*, **8**, 261-276.

Froeb, L. M. and B. H. Kobayashi (1996). Naive, biased, yet Bayesian: can juries interpret selectively produced evidence? *The Journal of Law, Economics and Organization*, **12**, 257-276.

Kelly, G. A. (1981). From Lèse-Majesté to Lèse-Nation: treason in eighteenth-century France. *Journal of the History of Ideas*, **42**, 269-286.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, **73**, 31-43.

Levy, D. M. and S. J. Peart (2006). Inducing greater transparency: towards the establishment of ethical rules for econometrics. *Eastern Economic Journal.* Forthcoming.

Lewis, D. (1969). *Convention: A Philosophical Study.* Harvard University Press, Cambridge, Mass.

Luschei, E. C. (1962). *The Logical Systems of Lesniewski*. North-Holland, Amsterdam.

Peart, S. J. and D. M. Levy (2005). *The "Vanity of the Philosopher": From Equality to Hierarchy in Post-Classical Economics*. University of Michigan, Ann Arbor.

Peart, S. J. and D. M. Levy (2006). Discussion, construction and evolution: Mill, Buchanan and Hayek on the constitutional order. *Allied Social Sciences Association*. Boston.

Phillips, P. C. B. (1983). Exact small sample theory in the simultaneous equations models. In: *Handbook of Econometrics* (Z. Griliches and M. D. Intriligator, eds.). North-Holland, Amsterdam.

Plato (1937). *The Republic* (P. Shorey, trans.). Loeb Classical Library, Cambridge, Mass.

Quine, W. van O. (1940). *Mathematical Logic* (Revised edition 1951). Harvard University Press, Cambridge, Mass.

Posner, R. A. (1999). An economic approach to legal evidence. *Stanford Law Review*, **51**, 1477-1546.

Rubinstein, A. (2000). *Economics and Language*. Cambridge University Press, Cambridge.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations* (E. Cannan, ed., 1904). Methuen and Co., London. http://www.econlib.org/library/Smith/smWN.html

Smith, A. (1978). *Lectures on Jurisprudence* (R. L. Meek, D. D. Raphael and P. G. Stein, eds.). Clarendon Press, Oxford.

Stigler, G. J. (1971). Smith's travels on the ship of state. *The Economist as Preacher, and Other Essays* (G. J. Stigler, 1982). University of Chicago Press, Chicago.

White, K. J. (1997). *Shazam: Econometrics Computer Program (8.0)*. McGraw-Hill, New York.

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 4

FROM SIGNALS TO SYMBOLS: GROUNDING LANGUAGE ORIGINS IN COMMUNICATION GAMES

Ángel Alonso-Cortés
Universidad Complutense de Madrid

This chapter brings together the fields of economics and linguistics on the topic of the origins of language. It concerns properties of human language, particularly how linguistic signs or symbols have inherited design features present in linguistic communication. I will show how some features of language can be adequately understood as a result of coordination games. I will argue that modern language originated as a consequence of trade relationships and the division of labour involved by early humans around 40,000 years ago. As an economic activity, both trade (or exchange) relationships and the division of labour call for coordination. The outcome is that games and economic behaviour have a significant causal relationship to general properties of the linguistic symbol.

1 ADAM SMITH'S DOG

Language and economics have been related since at least Adam Smith's reflections on the origin of the division of labour. Smith attributes the division of labour to language and to the faculty of reason. In his *Wealth of Nations* of 1776 Smith writes:

> The division of labour, from which so many advantages are derived, is not originally the effect of any human wisdom.... [I]t is the necessary consequence of a certain propensity in human nature: the propensity to truck, barter, and exchange one thing for another...This propensity...seems to be the necessary consequence of the faculties of reason and speech. (Smith, 1776, p. 25)

Smith goes on to assert that this propensity is unique to man, thus writing these famous words:

> No body ever saw a dog make a fair and deliberate exchange of one bone for another with another dog. (Smith, 1776, p. 26)

According to Smith, the division of labour, the exchange of goods and language could all be causally related. The division of labour produces a diversity of goods that could be exchanged.

The exchange of goods creates the necessity of a contract, and contracts require concerted or coordinated actions among the contracting individuals. In coordinated actions, agents are involved in communication games, whereby they convey the information required for the exchange. A symbolic and complex language then subserves the communication of information.

Modern linguistics has also adopted a view that appeals to economics and even political theory and political philosophy. Let me mention Ferdinand de Saussure, who in his *Cours de Linguistique Générale* of 1916 asserted that the linguistic signs, or "la langue", had originated in a social contract: "There is a language", Saussure states,[1] "only in virtue of a kind of social contract handed on among members of a community." Moreover, the Swiss linguist was the first to establish that language was comprised of interrelated signs that form a system. The Saussurean sign is a one-to-one mapping from meaning to sound that is lodged in the brains of at least two speakers. All individuals bound by language, Saussure avows, reproduce the same sounds[2] mapped onto the same concepts. The origin of this social crystallization, he goes on to explain, lies in the fact that the meaning-sound mapping is the same for all the individuals sharing a language, because there is a coordination faculty that makes such coordination possible.

Some years later, in 1933, the American linguist Leonard Bloomfield, in his *Language*, a work resting on Saussure, emphasised that language is a coordination problem between sound and meaning, and that this coordination "makes it possible for man to interact with great precision" (Bloomfield, 1933, §2.2.). He bolstered Smith's speculation on the relatedness of language to the division of labour, when he asserted that language always accompanies every human action. Bloomfield argues that:

> In the ideal case, within a group of people who speak to each other, each person has as its disposal the strength and skill of every person in the group. The more these persons differ as to special skills, the wider a range of power does each one person control. The division of labor, and with it, the whole working of human society, is due to language. (Bloomfield, 1933, §2.3, p. 24)

Bloomfield's approach to the function of language calls to mind Smith's speculation on language and the division of labour. As economist Karl Wärneryd remarked, there is no logical reason to expect that language is what makes possible the exchange.[3] For one thing, the division of labour—although not as in humans—occurs in animals without a complex language such as ants, wasps, bees and wolf packs.[4] Specialisation in social insects is so surprising that Dawkins (1989, p. 180) asserts that these insects discovered—before man!—that cultivation of food is more efficient than hunting-gathering.[5] Therefore, it is difficult to attribute to the faculty

[1] F. de Saussure, *Cours*, Intro. III, §2: "[La langue] n'existe qu'en vertu d'un sorte de contrat passé entre les membres de la communauté."

[2] Strictly speaking, it is a mental representation of the articulated sounds what is mapped into a concept or meaning. Both sound and meaning have a mental reality.

[3] Wärneryd (1995) tackles the relationship between exchange and language in a different but insightful way.

[4] Smith's omission of the social features of insects was noticed by Houthakker (1956). Recently, zoologist L. David Mech has added more evidence on the division of labour in wolf packs: "The typical wolf pack, then, should be viewed as a family with adult parents guiding the activities of the group and sharing group leadership in a division-of-labor system in which the female predominates primarily in such activities as pup care and defense and the male primarily during foraging and food provisioning and travels associated with them" (Mech, 1999).

[5] Slavery, warfare, and robbery can be found among social insects as well as in humans. See Hamilton (1995, p. 216).

of language the main motivation that led to the division of labour.[6]

As the division of labour may occur without language, it would behove us to look back to trade, or to the deliberate exchange of goods as a reasonable hypothesis to explain how language originated and acquired its properties.

In a recent paper, Horan et al. (2005) developed a mathematical model ("Shogren's model") to explain why Neanderthal man went extinct while coexisting with *Homo sapiens*. The title of their paper is fairly suggestive to my own present purpose: "How trade[7] saved humanity from biological exclusion..." They explore two hypotheses: biological exclusion and behavioural exclusion.

Biological exclusion predicts that the Neanderthal extinction would have been slower than it actually was. Also, if Neanderthals were biologically more efficient, Shogren's model predicts, contrary to fact, that humans would not have coexisted with Neanderthals.

The reason why humans survived, although they were biologically inferior to Neanderthals, is better explained by the behavioural exclusion theory. Behavioural exclusion theory proposes that humans survived due to the division of labour and specialisation, which Neanderthals lacked. The most plausible scenario envisaged by Shogren's model is one in which there is a complete division of labour within two groups of humans: skilled hunters that harvested meat and unskilled hunters who produced other goods. Incidentally, these two groups of humans were already envisioned by Smith in the *Wealth of Nations*.[8]

Even with a modicum of trade in Neanderthals, humans overcame them. Their model proves that humans survived because of the availability of meat consumption was greater due to the division of labour. Horan et al. (2005, p. 21) conclude that "A crucial issue remains unresolved: it is an open question why the early humans first realized the competitive edge from trade. Some attribute the edge to differences in cognition or language abilities or both, but the jury is still out."

The issue may be elucidated by looking into Neanderthal language. As there is no evidence that Neanderthals had a complex language unlike there is of early humans,[9] the hypothesis that the competitive edge could be realised by developing abstract symbols becomes compelling. The conclusion that language and trade co-existed seems inescapable. It seems reasonable that all cognitive capacities involved in trade (such as the designing of tools for manufacturing exchangeable goods, the exchange value of goods, and the ability to make decisions on goods) should be observable in language.

The next step involves determining which came first, language or trade? Although no definite answer can be given, some logical priority goes to trade. Three arguments may be adduced. First,

[6] Also, L. von Mises asserted that the division of labour makes man distinct from animals: "It is the division of labor that has made feeble man, far inferior to most animals in physical strength, the lord of earth and the creator of the marvels of technology" (von Mises, 2005, p. 18). Notwithstanding the core role of the division of labour, neoclassical and modern economists have observed that Smith's theory would lead to an organisation of the market dominated by increasing returns, which is not borne out; see Buchanan (1999).

[7] Trade means in Shogren's model "exchange", be it voluntary or involuntary (centralised or dictatorial).

[8] "In a tribe of hunters or shepherds a particular person makes bows and arrows, for example, with more readiness and dexterity than any other. He frequently exchanges them for cattle or for venison with his companions; and he finds at last that he can in this manner get more cattle and venison, than if he himself went to the field to catch them" (Smith 1776, I.ii.3).

[9] There has been a hot debate on the issue of Neanderthal language, which was settled by Lieberman (1984) and Mithen (2006, p. 221), who both argue that Neanderthals at most had an inferior linguistic capacity than *Homo sapiens*. It should be emphasised that no real evidence for a Neanderthal language has been offered.

language is a necessary condition neither for the division of labour nor for trade. In the Shogren-Smith model, it is meat consumption and a previous division among members of the tribe (skilled versus unskilled individuals) that triggered the division of labour. According to Shogren, the assumption that early humans were more skilled hunters than Neanderthals, allowed them to produce meat enough to exchange for goods produced by unskilled hunters.

Second, as language basically involves coordination problems just as trade and the division of labour do, it is plausible to assume that language depends on trade and the division of labour as well as on the more complex social relations added by trade. The ground for this dependency lies in the fact that the division of labour leads to coordination between (at least) two individuals thus incurring external coordination costs (Houthakker, 1956). Then language could have evolved in order to set off such costs.

The third argument is that some games can be played (or pre-played) using communication and cheap talk, which does not add more or less value to payoffs.

So trade may occur without language, but language must be motivated in the sense that a speaker S sends a message μ to a receiver R with a particular intention.

The scenario set up by trading can boost a symbolic communication system as rich as modern human language. Wärneryd (1995) addresses the role of language in economic activities reminding that neoclassical economists start from the premise that exchange follows from well-defined preferences of individuals with a basket of consumption goods. When preferences (or payoffs) are in equilibrium, however, it may occur that some equilibria are more efficient and stable than others. Communication selects the more efficient equilibrium if it is costless. Exchange, then, triggers or motivates language, not the other way around. Consequently, if animals do not have full symbolic communication it is because they do not exchange goods, which in turn motivates the existence of a language.[10] Smith's dog has not evolved language because that would require exchange and coordination. As he has nothing to coordinate, he needs no language. The dog is tied to its costly signals.

I will conclude, then, that trade is a robust candidate for the origins of a modern symbolic language.

2 Games and Symbols

Next, I will take up a subset of Hockett's design features and show how they fit into the coordination game framework. We should bear in mind the main difference between traditional game theory and coordination game theory: the former deals with winning strategies, solution concepts and equilibria, and the latter with players' common-interest strategies and multiple equilibria. Consequently, players in common-interest games use cognitive strategies such as imitation, analogy, reasoning, guessing, imagination and common knowledge.

Design features are properties that characterise language as a communication system to compare language and signals of other nonhuman communication systems. For the moment, I ignore animal signals and focus on linguistic symbols originated in the coordination game of trade and the division of labour.

I deal with the following design features proposed by Hockett (1960):[11] duality, semanticity,

[10] Hamilton (1995, p. 342) makes a case for the idea that tools and language confer benefits to a cooperative hunter.

[11] Some of these features were previously studied by Saussure, Bloomfield and Martinet, but are usually known

parity, specialisation, prevarication, cultural transmission, and displacement of reference.

Let us look at each of these features to see how they might be construed as games.

2.1 Duality

Def.: In Saussure (*Cours*, I.1. §1), duality is defined as follows: "The linguistic sign [i.e., symbol] is a mental entity with two faces: a concept [meaning] and an acoustic image [sound]. These two elements are tightly joined and one demands the other [bidirectional mapping]." Idem I.1. §2: "The tie [the mapping] joining meaning and sound is arbitrary."

The first conundrum that the Saussurean sign poses is a coordination problem. In order to communicate, the agents or the speakers of a community must make the same associations between sound and meaning. Such coordination is solved by means of a coordination game between meaning and sound.[12] It must be noticed that Saussure (*Cours*, Intro III. §2) put forth that speakers in a population P must be endowed with "receptive and coordinating faculties" to attain the same one-to-one mapping. Therefore, meaning and sound must be coordinated in a communicating population P of senders and receivers because both meaning and sound are unattached to each other. Meaning of a sign S_1 could, a priori, be attached to any other string of sounds σ_n and vice versa. This coordination problem can thus be formulated in the following way: how do the sender and the receiver of a message assign the same bidirectional mapping from meaning into sound and from sound into meaning?

As all members of the population P want to use the same signs to communicate, they all share a common interest and therefore must coordinate their choice. This is, in essence, a coordination game in the sense of Schelling (1980, pp. 83–118). More specifically, he characterises a coordination game as follows:

1. Players' preferences are identical, and so there is no conflict of interest.
2. Each player's best choice depends on the action he expects the other to take, which in turn depends on the other's expectations of his own. In other words, the game is based upon the players' mutual expectations.
3. The players' goal is to share some common-interest activity by means of a cognitive process (such as imagination, poetry and humour). In the case of language, players want to use the same signs to communicate with each other.

Let us look at Table 1. One player chooses a Row and other player chooses a Column. Row and Column represent tacit processes determining the payoffs. Since unlike in zero-sum games the players' goal is not to win but to share some common interest by tacitly searching through cognitive processes, payoffs only represent the degree of coordination attained by the players.[13] So the payoff matrix for a coordination game is different from zero-sum games and nonzero-sum games. If players combine $\langle R_1, C_1\rangle$ they are better off than combining $\langle R_1, C_2\rangle$ and better

as Hockett's design features.

[12] Wittgenstein's language games may be construed as coordination games. See Wittgenstein (1953, §§2, 8, 21, 48–51).

[13] Such processes may equal the usual strategies in conflict games, but contrary to conflict games, no minimax solution exists.

	C_1	C_2	C_3	C_4	C_5
R_1	1 1	0 0	0 0	0 0	0 0
R_2	0 0	1 1	0 0	0 0	0 0
R_3	0 0	0 0	1 1	0 0	0 0
R_4	0 0	0 0	0 0	1 1	0 0
R_5	0 0	0 0	0 0	0 0	1 1

Table 1: Lower left entry in cells is payoff to the row-player, upper right to the column-player

off than combining $\langle R_2, C_1 \rangle$ and so on. As it is possible that, whenever choosing one Row and choosing one Column players 'win', that is, they guess what each other is thinking, the winning results $\langle 1, 1 \rangle$ can be arranged in a diagonal line.

Let us get back to the design features. Duality is conceived in Saussure's sense as a bidirectional mapping from sound and meaning such that both sound and meaning are autonomous of each other but must be coordinated by the senders and the receivers in order for them to attain optimal communication. What cognitive strategies are involved in such a duality? Some tacit strategies that come to mind are random mapping, imitation, probabilistic mapping, and knowledge of convention in the sense of David Lewis (1969).

Linguistic conventions are not explicit but tacit agreements. This means that speakers must use cognitive strategies to coordinate sound and meaning. Convention can be arrived at by calling on a variety of such strategies. Saussure assumed the existence of a coordinative capacity in humans. This assumption, however, sets up a circular argument. A much more adequate explanation is provided by Lewis' convention.

2.2 Semanticity

Def.: "The elements of a communicative system [linguistic symbols] have associative ties with things and situations, or types of things and situations, in the environment of its users... such ties are semantic conventions shared by speakers." (Hockett, 1960, p. 41)

The bidirectional mapping sound-meaning should be distinguished from the mapping symbol-denotation (things, situations, or simply, actions). Adopting a Lewisian theory of meaning, symbols (or signals in the sense of game theory) are mapped into actions so that actions can be true or false if they establish a coordinating equilibrium. Table 2 shows such an equilibrium. Signal A means (is mapped onto) action X, with payoff (1, 1), while signal B means action Y with payoff (1, 1). Mapping is established by convention.

			Receiver Action	
			X	Y
	Signal	A	1,1	0,0
Sender				
	Type	B	0,0	1,1

Table 2: A Coordinating Equilibrium

2.3 Interchangeability (parity)

Def.: "Adult members of any speech community are interchangeably transmitters and receivers of linguistic signals." (Hockett, 1960, p. 139)

This feature derives from the definition of a coordination game without proof, as this game is played by pairs of speakers. Yet parity has been challenged by rationalist philosophers and linguists. Rationalists claim that language is used only for the expression and representation of thought, not for social communication. However, game-theoretic approach to language requires that language be strictly used and motivated for the communication of intentions. Besides, this should be taken not only as its current function, but as the original function.[14] Since the communicative function overrides the representational function in the efficiency of coordination, the claim that language is for the expression of thought is not motivated by game theory. Communication, not the expression of thinking, subserves coordination.

2.4 Specialisation

Def.: "A communicative act, or a whole communicative system, is specialized to the extent that its direct energetic consequences are biologically irrelevant. Obviously language is a specialized communicative system."

Contrary to human language, animal signals have direct biological consequences as well as energetic costs. In insects, signals (calls and songs) emitted by a male insect serve to attract females as sexual mates.[15] The bees' dance informs only about the food source.[16] Also birds' alarm signals alert other conspecifics to flee. The bird that warns its conspecifics by emitting an alarm call is in grave danger of dying because it attracts the predator's attention. This shows

[14] Assuming that communication is both the original and the current function of language avoids the issue (for which Darwinism lacks an adequate response) of how an original organ transforms its original function into another function, contrary to Chomsky who asserts that we do not know the original purpose of language, although he assumes a transformation of the original function into the "expression of thought" function; see Kirschner & Gerhart (2005) and Hauser et al. (2002).

[15] Gerhardt & Huber (2002) observed that some insects lose weight during call transmission.

[16] For these and other examples of animal calls, see Dawkins (1989).

that communicative behaviour in animals adopts strategies that incur costs and benefits just as in the conflict-of-interest game. Dawkins points out that "the belief that animal communication signals originally evolve to foster mutual benefits, is too simple." Rather, he continues, "all animal interactions involve at least some conflict of interest" (Dawkins, 1989, pp. 68–87). Since linguistic communication is basically a coordination game, it is costless or cheap; costs and benefits of sending and receiving signals are irrelevant. Language, then, may be conceived as a signalling game in which both the sender and the receiver obtain equal payoffs because they share the same interests.[17] Moreover, animal signals may be dishonest, while language lacks dishonest signals. Language evolved for coordination, which sets a barrier for a strict Darwinian view on language origins.[18]

2.5 Prevarication

Def.: "Linguistic messages can be false, and they can be meaningless in the logician's sense." (Hockett, 1960, p. 14)

One of the main differences between animal signals and linguistic symbols is in animal's signal communication being truthful while communication by linguistic symbols possibly not. Signals correspond to a set of fixed states either of the animal type (hunger, sex) or the environmental type (danger). Therefore, prevarication or lying is not a real option for animals.[19] However, the possibility of the receiver being manipulated by the sender has been emphasised as an option in animal communication (Dawkins, 1989, p. 64). On the other hand, linguistic communication assumes truthful messages sent by truthful senders. This is called the "truth bias" by game theorists. The speaker, in turn, is committed to the truth of his messages.

The nature of lying is due to the symbolic character of human communication comprising conventionality and unboundedness. Biologically, lying is a cost for a symbolic system because it contributes to the selfish and parasitical but non-coordinating behaviour (Hamilton, 1995, p. 332).

Game theory recognises both the existence of lying and "The Decay of Lying", as Oscar Wilde put it in his comedy.[20] Lying is a behaviour that fits into a two-person partial-interest game, that is, a game in which some agent is coordinating non-strictly. Table 3 represents such a game, in which the sender sends a signal which triggers the best action by the receiver.

This matrix enlists values of common interests as well as of conflict of interests. The combination $\langle A, Z \rangle = (6, 3)$ and the combination $\langle B, Z \rangle$ represent cases in which the sender has obtained more profit than the receiver.[21]

Note, however, that lying violates linguistic conventions, but these conventions cannot be associated with lying because there would then be a winning strategy for the receivers such as "If the sender lies—using a lying strategy—do not act as the sender expects." Thus a better winning

[17] Otherwise said, the utility function of Sender $u(s)$ and Receiver $u(a)$ are equal.

[18] Because linguistic communication is a pure coordination, mass phenomenon (individuals being genetically unrelated), it presents a real conundrum for natural-selection accounts of language origins and evolution concerning individuals and genes. For instance, Pinker (1995) does not take up these issues.

[19] Hamilton (1995, p. 218), who writes on "Selection of selfish and altruistic behavior in some extreme models", remarks that "by our lofty standards, animals are poor liers." In turn, Karl Popper (1974, pp. 1112–3) suggested that "human language evolved because it made lying possible."

[20] Wilde's words wittily express the non-predominance of lying: "With the possible exceptions of barristers, lying as an art has decayed." On lying as a game, see Wittgenstein (1953, §249).

[21] Experimental work by Kawagoe & Takizawa (2005) shows that lying pays as well as the truth bias of agents.

				Receiver Action	
			X	Y	Z
	Signal	A	4,4	1,1	6,3
Sender					
	Type	B	1,1	4,4	6,3

Table 3: Partial interest game (lying)

strategy would eventually evolve. Therefore, one can deduce that lying cannot be evolutionary stable (Dawkins, 1989, p. 77). This evolutionary game explains why there are no markers (conventions) for lying in human languages.

2.6 Cultural transmission

Def.: "The continuity of language from generation to generation is provided by tradition. All traditional behaviour is learned [from others]. Tradition becomes transformed into cultural transmission when the passing down of traditional habits is mediated by symbols." (Hockett, 1960, p. 155)

Symbols are learned from generations to generations, and they constitute grammatical patterns. Linguists and psychologists argue whether an innate, not culturally but genetically transmitted device exists that makes the learning of grammar possible. Supporters of an innate device assume the existence of an absolute invariant Universal Grammar (UG) genetically transmitted that would explain language learning with no resort to cultural transmission.[22] UG is conceived as something like a random generator or an automaton.

The UG hypothesis, however, has not found observable or empirical universals that would account for overt and regular crosslinguistic variation.[23]

A different way to tackle regular variation (sometimes termed Greenberg universals) is to look at it as a coordination game problem in Schelling's sense. Language learning requires the input from the learner's community. All learners must converge on the input grammar, that is, they must coordinate their grammars with those of the input. When coordination problems persist among the members of a community, that community yields regular patterns to solve such problems or otherwise adopt them from other communities (for example, by cultural diffusion). These regular patterns turn into common knowledge within the community.[24] Note also that in a

[22] Note, however, that is false in a strict (neo)darwinian view.

[23] Universals of the kind required by the supporters of the random generator grammar are located at the neurobiological level ignoring overt linguistic properties and offering no general account of crosslinguistic variation. Apart from the automaton, such universals are missing at present.

[24] Lewis (1969) adopted this view which can be extended to the realms of language learning and language evolution.

coordination game an agent selects an action in an undetermined way within a bounded set. Thus, using a bounded number of actions, we expect different conventions for different communities.

In fact, some computational models of language evolution suggest that overt empirical universals do arise out of multiple agents that evolve across generations (Kirby & Hurford, 2001). Here linguistic generalisations (grammatical rules) emerge from cultural transmission, making the assumption of innate Universal Grammar unnecessary.

2.7 DISPLACEMENT OF REFERENCE

Def.: Bloomfield (1933, §9.3):

> If we had perfect definitions [of words], we should still discover that during many utterances the speaker was not at all in the situation which we had defined. People very often utter a word like *apple* when no apple at all is present. We may call this *displaced speech*. The frequency and importance of displaced speech is obvious. Relayed speech embodies a very important use of language: speaker A sees some apples and mentions them to speaker B, who has not seen them; speaker B relays this news to C, C to D, D to E and so on, and it may be that none of these persons has seen them, when finally speaker goes and eats some. In other ways, too, we utter linguistic forms when the typical stimulus is absent. (Bloomfield, 1933, p. 141)

The displacement of reference has been taken to be a key property of language. Chomsky (1966) highlights displacement under the "absence from stimulus" argument, which he uses against the behavioural account of language use. Displacement (or absence of stimulus) can be derived from (i) semanticity and (ii) specialisation. As we have seen, semanticity is the result of conventions under a coordinating equilibrium, while specialisation yields costless communication (cheap talk). Semanticity provides for conventional and arbitrary symbols that can be stored in memory, while specialisation makes cheap the use of symbols so that agent A can relay information (at no cost) to agent B, agent B to agent C and so on, so that the whole population of agents can exchange information not perceived at the moment of the utterance.

The fact that symbols can be relayed accounts for one crucial property of language: sentence recursion. If agent A relays to B "John ran away", B can relay this information to agent C as embedded into another symbol: B says: "John ran away", and C relays to D: B says "A says 'John ran away'", and so on. Recursion, then, is a property that emerges from displaced reference and is not imposed by a Universal Grammar. The case in which knowledge of an event is acquired from hearsay (i.e., displaced from the speaker) sets up the evidential modality. Some languages morphologically mark events known from evidence acquired in this way. Thus Tunica, Bulgarian, and Kwakiutl—among a wide set of languages—use evidentiality markers to signal that the speaker knows the information from others. Other linguistic processes are direct consequences of displaced reference such as indirect questions, quoted speech, discourse representation or free indirect discourse. Moreover, displacement adds a significant edge to the population of agents using referential symbols: it spares time invested in searching for information that otherwise an agent needs to obtain in the presence of stimulus. The spared time can be invested in other activities increasing the number of activities that the population can engage in. Displacement increases the production possibility curve.[25]

[25] I deal with time allocation related to displacement in Alonso-Cortés & Cabrillo (2006).

3 CONCLUSIONS

Language and trade are related to each other because both involve the kinds of exchange problems that may be solved by coordination games. Modern symbolic language might have been boosted as a tool to set off external coordination costs incurred by trade (goods exchange) in modern human populations. From coordination games and the cost-benefit analysis one can derive a subset of design features of language. Some significant and crucial features such as duality, semanticity, displacement of reference and prevarication are a direct consequence of coordination among members of a population, while coordination through evolutionary games accounts for cultural transmission.

ACKNOWLEDGMENTS

Paper presented at the 6th Winter Workshop on Economics and Philosophy, "Economics and Language", Madrid, Spain, 2006, at Universidad Nacional de Educación a Distancia. Sponsored by Fundación Urrutia-Elejalde, Ministerio de Educación y Ciencia, Project HUM2005-25447-E/FISO. I thank Paco Cabrillo (UComplutense) and Pedro Schwarz (USan Pablo-CEU) for their valuable comments to an earlier draft of this paper. Matilde Vivancos has contributed generously to the final version.

REFERENCES

Alonso-Cortés, Á. and F. Cabrillo (2006). *The Economics of Language: A Coordination Game Approach*. Manuscript, UCM, Madrid.

Bloomfield, L. (1933). *Language*. Chicago University Press, Chicago.

Buchanan, J. M. (1999). Generalized increasing returns, Euler's theorem, and competitive equilibrium. *History of Political Economy*, **31**, 511-523.

Chomsky, N. (1966). *Cartesian Linguistics*. New York: Harper.

Dawkins, R. (1989). *The Selfish Gene*. Oxford University Press, Oxford.

Gerhardt, H. C. and F. Huber (2002). *Acoustic Communication in Insects and Anurans*. University of Chicago Press, Chicago.

Hamilton, W. D. (1995). Selection of selfish and altruistic behavior in some extreme models. In: *Narrow Roads of Gene Land* (William D. Hamilton). W. H. Freeman, Oxford.

Hamilton, W. D. (1995). Innate social aptitudes of man.... In: *Narrow Roads of Gene Land* (W. D. Hamilton). W. H. Freeman, Oxford.

Hockett, C. F. (1960). Logical considerations in the study of animal communication. In *Animal Sounds and Communication* (W. F. Lanyon and W. N. Tavolga, eds.), pp. 392-430. American Institute of Biological Sciences, Symposium Series 7, Washington, D.C.

Hockett, C. F. (1963). The problem of universals of language. In: *Universals of Language* (J. Greenberg, ed.). MIT Press, Cambridge, Mass.

Horan, R. D. E. Bulte and J. S. Shogren (2005). How trade saved humanity from biological exclusion: An economic theory of Neanderthal extinction. *Journal of Economic Behavior and Organization*, **58**, 1-29.

Hauser, M., N. Chomsky and T. Finch (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, **298**, 1569-1579.

Houthakker, H. (1956). Economics and biology: Specialization and speciation. *Kyklos*, **9-2**, 181-189.

Kawagoe, T. and H. Takizawa (2005). Why lying pays: Truth bias in the communication with conflicting interests. Manuscript, Tokyo. http://ssm.com/abstract=696141

Kirby, S. and J. Hurford (2001). The emergence of linguistic structure: An iterated learning model. In: *Simulating the Evolution of Language* (A. Cangelosi and D. Parisi, eds.), pp. 121-148. Springer, London.

Kirschner, M. W. and J. C. Gerhart (2005). *The Possibility of Life.* Yale University Press, New Haven.

Lewis, D. (1969). *Convention.* Harvard University Press, Cambridge, Mass.

Lieberman, P. (1984). *The Biology and Evolution of Language.* Harvard University Press, Cambridge, Mass.

Mech, L. D. (1999). Alpha status, dominance and division of labor in wolf packs. *Canadian Journal of Zoology*, **77**, 1196-1203.

von Mises, L. (2005). *Liberalism: The Classical Tradition.* Liberty Fund, Indianapolis.

Mithen, S. (2006). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body.* Harvard University Press, Cambridge, Mass.

Pinker, S. (1995). *The Language Instinct.* Harper, New York.

Popper, K. (1974). Reply to my critics. In: *The Philosophy of Karl Popper* (P. A. Schilpp, ed.). Open Court, La Salle.

de Saussure, F. (1916/1967). *Cours de Linquistique Génèrale.* Payot, Paris.

Schelling, T. C. (1980). *The Strategy of Conflict.* Harvard University Press, Cambridge, Mass.

Smith, A. (1776/1981). *An Inquiry into the Nature and Causes of the Wealth of Nations.* Liberty Fund, Indianapolis.

Wärneryd, K. (1995). Language, evolution and the theory of games. In: *Cooperation and Conflict in General Evolutionary Processes* (J. L. Casti and A. Karlqvist, eds.). pp. 405-421. John Wiley, New York.

Wittgenstein, L. (1953). *Philosophical Investigations.* Blackwell, Oxford.

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 5

EVOLUTIONARY GAMES AND SOCIAL CONVENTIONS

Pelle Guldborg Hansen
Roskilde University

1 INTRODUCTION

Some thirty years ago Lewis published his *Convention: A Philosophical Study* (Lewis, 1969). This laid the foundation for a game-theoretic approach to social conventions, but became more famously known for its seminal analysis of common knowledge; the concept receiving its canonical analysis in Aumann (1976) and which, together with the assumptions of perfect rationality, came to be defining of classical game theory.

However, classical game theory is currently undergoing severe crisis as a tool for exploring social phenomena; a crisis emerging from the problem of equilibrium selection around which any theory of convention must revolve. In response, the so-called *evolutionary turn* has developed. While retaining the broad framework, in which games are described in terms of strategies and payoffs, this marks a transition from the classical assumptions of perfect rationality and common knowledge to assumptions characterising agents as conditioned for playing certain strategies upon the population of which evolutionary processes operate.

By providing accounts of equilibrium selection and stability properties of behaviours, the resulting frameworks have been brought to work as well-defined metaphors of individual learning and social imitation processes, from which a revised theory of convention may be erected (see Sugden 1986, Binmore 1994 and Young 1998).

This paper makes a general argument in support of the evolutionary turn in the theory of convention by a progressive exposition of its successful application to a variety of simple, but paradigmatic games. In doing this, it examines and qualifies on what may be said within this framework about the relations between social conventions on the one hand, and phenomena such as Pareto-efficiency, risk, discrimination, self-interest and cooperation on the other. For most of the arguments, the formalisation will be kept at a minimum as well as restricted to models based on two-player interactions.

2 Social Conventions

It has long been recognised within the social sciences that a certain type of behavioural patterns making up the anatomies of social systems may be given informal descriptions as *social conventions*, in the sense that (i) they depend on the interdependency of individual actions, and (ii) comparative analysis may reveal them to be *contingent* relative to some attributed functional description. For instance, the regularity of keeping to the left in the UK is conformed to by drivers as they expect from past interactions that other drivers keep to the left as well. Yet on the functional description of avoiding collisions, the purpose of such conformity could also be served by everyone keeping to the right, as is done, for example, in the US.

It is just as well recognised, though, that conventions often enjoy complex natures as well as an intimate relationship with social norms and institutions. This is especially true when these incorporate aspects of conflict, risk or discrimination. Thus, apart from the two common features mentioned, it has to be granted that social conventions differ from each other in many respects. Some are institutionally engineered, carefully codified and severely enforced, while others are so fundamental that thinking about them as a product of man is intriguingly difficult. Some enable the utilisation of potentials for cooperation, yet sometimes cease to exist, while others, though socially ineffective, prove utterly hard to dissolve. Consequently, though conventions provide the structure within which social life is led and institutions operate, a unified approach has proven difficult to develop. For this reason, a constant issue of the philosophy of social science has been that of explaining their nature and dynamics. That is, to explain what causes conventions *to*, and how conventions *do*, emerge, stabilise and in some cases change or deteriorate.

3 Lewis' Theory of Convention

Perhaps the largest obstacle in answering these questions has been the lack of a rigorous framework for systematically exploring conceptual hypotheses and their implications. Since Lewis' *Convention*, however, game theory has been thought by many to provide one particularly interesting candidate as a framework of thought.

Convention begins by extrapolating from a series of paradigm conventions a shared function of being coordination devises in situations presenting some recurrent *coordination problem* to the agents involved. It then utilises classical game theory to model the most simple strategic structure of this type of problem as a pure-coordination game like that of Matrix 1.

		Player 2	
		a	b
Player 1	a	**1,1**	0,0
	b	0,0	**1,1**

Matrix 1: A pure-coordination game

For the purpose of exposition, this two-player game may be interpreted as representing any instance of a recurrent situation G presenting a coordination problem in a population P, where $P \geqq 2$. This game, then, is played by agents i and j, where $i, j \in P$. Further, i and j are randomly drawn from P and assigned at random to either the role of row player, Player 1, or column player,

Player 2. Due to the symmetry of the game, i and j then face an identical strategy set $S = \{a, b\}$ no matter which player position they are assigned to.

In classical game theory, each agent is assumed to choose rationally from S a pure strategy s, or a mixed strategy x assigning some probability mix to the elements of S, with the aim of maximising his expected payoffs. As expected payoffs are dependent on combinations of strategies, referred to as *strategy profiles*, each agent is assumed to reason rationally about his choice of strategy under the assumption that other agents are rational as well and that this is common knowledge in P in order to pursue maximisation. In the matrix like that above, Player 1's payoffs are given first, then the payoffs of Player 2 for each strategy profile. In classical game theory, an acceptable solution to such a game is required, at least, to be a Nash equilibrium; that is, a strategy profile in which each agent has done as well as he can given the actions of the other agents. Thus in the game of Matrix 1 there are two pure-strategy Nash equilibria (marked in bold) and one mixed strategy Nash equilibrium in which the agents randomise over a and b with the probability of 0.5. The coordination problem is then constituted by the problem of agents having to coordinate on one particular out of the multiple available Nash equilibria.

According to Lewis' analysis, a regularity R in the members of a population P when they are agents in a recurrent situation G is a convention if and only if it is true that, and it is common knowledge in P that, in almost any instance of G among members of P,

1. almost everyone conforms to R;
2. almost everyone expects almost everyone else to conform to R;
3. almost everyone has approximately the same preferences regarding all possible combinations of actions;
4. almost everyone prefers that any one more conform to R, on condition that almost everyone conforms to R;
5. almost everyone would prefer that any one more conform to R' on condition that almost everyone conform to R';

where R' is some possible regularity in the behaviour of members of P in G, such that almost no one in almost any instance of G among members of P could conform both to R' and to R (Lewis, 1969, p. 78).

This definition not only captures and elaborates on the characteristics (i) and (ii) in the informal description of conventions of Section 2. If, as done by Lewis, R is constructed as the repeated selection of a particular strategy profile in a game, it effectually attributes to conventions the property of being behaviour convergent to what he calls *proper coordination equilibria*. That is, in game-theoretic terminology, behaviour convergent to one out of multiple available *strict* Nash equilibria (a Nash equilibrium is strict when each player likes this not only *at least as well* but *better than* any other strategy profile he could have reached given the actions of the other players). In the game of Matrix 1 this makes both pure strategy Nash equilibria proper coordination equilibria, while the one in mixed strategies does not qualify as it does not satisfy the second requirement of strictness. In this way, a game like that of Matrix 1 may work as a framework of thought for exploring conventions as defined by Lewis.

As Lewis himself recognises, however, his definition gives rise to at least two fundamental questions pertaining to the dynamics of conventions. First, what keeps the set of expectations leading to recurrent coordination on a specific proper coordination equilibrium *stable*? That is, as Player 1's choice of action depends on what he expects Player 2 to do and vice versa, what ensures that Player 1 expects Player 2 to conform prior to his own decision to also do so?

Second, how does a specific convention *emerge* for a recurrent coordination problem, now that alternatives by definition are said to have been open prior to their establishment? For most conventions no agreement seem to have ever been made. Further, the social processes in which they emerge and operate are so large and complex that communication rarely seems to have been extensible to such a degree as to provide for simultaneous and unambiguous agreement. Not to mention that if the semantic rules of natural languages themselves are taken to be conventions of coordination—as they indeed are by Lewis—then how did they come about, if not by agreement? As Quine recounts in his foreword to Lewis' *Convention*:

> When I was a child I pictured our language as settled and passed down by a board of syndics, seated in grave convention along a table in the style of Rembrandt. The picture remained for a while undisturbed by the question of what language the syndics might have used in their deliberations, or by dread of vicious regress. (Lewis, 1969, p. xi)

For theoretical purposes, these empirical considerations make for the general adoption of a *non-cooperative contest* interpretation of the game-theoretic models, utilised in analysing social conventions.[1] In other words, games are modelled under the assumption that players do not have the possibility of making binding commitments or engage in pre-play communication (cf. Binmore 1990). Recognising this allows one to pose the questions of emergence and stability as a single question within the framework of classical game theory: how may strategically rational players come to coordinate their choices repeatedly on particular out of multiple available proper coordination equilibria in a recurrent coordination problem G, where any instance of G is a non-cooperative pure coordination game of contest?

Unfortunately, it has turned out that classical game theory is inherently and unhelpfully indeterminate in explaining this. Any proper coordination equilibrium in any instance of G is by definition always just one out of multiple Nash equilibria. Consequently, the general Nash equilibrium selection problem applies. This states that play of *any* available Nash equilibrium is consistent with the assumptions of perfect rationality and common knowledge, thereby making the players of classical game theory indeterminate in selecting between proper coordination equilibria like those in the game of Matrix 1. But not only this. It also leaves them without any reason for disfavouring the mixed-strategy Nash equilibrium of the game relative to the proper coordination equilibria. Within Lewis' theory of convention, this raises the additional and fundamental question of why only proper coordination equilibria should arise; or reversely, why conventions should only be attributed the nature of proper coordination equilibria? The fact that since then, the *folk theorem* has revealed that *any* outcome securing the minimax outcome for each player is an equilibrium when a game is repeated—like it is presumed to be in the theory of convention—only serves further disillusion.

Lewis seemed to recognise the basics of this problem. In dealing with the equilibrium selection problem in his theory of convention, he adopted his own version of Thomas Schelling's

[1] Obviously it is not necessary to make this interpretation for all practical purposes. It is not the purpose of a theory of convention to explore every particular social convention without recourse to other social conventions such as language or promise making. In order to explore complicated conventions like human languages, one obviously has to consider how one convention can evolve from another: an exploration of present conventions will normally refer to the conventions of yesterday. Still, what is required theoretically is that social conventions are not among the deepest foundations of the theory itself (cf. Sugden 2000, p. 104).

focal point theory as part of his answer. In *The Strategy of Conflict* Schelling (1960) had argued, roughly, that certain psychological and logical associations that agents might have with particular actions or outcomes could make them *salient* or *prominent* in a way that would serve to focus expectations. In order to show this, Schelling had conducted a series of experiments on pure coordination problems. In these, subjects could not communicate, yet succeeded in coordinating far beyond what chance would prescribe. However, as the details underlying focal points are usually abstracted away in the process of constructing the mathematical game models, Schelling had argued that an explanation of coordination could only be provided by the addition to game theory of an empirically based theory of focal points.

Picking up on Schelling's line of thinking, Lewis argued that salience could be used to explain not only the emergence (including the de-selection of mixed-strategy equilibria, which Lewis found could hardly become salient), but especially the stability of conventions.[2] He hypothesised that agents "will tend to pick the salient as the last resort, when they have no stronger grounds for choice" (Lewis, 1969, p. 35), and that this tendency is a matter of common knowledge. Lewis then claimed that, given salience of a strategy profile or the strategies associated with this, a basis was provided for agents to achieve coordination *rationally*; that is, a basis from which systems of concordant mutual expectations (that everyone will do his part in pursuing the strategies associated with a salient outcome) can be derived (Lewis, 1969, pp. 27–33).

If granted the truth of this argument, Lewis' theory of convention is able to answer the question posed above with regard to explaining the emergence and stability of conventions. Conventions may emerge as results of agreement, coincidence, imagination or the like bestowing salience upon particular strategies or their combinations that make up proper coordination equilibria. That is, given a salient proper coordination equilibrium and common knowledge of this, each agent will tend to choose this; if not only from his own tendency to do so (acting from primary salience), then from the fact that he has reason to expect everyone else to have this tendency (acting from secondary salience); and if not solely from his knowledge of their tendency to make such choices in general, then also from him inferring that they have reason to expect him to make such choices and expect the same of them, and so on. Once successful coordination is achieved, a case of precedent from which to project salience onto future outcomes or strategies will then have been established. In this way, conventions become self-perpetuating due to *salience by precedence* and common knowledge of this—ultimately almost everyone expects almost everyone else to conform to R, and given these expectations stability is trivially ensured through each agents' rational choice of action.

4 The Problem of Salience

In this formulation, however, the role ascribed to salience by Lewis marks a departure from, or addition to, the notion of strategic rationality characterising what came to be the 'classical' approach in game theory. If Player 1 of the game of Matrix 1 thinks that Player 2 has some tendency to aim for a particular strategy or strategy profile, his rational choice of aiming at this as well should be characterised as parametric rather than strategic: he is reasoning from the assumption that the parameters of his situation is given prior to his decision, rather than from

[2] In fact, Lewis was not particularly interested in explaining the emergence of conventions or de-selection of mixed-strategy equilibria. His primary aim was to analyse language as based on convention, which led him to consider these issues only peripherally and concentrate on stability instead (cf. Cubitt et al. 2003).

assumptions about the rationality of Player 2 and common knowledge. The same is ultimately the case if he acts on secondary salience and so on. To that extent *salience* is a feature extraneous to the strategic structure of a game as well as the framework of classical game theory. Consequently, though almost all game-theorists recognise the important role played by this feature, its existence has generally been used in an ad hoc way to rationalise intuitively plausible but theoretically unsupported claims about equilibrium selection (Sugden, 2001, p. 116).

Still, some attempts has been made since then to annex the role of salience into the realm of strategic rationality. Common to these are that they operate with two key elements (cf. Janssen 1998): (a) the idea that salience functions by transforming the personal description of coordination games so as to make the salient outcome a uniquely Pareto-efficient equilibrium; and (b) some kind of principle of coordination according to which strategically rational agents have reason to play their part in such an equilibrium (see also Gauthier 1975, Crawford et al. 1990, Bacharach 1993, Colman 1997 and Janssen 1998).

For instance, Bacharach (1993) and Janssen (1995) have concentrated on explaining the nature of salience for observations such as Schelling on focal points and more controlled experimental replications and developments of this like those of Metha et al. (1994a,b). Common to their explanations is a reliance on the notion of the *availability* of clues of salience due to certain physical, logical or other features associated with certain strategies or outcomes. To be specific, each player is assumed to observe a certain number of such features having some kind of primary salience. Each of these features are then ascribed with a commonly known or approximately shared probability stipulating its availability—that is, its potential for being recognised by each player. From this a personal description of a coordination game may then be constructed on the basis of the expected payoffs associated with choosing in accordance with strategies based on salience by each player. Given that choosing according to one salient feature is associated with a higher expected payoff than any other, rational agents are then asserted to choose this from a principle of coordination.

Before discussing the principle of coordination it should be noticed that this type of explanation may be considered as a rationalisation of what Lewis might have meant by saying that when given no other reason agents tend to chose the salient and that this is common knowledge up to a certain level; and it definitely seems plausible for some situations where logic applies like in Schelling's coordination problem of choosing the same out of all positive numbers. However, it relies on some quite strong assumptions that can hardly be expected to hold in general for most of those situations in which real world social conventions emerge and operate. First of all, the explanation requires, roughly, that all features should be recognised by all players and that the probabilities of the availability of these features should be either common knowledge or at least approximately shared. However, none of these two assumptions seem to plausibly apply for real world social interaction in the complex settings governed by social conventions. For these, people are usually quite unaware of what they are doing and why, other than following *precedent*. In fact, it seems plausible that one significant reason for following precedent is, precisely, that real world people would be unable in real world situations to satisfy assumptions like these two.

Also, if this kind of explanation is taken to apply to particular regularities an interesting point follows. If a regularity for solving a coordination problem attaches to a *particular* clue of salience which covers all instances of G, the behavioural regularity R observed is ultimately not to be regarded as a convention as it rules out contingency; that is, unless the function attributed is a generalised one covering types. Though attributing from types may be regarded as a fair methodological move, it significantly detracts from the explanatory adequacy and should be remembered when carefully exploring conventions on this level of abstraction. Keeping to particular regularities, it is for considerations like these that some form of minimal functionalism in the definition of conventions is usually pressed—in some way, social conventions are part of the causal chain resulting in their own stability. Still, this does not exclude the possibility that non-contingent regularities may become conventional, and that is without changing appearance to an observer. If wanting to give this phenomenon which has hitherto gone unnoticed a name we might consider referring to it as that of 'paragliding conventions'. Lifted by other means, such as rationality, it flies on the air of something else; and that is most likely *precedence*. This, however, turns the discussion of salience into one independent of the transformation argument based on the availability of clues or other transformation arguments. Instead, the discussion becomes one concerning the notion of precedent alone.

5 Precedent and Coordination

The seminal account of salience provided by Gauthier (1975) illustrates how salience by precedent may work on this approach. His argument may be reworked by assuming salience of (a, a) in a pure-coordination game like that of Matrix 1. Gauthier claims that salience transforms the personal description of this game for each of the players to that of Matrix 2. In this game each player faces two options: 'seeking salience' and 'ignoring salience'; or in the case of precedent: 'seeking precedent' and 'ignoring precedent'. The former leads to the realisation of (a, a) in the original game of Matrix 1, while the latter is likely to lead each agent to randomise over a and b. In the transformed game two Nash equilibria exist—one proper coordination equilibrium, and one non-strict Nash equilibrium—but according to the principle of coordination invoked by Gauthier, similar to that of Bacharach and Janssen, rational players now have reasons for choosing their 'seeking salience' strategy since this leads to a Pareto-efficient outcome. Hence successful coordination in the original game is explained in consistency with the assumption that the regularities regarded as conventions represent contingent behaviour.

		Player 2	
		'seeking salience'	*'ignoring salience'*
Player 1	*'seeking salience'*	**1, 1**	0.5, 0.5
	'ignoring salience'	0.5, 0.5	**0.5, 0.5**

Matrix 2: Gauthier transformation of the game in Matrix 1

Against this argument of transformation, Gilbert (1989) has argued quite convincingly that Gauthier does not seem to have any good reason to restrict the options of the transformed game to two. In particular, sticking to the two-strategy case utilised here as well as in Gauthier (1975), there seems to be at least one other alternative, namely that of seeking the 'non-salient'—in casu the 'non-precedent'. Allowing for this third possibility ultimately destroys Gauthier's account. This is revealed by the Gilbert-based correction of the transformed game of Matrix 2 to Matrix 3. In this game no unique Pareto-efficient equilibrium exists. Hence Gauthier's Principle of Coordination cannot operate successfully.

		Player 2		
		'seeking salience'	*'ignoring salience'*	*'s. the non-salient'*
Player 1	*'seeking salience'*	**1,1**	0.5,0.5	0,0
	'ignoring salience'	0.5,0.5	0.5,0.5	0.5,0.5
	's. the non-salient'	0,0	0.5,0.5	**1,1**

Matrix 3: Gilbert-correction of Gauthier's transformation

Now, one could retort that this argument does not apply when more than two strategies are available (Janssen, 1995). This makes for asking whether it is plausible that only one strategy salient by precedent exists—in particular, in the case of the hyperrational agency that must be invoked to make the calculations needed when many strategies are available.

One may answer by noticing that a hyperrational player would be capable of recognising all possible patterns that precedent could be taken to follow. Following (Sugden, 1989, p. 190), imagine a player who has played a pure coordination game like that of Matrix 1 a 1.000 times, and on every repetition has met players choosing strategy a of the original game. It may seem obvious that the rational inference for him to draw is that the next player he encounters will very probably choose a as well. But although this inference is obvious in the perspective of common-sense, it is not accessible to perfectly rational players. For real world individuals, the fact that all 1.000 encounters fit the pattern 'always a' is a remarkable fact, which calls for some explanation beyond pure chance. But, from the perspective of a hyperrational player, such reasoning is merely a betrayal of the lack of imagination. For him, every sequence of 1.000 instances of 'a' and 'b' has some pattern that can be projected into the future. Consequently, within the framework of classical game theory, transformations like that of Gauthier's are infeasible whether only two or more strategies are available.

However, ultimately one may still entertain the idea that some Principle of Coordination may facilitate coordination in non-transformed games possessing structures similar to the game of Matrix 2, where multiple proper coordination equilibria exist, but where one of these Pareto-dominates all others—if only to exhaust any transformation argument along the lines of the two above. This raises the general question of whether Pareto-efficiency somehow provides strategically rational players with reasons to pursue such an equilibrium if unique. According to Heal (1978) intuitively this must be case.[3] Choosing a salient strategy, in casu one salient by being the unique Pareto-efficient equilibrium, seems the rational thing to do. The players "know that they can coordinate their choices if they can single out one [strategy] from the rest". They also know that "by choosing a [strategy] which does stand out for both of them, and *only* by doing this, can they hope to coordinate. This provides a reason for each to make the choice of the outstanding [strategy], which is reinforced by knowledge that the other also has that reason" (Heal, 1978, p. 129).

Gilbert (1989), Sugden (1991) and Colman (1999) have pointed out that arguments along this line fail. It does not explain why it is rational for a strategically reasoning player to choose a salient strategy without any reason for assuming that other players will also choose the salient strategy, aside from the knowledge that the other agents confront the same coordination problem. The fact that choosing a salient strategy has powerful intuitive appeal does not, in itself, constitute

[3] Heal does not make her argument with special regard to Pareto-efficient equilibria, but to salient equilibria in general, see Heal (1978).

a rational reason for choosing it. The point is that although it is obviously true that the players succeed in coordinating if they both choose the same salient strategy, and although it is true by definition that successful coordination is a good outcome for them, that does not provide them with any rational grounds for behaving in that way. As Gilbert puts it, "the fact that a good outcome would be reached if *both* did something cannot by itself be a reason for either one individually why he should do it. For his doing it cannot ensure that the other does it" (Gilbert, 1989, p. 72). Thus, "if human beings are—happily—guided by salience, it appears that this is not a consequence of their rationality" (Gilbert, 1989, p. 61). Consequently, any principle of coordination assuming that agents will automatically play their part of a Pareto-efficient equilibrium cannot be derived from premises of rationality and common knowledge.

Ultimately, what this argument amounts to is the claim that salience, whether by precedent or any other means, may not facilitate strategically rational players with reasons for acting in conformity with conventions. Annexing the relevant notion of salience in relation to social conventions into the realm of strategic rationality ultimately must fail due to strategic rationality being purely forward looking. If seeking to understand the dynamics of social conventions, or for that matter, the nature of salience in the context of these, one has to look somewhere else.

6 The Evolutionary Turn and the Theory of Convention

During the 70s and 80s evolutionary game theory arose as a result of applying the game-theoretic framework to problems in evolutionary biology. Though classical game theory was developed for approaching social phenomena as aggregate products of individuals' strategic decision making, evolutionary biologist John Maynard Smith and colleagues demonstrated that it also provided a powerful framework for explaining various aspects of animal behaviour and evolution, see Maynard Smith & Price (1973) and Maynard Smith (1982).

Their utilisation of game theory, however, was not just carbon copied. They adjusted and developed the framework in important ways. Instead of assuming agents to be fully informed and hyperrational, the agents of evolutionary game theory came to be understood as biologically or socially "pre-programmed" (conditioned) for certain strategies. Also, where the baseline models of classical game theory are games played exactly once, the baseline models of evolutionary game theory came explicitly to be games played over and over again by agents randomly drawn from large populations on which some evolutionary selection process operates over time on the population distribution of strategies.

Besides providing insight into problems of evolutionary biology it was soon realised that the development of evolutionary game theory also provided a novel way of dealing with what figures as the equilibrium selection problem within the classical approach. The first large achievement came by the provision of the concept of *evolutionary stability* pointing to hitherto neglected stability features of the strategy profiles figuring as Nash equilibria on the classical approach, see Maynard Smith & Price (1973). In particular, while some Nash equilibria turns out to be evolutionary stable, others are revealed to be evolutionary unstable, why behaviour convergent to these should not be expected to persist in the long run.

Formally, an *evolutionary stable strategy* (ESS) may be defined as follows. For any two-player symmetric game G with the finite set of pure strategies, $S = \{\mathrm{a}, \mathrm{b} \ldots \mathrm{m}\}$, the same for all players, and a corresponding set Δ of mixed strategies, any mixed or pure strategy, x, is an ESS iff

$$u(x,x) > u(y,x), \text{ or} \tag{1}$$

$$u(x,x) = u(y,x) \text{ and } u(x,y) > u(y,y), \tag{2}$$

where, $x, y \in \Delta$, $x \neq y$ and $u(x,x)$ and $u(x,y)$ is the expected payoff from playing strategy x against strategies x and y, respectively, and $u(y,y)$ of playing strategy y against y.

From this definition it follows that a population in an evolutionary stable state, a state where all agents play the same ESS, has converged to what amounts to a Nash equilibrium of the game on the approach of classical game theory. That is, an ESS is always a Nash equilibrium by definition. However, the opposite does not hold. The definition of a Nash equilibrium does not exclude the possibility of such to rely on a weakly dominated strategy. In cases like this the weakly dominated strategy may do just as good against the weakly dominant strategy, but better against itself than the weakly dominant strategy does against itself. Consequently, the behaviour of the population may exhibit a phenomenon called *drift*. Here individual changes in strategy by error, creativity or experimentation does not inflict a payoff loss to the 'deviator', and when meeting other 'deviators' these may do better against each other than the 'conformists' do against each other, whereby the original state is disrupted. What this reveals is that the set of evolutionary stable states identifies a subset of the set of Nash equilibria in a game, that is, the ESS criterion refines the Nash equilibrium concept in an evolutionary setting.

At least three features of the ESS concept are, however, important to notice. First, the ESS concept refers implicitly to a close connection between the utilities in a game and the spreading of a strategy in a population. This presupposes that the payoffs in a game are supposed somehow to represent a gain in social reproductive fitness of a strategy from the interaction in question. Second, the ESS concept only applies when the population is large and the 'mutation-rate' or 'experimentation' is low (cf. Weibull 1995, pp. 33–35). Although credible within biology, this assumption may raise some questions when transferred to social phenomena. However, third, and perhaps most important, as with the Nash equilibrium concept, the ESS concept does not explain *how* and *how likely* a population arrives at an associated evolutionary stable state. Instead it asks whether, once reached, such a state is robust to evolutionary pressures. Hence, it provides a refinement of the Nash equilibrium concept in an evolutionary setting, rather than a real alternative to this.

Such an alternative was, however, provided by the second large achievement—that of the particular population dynamics first formulated by Taylor & Jonker (1978), which later came to be dubbed the *replicator dynamics* in Schuster & Sigmund (1983). Other dynamics has been developed since, but this is the most widely used. While the criterion of evolutionary stability highlights the role of mutations, the replicator dynamics highlights the role of selection. It does this by providing a model of such a selection dynamics capable of describing how the distributions of different strategies evolve over continuous time. Specifically, it takes the rate of growth of the frequency with which any given strategy is played in a population state to be proportional to the difference between the expected payoff of playing that strategy in this state and the weighted average of the expected payoffs of all strategies played in that state (each strategy being weighted by the frequency with which it is played). Formally, then, the replicator dynamics may be defined by the differential equation

$$x_{t+1} = [u(x,y) - u(y,y)]x_t,$$

where $u(x,y)$ is the expected payoff to any strategy x at a random match, when the population is in state $y \in \Delta$; $u(y,y)$ is the expected payoff to any mixed or pure strategy y mirroring the population distribution of strategies when played against itself (i.e., the current average payoff in the population); x_t is the population share playing strategy x at time t; and x_{t+1} is the population share playing strategy x at time $t + 1$. Consequently, the *growth rate* x_{t+1}/x_t of the population share using strategy x equals the difference between the strategy's current payoff and the current average payoff in the population.

Features of dynamic processes specified for games by equations like this may be described in the following terminology (cf. Binmore 1990). An *initial point* of a dynamic process is the point from which it begins at $t = 0$. A process such as the replicator dynamics then describes a *trajectory*. A trajectory may do various things. In particular, it may converge or diverge. Except for pathological cases, a convergent trajectory converges to a *fixpoint*. Such are defined by being initial points from which the dynamic process never moves. The *basin of attraction* of a fixpoint is the set of initial points from which the dynamic process converges to this. If a fixpoint's basin of attraction consists of every possible initial point, then the fixpoint is a *global attractor*. A *local attractor* is a fixpoint that lies in the interior of its basin of attraction. Henceforth attractors are referred to as evolutionary stable states. Finally, some fixpoints are not evolutionary stable at all. Their basin of attraction is a singleton. Hence, no-one would ever want to be found predicting that the long-run outcome of a dynamic process will be such a non-stable state. Even if the process started at such a point, any small perturbation could push it into a trajectory in the basin of attraction of evolutionary stable state making the prediction wildly wrong.

As it turns out, fixpoints that are evolutionary stable states always correspond to what amounts to a Nash equilibrium of the game on the classical approach. However, as in the case of the ESS the reverse does not hold (cf. Weibull 1995). On this background it is possible to see how evolutionary game theory might provide a framework for solving what amounts to the equilibrium selection problem on the classical game-theoretic approach. Given an evolutionary dynamics such as the replicator dynamics, plus some slight mutation rate (notice the replicator dynamics is deterministic and hence does not incorporate mutation by itself) it may be explained how a population converges to one particular out of multiple available evolutionary stable states. Granted some initial point, high-performing strategies may be observed to increase, whereas low-performing strategies decrease and eventually disappear, depending on which basin of attraction the initial population state is located in. This introduces a novel factor into the analysis of games. The particular initial point of the process determines which stable state, if any, the process converges to.

7 Evolutionary Games and Conventions

Having become acquainted with the basics of evolutionary game theory, the details of how the theory of convention might be reconstructed and developed within this framework may now be outlined. This section argues in the context of a variety of simple but paradigmatic games of conventions that the framework of evolutionary game theory may be used to explain the emergence and stability of social conventions, as well as explore several features of conventions that it is hard if not impossible to accommodate within the classical approach.

7.1 Evolution and Coordination

In *Convention*, Lewis modelled a simple coordination problem underlying conventions, the symmetric two-player game of pure coordination (Matrix 1). When speaking of a symmetric two-player game one standardly assumes that there are precisely two player 'positions', that each position has the same number of pure strategies (in the sense that they are identical), and that the utility to any strategy is independent of which player position it is played in (Weibull, 1995). In the context of the theory of convention, however, it makes sense, for reasons to come, to follow Sugden (1989, p. 14) and further require of a symmetric game that it looks exactly the same from the viewpoints of the two players; that is, a game where the agents do not know, or alternatively, do not attach any significance to their assigned 'position' as Player 1 or Player 2. This latter requirement makes no difference to the perfectly rational agents of classical game theory who are responsive only to the strategic structure of a game. As it will appear, however, it makes an important difference in a theory of convention, in so far as evolutionary game theory is taken to provide a rigid way of thinking about individual learning or social imitation processes.

The single most referred to example of a regularity taken to qualify as a convention under Lewis' definition is that of drivers keeping to one particular side of the road—*the rule of the road*. Conceptualising the underlying problem solved by conformity as one of choosing between *keep left* or *keep right* this coordination problem clearly qualifies as one to be modelled as the simple symmetric pure coordination game of Matrix 1 (reproduced in the present context below).

		Player 2	
		keep left	*keep right*
Player 1	*keep left*	**1,1**	0,0
	keep right	0,0	**1,1**

Matrix 4: The driving game

On Lewis' definition of a social convention two potential conventions exist in this game. These are the two proper coordination equilibria where either both players keep left or both players keep right. Still, besides revealing that the framework of classical game theory could not answer the question pertaining to the possible emergence and stability of conventions like these (even if salience is granted of a particular outcome or strategy), the argument of Sections 4 and 5 also revealed no credible reason beyond the definition as to why the mixed strategy Nash equilibrium of this game should be denied the status of a potential convention.

Now, turning to an evolutionary analysis, things look quite different. First, it may be noticed that in the game of Matrix 4, if played recurrently within a single population by pairs of randomly matched agents, only the two Nash equilibria in pure strategies—the two proper coordination equilibria—but not the one in mixed strategies correspond to evolutionary stable states. This is confirmed by the replicator dynamics, which drives the population to one of the population states in which everyone either plays strategy a (*keep left*) or strategy b (*keep right*). In particular, this is also the case if we start the population in the mixed strategy Nash equilibrium and invoke slight perturbations. Sooner or later one agent will switch strategy by error, choice, or whatever, whereby the average payoffs of all strategies in the game changes so as to favour the strategy that the 'switching' agent adopted. That is, the population is pushed into either of the basins of attraction belonging to one of the evolutionary stable states and separated by the mixed strategy

Nash equilibrium. Given a low perturbation rate and the positive feedback loop between payoffs and play of particular strategies stipulated by the replicator dynamics, more and more agents will then adopt the strategy in question whereby the incentive for others to do so as well is raised even more. Ultimately, a convention emerges and is kept stable by evolutionary forces.

Now, what is interesting to notice here, besides the emergence of conventions, is that as conventions like these become established, each agent has an increasing reason to expect other agents to have a tendency to aim for the strategy associated with an emerging convention. Though such reasons have no effect on the pre-programmed agents of evolutionary game theory, should any one agent bite the Apple of Eden he would surely recognise this and act in accordance from considerations based on parametric rationality. In this way Lewis' notion of salience may partly be accommodated with an evolutionary framework.

7.2 Evolution and Pareto-Efficiency

But what about the idea entertained by the coordination principle saying that Pareto-efficient conventions should be observed to be favorued relative to inefficient ones?

		Player 2	
		a	b
Player 1	a	2,2	0,0
	b	0,0	1,1

Matrix 5: A pure-coordination game with a unique Pareto-efficient equilibrium

Peter Kincaid (1986) gives examples that lead to the conclusion that payoff equivalence should not be attributed potential conventions for most kinds of pre-automobile traffic facing the pure-coordination problem of which side of the road to drive on. For instance, when leading a horse with a hand, the fact that most people are right-handed, together with the fact that a hand-led horse tends to kick away from its leader, makes staying on the right hand side of the road a Pareto-efficient solution relative to left-side 'driving'. The opposite situation is the case for traffic dominated by riders, because right-leggedness in combination with the practice of mounting and dismounting makes left-side riding a Pareto-efficient solution. Still, it may be maintained that, despite the Pareto-efficiency of a particular solution, the recurrent situation is still one of a coordination problem, captured by games like that of Matrix 5 and that such efficiency is in fact a clear-cut example of how salience may not only arise from the labeling of strategies or the history of a game, but also from the very strategic structure of this itself. Yet, though salience in this case is inherent to the strategic structure of the coordination problem, the arguments of Sections 4 and 5 still hold: either this regularity is not a convention, or it is a convention by definition, working through precedent, why classical game theory is unable to accommodate an explanation of its stability.

On an evolutionary analysis, however, the emergence of conventions again follows. Just as for the pure coordination game of Matrix 1, only the two pure strategy Nash equilibria, but not the one in mixed strategies, correspond to evolutionary stable states if played recurrently within a single population by pairs of randomly matched agents. However, as the state forming the separating point for the basins of attraction of the potential conventions again corresponds to the mixed strategy Nash equilibrium of the game, it turns out that the replicator dynamics drives the population to the Pareto-efficient convention as soon as more than $\frac{1}{3}$ of the population is made up of agents playing a. Consequently, if the initial distributions of agents playing either of the strategies a and b are taken to be formed at random, then two-thirds of these populations will be driven to the Pareto-efficient convention. In conclusion, it turns out that evolution tends to select Pareto-efficient conventions.

7.3 Evolution and Risk

Now, it would be good if this conclusion would cover all cases of social conventions. Unfortunately, the game of Matrix 6 gives evidence to the contrary.

		Player 2	
		attending	*staying home*
Player 1	*attending*	**2,2**	0,1
	staying home	1,0	**1,1**

Matrix 6: The stag hunt game

This game illustrate the common predicament of considering whether to attend some collective activity (e.g., soccer training) whose payoff depends upon the attendance of others, or attending some individual one with a guaranteed payoff (e.g., staying home and watching TV). In such situations, one would have preferred to stay at home if others do not show up, but if they do show up one would have preferred to go as well. The game associated with this type of problem is usually referred to as *The Stag Hunt* game due to a story by Rousseau and is a sub-type of the general class of pure coordination games. Special to it is that although agents in this type of game agree on their preference for one particular equilibrium, (a, a), alternatives are less risky, (b, b). In the terminology of Harsanyi & Selten (1988) the latter equilibrium is *risk-dominant*. It instantiates a strategy profile of best responses between opponents who are equally likely to play either of their strategies a and b. Coordination on the Pareto-efficient equilibrium is thus a matter of confidence in other agents' intentions to coordinate on this as well, which again turns on their confidence in one's own intention.

Conventions for solving this type of coordination problem are abundant, and so are conventions for what is to count as acceptable excuses from or credible signals for attendance; and all of these and their properties seem intuitively to revolve around the tension between the salience of Pareto-efficiency and relative to risk dominance. Yet, as is the case for the two previous games classical game theory is unable to explain their emergence and stability. That is, unless the contingent nature of such conventions are derived from variations in the de facto confidence that agents assign to each others intentions for coordinating on the Pareto-efficient equilibrium. But then, again, such variations would be derived from earlier experience, namely precedent.

On an evolutionary analysis, however, the emergence and stability of conventions may once again be accommodated. For the stag hunt game, only conventions corresponding to the two pure-strategy Nash equilibria, but not the one in mixed strategies, are evolutionary stable states if played recurrently within a single population by pairs of randomly matched agents; and again, the state forming the separating point of the basins of attraction of the potential conventions corresponds to the mixed strategy Nash equilibrium. However, for the stag hunt game of Matrix

6, this is located when the probability of playing a is 0.5. Now notice that, while the equilibrium-payoffs in this game are identical to those of the game of Matrix 5, the basin of attraction for the Pareto-efficient convention has shrunk. That is, the risk dominance of the Pareto-dominated equilibrium (b, b) is counteracting the attraction of the Pareto-efficient convention. In fact, it turns out that if individual learning or social imitation processes are portrayable as evolutionary processes, learning and imitation dynamics in general tends to select risk-dominant conventions relative to Pareto-efficient ones (cf. Sugden 1999, p. 458).

In conclusion, what the shift to an evolutionary framework shows us is that any learning or social imitation dynamics for which the evolutionary dynamics may function as a metaphor will tend to drive a population playing a symmetric coordination game towards evolutionary stable conventions instantiating one of the pure-strategy Nash equilibria on the classical approach. In particular, it shows that such coordination may be achieved in the absence of strategic rationality. Further, the evolutionary reconstruction seems to lend substance to Lewis' claim that we should not expect to see behavioural patterns corresponding to the mixed strategy equilibrium materialise; or alternatively, why we should not attribute to behavioural patterns the nature of mixed strategy equilibria in this type of games. Thus, fundamental assumptions of the theory of convention are apparently saved.

7.4 Evolution and Labelling Asymmetries

Things, however, are not that simple when trying to extend the evolutionary framework so as to include paradigm examples of what has been referred to as *labelling asymmetric* or discriminatory conventions of coordination (cf. Sugden 1989 and Hansen 2006). Consider the division of labour game of Matrix 7.

		Player 1	
		a	*b*
Player 2	*a*	0,0	**1,2**
	b	**2,1**	0,0

Matrix 7: The division of labour game

This game, usually referred to as *the battle of the sexes*, is a game of partial coordination and has often been attributed to social systems in which conventions for a division of labour is established (see e.g., Ullman-Margalit 1977). To see why, notice that this game illustrates why even disfavoured individuals may be reluctant to challenge a given division of labour and try to prevent members sharing their situation to do so, thereby keeping the convention stable. Given the fact that an established convention for a division of labour instantiate a strict Nash equilibrium, deviation by disfavoured individuals inflicts a payoff loss on themselves. Likewise, if the game is taken to be played by individuals associated with a group constituted by some arbitrary feature individual deviation may affect the expectations dependent on this feature by other groups of players, such that deviation may end up inflicting a payoff loss on the rest of the group (this latter intuition, not facilitated by classical game theory, turns out to be supported by the evolutionary approach below).

However, if taking an identical evolutionary modelling approach to this game as to the previously analysed games, it turns out that the only evolutionary stable state is that corresponding to the mixed strategy Nash equilibrium. Consequently, the mixed strategy Nash equilibrium is also the global attractor to which any initial population converges, including initial populations instantiating one of the pure strategy Nash equilibria. Consequently, evolutionary dynamics like the replicator dynamics inevitably drives any population to the state assigning the probability of $\frac{1}{3}$ to the play of strategy $\mathtt{a}$ and the probability of $\frac{2}{3}$ to the play of strategy $\mathtt{b}$. Apart from appearing utterly disappointing in the context of the theory of convention, the agents playing this game have also reason to be disappointed. This state leaves them with the mere expected, though more 'egalitarian', payoff of $\frac{2}{3}$, which is significantly worse than if they were able to coordinate on one of the potential conventions of the game. But why does evolution not favour the emergence of conventions in this game?

One may see why by considering a case of a discriminatory convention reported by Lewis of re-establishing cut off phone calls. For a period, all local phone calls were cut off without warning after three minutes in his hometown Oberlin, Ohio due to technical problems. Soon, Lewis reports, "a convention grew up among Oberlin residents that when a call was cut off the original caller would call back while the called party waited" (Lewis, 1969, p. 43). Now, intuitively this game should be formalised as the Telephone Game in Matrix 8 below.

		Original Receiver	
		call back	*wait*
Original Caller	*call back*	0, 0	**1, 1**
	wait	**1, 1**	0, 0

Matrix 8: The telephone game

This game shares a crucial feature of the strategic structure with the division of labour. Just as in the division of labour the players have to coordinate on playing *different* strategies relative to each other in the telephone game; and in particular, just like for the division of labour game it turns out that the replicator dynamics carries any initial population to the mixed strategy Nash equilibrium. This is located in the state where everyone is conditioned for playing *call back* half of the time and *wait* the other half (or, alternatively, the polymorphic population state where half of the agents play *call back* all the time, while the other half play *wait*), a state yielding the average payoff of 0.5 to each agent. However, such a formulation makes it clear what is wrong with directly applying the single population replicator dynamics to the division of labour game as well as the telephone game. People in Lewis' hometown where obviously able to coordinate because they where able to *condition* their choice of strategy on whether they were the original caller or not. This is equivalent in the game-theoretic framework to saying that they were able to condition their choice of strategy on their player position, a labelling asymmetry of the game which according to classical game theory is irrelevant. Unfortunately the single population replicator dynamics does not take this possibility into account. Agents are conditioned to playing some pure or mixed strategy of *call back* or *wait* without giving any attention to whether they are the original caller or not.

7.5 Discriminatory Conventions and Multi-population Models

Fortunately the evolutionary framework may be tailored to deal with this problem. If agents begin to condition their strategies on their player positions, then their individual learning processes will have to operate within two different scenarios: one in which an agent is the original caller and one in which he is the original receiver. This leads the exploration into the field of evolutionary selection in multi-population models.

In multi-population models it is assumed that large (technically infinite) populations of agents interact, one such population for each player position of the game. Repeatedly, agents are randomly drawn—one for each population—to play the game. Formally, a population state is identical with a pure- or mixed strategy for a player position. It is these population states that are modelled as interacting. Taken together, such states (one for each population) constitute a pure- or mixed strategy profile of the game. However, little may be said in general about multi-population modeling. For instance, there appears to be no strict consensus as to how the criterion of evolutionary stability should be extended to multi-population interactions. Even further, multiple extension of the replicator dynamics exist (cf. Weibull 1995, p. 165).

Still, some interesting conclusions may be drawn. For instance, when some fraction of a population state changes strategy by creativity or error, this fraction will never meet members of its own population, for the simple reason that each agent in any of the interacting populations is always matched with agents from the other population(s). Thus, where such strategies may have done poorly against themselves in single population models, this is not an issue in a multi-population model. They may survive and invade their population due to them doing quite well against the strategies of another population. Consequently, non-strict Nash equilibria like the mixed strategy equilibria of Matrix 7 and 8 become vulnerable to invasions. On this background, the different criteria for multi-populational evolutionary stability are formulated so that they are met *only* by strict Nash equilibria (Weibull, 1995, p. 163). That is, for games of asymmetric labelling such as that of Matrix 7 and 8, the mixed strategy Nash equilibria turns out to be unstable in multi-population models on the evolutionary approach, while only the pure strategy equilibria facilitate stability as they correspond to strict Nash equilibria.

This gives way to an interesting conclusion. Using the 'standard' n-population replicator dynamics formulated by Weibull (1995, pp. 171–181) gives a dynamics for the game of Matrix 7 (Figure 1).

The vertical axis gives the frequency distribution of strategies a and b, respectively, within 'Population 1' corresponding to the player position of Player 1. Likewise, the horizontal axis gives the frequency distribution of strategies a and b, respectively, within 'Population 2', the player position of Player 2. Where the mixed strategy equilibrium profile was seen to correspond to a globally attractive population state for the replicator dynamics in the single population model, the portrayed trajectories of Figure 1 show that population states corresponding to pure strategy Nash equilibria constitute evolutionary stable states in the two-population replicator dynamics of the game of Matrix 7, while the state corresponding to the mixed strategy Nash equilibrium has disappeared as a stable state. The reason for this stark qualitative contrast between the single- and two-population models is that when interaction takes place between two distinct populations, there arises the possibility of *polarisation* in behaviours. The slightest deviation from identical population distributions corresponding to the mixed Nash equilibria may lead the player populations toward specialisation in *different* pure strategies (Weibull, 1995, p.

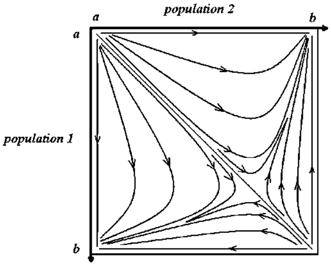


Figure 1: Two-population replicator dynamic solution trajectories in the game of Matrix 7

183). In the game of Matrix 7 this means that one population distinguishable by some arbitrary feature will specialise in (be carried by the dynamics towards) the state corresponding to the Nash equilibrium in which they are favoured, while the other population will move towards the state corresponding to the Nash equilibrium in which they are disfavoured. Which of the two possible evolutionary stable states, that is, which of the two potential conventions, will emerge depends on the initial population state of the game.

7.6 The Emergence of Discriminatory Conventions

However, Sugden (1989) has pointed out that it does not seem plausible that everyone in the population at the same time should come to conceptualise their coordination problem as one facilitating coordination by conditioning their strategies on exactly the same arbitrary feature. That is, it is most likely that to begin with only a small fraction of the population will come to conceptualise their problem as one in which who was the original caller or the like could be relevant instead of just ignoring this. Intuitively one may think of this scenario as a battle between two uneven forces working in opposite directions: one that operates on the majority of agents in the population (namely those who ignores such features) pushing towards the mixed strategy Nash equilibrium, and one that operates on a small minority who by accident attaches significance to this, pushing towards one of the potential conventions based in casu on who is the original caller. However, the evolutionary framework reveals a conclusion to the contrary.

Consider the single population of Oberlin residents playing the mixed strategy Nash equilibrium of the telephone game of Matrix 8. In this population everyone receives a payoff of 0.5 on average. If the population is taken to be monomorphic and any one agent or small fraction of agents should deviate from this equilibrium so as for instance to assign the probability of 0.4 to *call back* when cut off, they would still receive an average payoff of 0.5 when meeting a 'conformist'. Yet, they would only receive a payoff 0.48 when meeting other 'deviators' from their fraction. Consequently, in the long run the fraction of deviators will perish. The same argument pertains if the population is taken to be polymorphic. If any one agent or some small fraction p of agents should deviate from their role of playing either *call back* or *wait* in the state instantiating the mixed strategy Nash equilibrium, then they would receive a payoff of 0.5 in $(1 - p)$ encounters and 0 in p encounters. As p is positive, this will necessarily imply a payoff less than 0.5, the payoff that any conformist will continue to get. Consequently, in the long run the fraction of deviators in a polymorphic population will perish. All of this is as expected: for a single population evolutionary dynamics the mixed strategy Nash equilibrium of the game is an evolutionary stable state.

However, consider now that in the mixed strategy equilibrium 1.6% of the Oberlin residents will experience that in their last three games coordination was established by the original caller calling back while they waited (or the other way around). This might lead some of them to adopt the corresponding convention conditional on their player position on the false belief that other players slightly tend to follow this in general. If this is the case, they will thereby make what was originally a false belief true. Further, this fraction p of deviators or 'conventionalists' will on average receive a payoff approximating that of $0.5 \times (1 - p) + 1 \times p$, which will always be than what conformist receive on average in the mixed strategy Nash equilibrium. Hence, keeping to the intuitive interpretation of the dynamics, their belief in the tendency is likely to be reinforced. As p increases, this payoff increases as well. In the long run, the 'conventionalists' flourish, while the conformist will gradually die out. Now, of course, at some point 'conventionalists' conforming to different potential discriminating conventions will start meeting each other. In this case the initial distribution of these 'conformists' between the potential conventions will determine which of these will become established. Still, the picture is clear. Only a small fraction noticing the possibility of conditioning their strategies on player positions leads to the emergence of one of the potential discriminatory conventions.

7.7 Conventions and Individual Interests

So far, it appears that the emergence of some potential convention for solving a coordination problem is always in the interest of the members of a population. Consequently, conformity to a social convention is always in the interest of the individual parties of that convention. From this it may be thought, then, that if observing agents acting in conformity with some established convention, following this convention is always in their interest relative to a state of coordination failure. That is, should any one member complain about a given convention, one could rightfully remind him that he should be happy about the convention as conformity serves his own interest. However, a conclusion to the contrary follows by considering the *hawk-dove game* of Matrix 9; a slightly amended version of that utilised by Maynard Smith (1982) and Sugden (1989) to explore situations where two agents dispute over a given resource.

		Player 1	
		Dove	*Hawk*
Player 2	*Dove*	2,2	**0,3**
	Hawk	**3,0**	−2,−2

Matrix 9: A Hawk-Dove game

On the classical analysis three Nash equilibria exist: two in pure strategies, where one of the players play 'dove' while the other plays 'hawk', and one in mixed strategies where each player plays 'dove' two-thirds of the time. Further, it may be noticed that the more 'egalitarian' strategy profile where both players play 'dove' is not a Nash equilibrium. Each player would prefer to play 'hawk' as soon as they come to expect the other to play 'dove'.

On the single population evolutionary analysis any population state corresponding to the mixed strategy Nash equilibrium turns out to be the unique evolutionary stable state. In such states each agent receives an average payoff of $1\frac{1}{3}$. Now, it is obvious that everyone would benefit if everyone would instead play dove. In fact, each agent in the unique evolutionary stable state would benefit individually by defecting from this so as to adopt the dove strategy (in polymorphic populations) or, alternatively (in monomorphic populations), if everyone increase the probability of playing the dove strategy. Unfortunately, however, as soon as this happens, every agent will then benefit individually by making a change towards playing 'hawk'. In particular, in the state where everyone play 'dove' an intruding fraction of 'hawks' would flourish until the state corresponding to the mixed strategy Nash equilibrium is restored.

Next, consider the game when agents are given the opportunity of conditioning the play of strategy on player positions. In the case of fighting over a resource one labelling through which this possibility could be acquired is if some small fraction of agents come to believe in a tendency for 'first comers' to play hawk and 'last comers' to play dove. In classical game theory, re-labelling is irrelevant to the strategic structure of the game. But as the previous section revealed it may play an important role for evolutionary processes.

Now, in the hawk-dove game with labelling asymmetry populations states corresponding to the two pure strategy Nash equilibria (assumed to be potential conventions) are the only evolutionary stable states of the game. Consider a polymorphic population playing the state corresponding to the mixed strategy equilibrium of the game. Remember, in this state each agent receives a payoff of $1\frac{1}{3}$ on average. However, should a small fraction come to believe in a tendency for 'first comers' to play hawk and 'last comers' to play dove (or vice versa) and adopt the corresponding convention, these would receive the same payoff on average when meeting 'non-conventionalists', but a payoff of 1.5 on average when meeting other 'conventionalists' given that the asymmetry is perfectly *cross cutting* (the process by which the agents are assigned to either of the player positions is completely random). Consequently, relative to 'non-conventionalists', 'conventionalist' population shares will prosper and ultimately take over the population, which will then be in a state corresponding to one of the two pure strategy Nash equilibria, where 'first comers' play hawk while 'last comers' play dove, or vice versa. These states are evolutionary stable. If 'late comers' are always sure that their opponents will play hawk, their best reply is to play dove; and likewise, if 'first comers' are always sure that their opponents will play dove, their best reply is to play hawk. The same goes for the alternative convention by these labels, where 'late comers' play hawk and 'first comers' play dove. Which of the two conventions result depends once again on the initial distribution of conformist to each potential convention.

Returning to the question of whether following an established convention is always in each individual's interest relative to a pre-conventional state, it may be observed that this depends on whether the asymmetry on which it is based is stable, in other words whether the asymmetry continues to be perfectly cross-cutting. As long as this is the case, all agents will be happy that the convention emerged so as to solve their coordination problem. The established state of convention will be preferred by all to the pre-conventional state corresponding to the mixed strategy equilibrium of the game. If, however, the frequency by which agents are assigned to a player position changes sufficiently, the established convention may stop serving the interest of every participating agent relative to the pre-conventional state of coordination failure. In particular, if an agent is assigned less than 44% of the time in the current game to the player position favoured by the convention, he will receive an average payoff less than that associated with the pre-conventional state. However, if this should happen it is *still* in his interest to conform to the established convention, as this convention still instantiates an evolutionary stable state of the population. Though conformity to an established convention may always be in the local interest of the participating agents, some of these may rightfully come to regret its establishment.

7.8 Conventions of Cooperation

The hawk dove game immediately prompts the question whether it would be possible for a population to reach the socially optimal state where everyone play dove. Unfortunately this state is undermined by a collective action problem similar to that of the famous Prisoners Dilemma game of Matrix 10.

		Player 1	
		cooperate	*deviate*
Player 2	*cooperate*	3,3	0,4
	deviate	4,0	**1,1**

Matrix 10: A Prisoners dilemma game

Though this game is not positing potential conventions, the possibility of the evolution of cooperation has been studied extensively within evolutionary game theory, resulting in conclusions of much interest to the theory of convention (see Axelrod 1984, Sugden 1989, Jiborn 1999, Skyrms 2004, Hansen 2006).

Contrary to the games hitherto analysed, the problem in the Prisoners dilemma is not equilibrium selection. Rather, it is reaching a state of *cooperation*, where the pursuit of individual interest threatens to undermine such. That is, agents face a collective action problem, because the state of cooperation is not a Nash equilibrium in itself. Yet this game is often invoked to analyse fundamental preconditions of such phenomena as tax-payment, gun control, property rights, restricted parking behaviour, self-serviced supermarkets; in general, behavioural patterns that seemingly presuppose normative expectations or institutions prescribing behaviour from which unilateral deviation enables the deviator to enjoy the benefits generated by general or near general conformity without attributing himself in their absence. Though the lack of multiple equilibria disqualifies analysis of the associated behavioural patterns as *contingent* threatening their status as conventions, they may be regarded as contingent in a *derivative* sense: particular kinds of such preconditions have been established in some social systems but not in others.

A standard approach to explaining cooperative behaviour has called on the necessary imposition of sanction systems external to the situation of interaction, in order to make cooperative behaviour a strictly dominant strategy (Matrix 11). To be specific, the claim is that sanctions change the individually perceived payoffs and hence transform the preference structure to individually rational cooperation (cf. Kavka 1983, Ostrom 1990).

		Player 1	
		cooperate	*deviate*
Player 2	*cooperate*	**3,3**	0,2
	deviate	2,0	−1,−1

Matrix 11: Coercion in the Prisoners dilemma game of Matrix 10, with sanction −2

This strategy has a notable precedent in Hobbes (1968) who interpreted the state as a *Leviathan* based on contract: an absolute sovereign established by everyone agreeing to confer all of their powers and rights to this common power, which thereby becomes strong enough to "tie them by fear of punishment to the performance of their covenants" (Hobbes, 1968, Ch. 17). However, besides raising the question of how sanction systems emerge in a pre-institutional or pre-normative state and how they are kept stable, it may be argued that such as explanatory strategy overlooks a fundamental feature of institutional reality: no sanction seems capable to mount the power necessary to bind its 'subjects' by *fear of punishment* alone to the performance of some of the most fundamental kinds of cooperative behaviour. This problem may be illustrated by observing what individuals actually do when such systems exist, but when they at the same time expect almost nobody to conform to their prescriptions. For instance, during the Los Angeles Riots in 1992, chaos broke out and crowds looted supermarkets, violated traffic rules, disregarded gun-control, property-rights and law and order in general.[4] Situations like these show that formal sanction systems may be powerless against the overwhelming force deposited in a population. Thus, the effect of sanction systems appears to be *conditional* on the individual expectation of general or near-general conformity. Hence, invoking their imposition does not suffice to account for how collective action or cooperation problems are solved.

What this means is that the standard interpretation of the effect of sanction systems (modelled in Matrix 11) has by and large been wrong. In particular, the type of Prisoners Dilemma may be challenged as an appropriate analogy for cooperation patterns. Instead, the appropriate game model may be argued to be that of the *Stag Hunt*. Thus, Hansen (2005) argues that this game is both appropriate if sanction systems exist (by annulling the effects of sanctions in the strategy profile of mutual deviation as in Matrix 12) or, if not, by iterating the Prisoners Dilemma game indefinitely in the *shadow of the future*, yielding a structurally similar game of Matrix 13 under suitable assumptions (see also Jiborn 1999).[5]

		Player 2	
		c	*d*
Player 1	*c*	**3,3**	0,2
	d	2,0	**1,1**

Matrix 12: The Stag Hunt Game resulting from annulling the effects of sanctions in the strategy profile of mutual deviation

[4] Another good example is the mutiny of the French army under the Nivelle offensive in WWI. Here more than 20,000 soldiers refused to attack enemy lines, leaving the officers in recognition of the impossibility of punishing entire divisions or implement harsh measures.

[5] Matrix 13 is taken from Hansen (2005, p. 90), who reaches this game from the standard Prisoners dilemma game of Matrix 3 by setting the shadow of the future at $\frac{1}{3}$ and following strategy of Skyrms (2004, p. 5) of categorising the infinitude of available strategies in the game under one of two ideal-types of '*trigger*' or '*reciprocal*' and '*all* d'. Thus, if Matrix 13 is to be read precisely strategy c refers to all strategies approximating so-called trigger or reciprocal strategies of the indefinitely repeated Prisoners dilemma of Matrix 10, while strategy d refer to all strategies of this game approximating the so-called '*all* d' strategy.

$$\begin{array}{r|c|c|} \multicolumn{1}{r}{} & \multicolumn{2}{c}{\text{Player 2}} \\ \multicolumn{1}{r}{} & \multicolumn{1}{c}{c} & \multicolumn{1}{c}{d} \\ \cline{2-3} \text{Player 1} \quad c & \mathbf{6,6} & 1,5 \\ \cline{2-3} d & 5,1 & \mathbf{2,2} \\ \cline{2-3} \end{array}$$

Matrix 13: The Stag Hunt Game resulting from iterating the PD-game indefinitely in the shadow of the future

This reinterpretation makes the basic problem of cooperative behaviour a true coordination problem. By comprising multiple equilibria it reveals the surprising fact that cooperation may ultimately be a matter of *contingency* in a *non-derivative* sense. If this is true, exploring cooperative behaviour from the perspective of the theory of convention may turn out to have profound implications for understanding the nature and dynamics of such behaviour.

Turning to an evolutionary analysis, then, it is known from the above analysis of the Stag Hunt game that the Pareto-efficient equilibrium corresponds to an evolutionary stable population state. However, from that analysis it is also known that the population state corresponding to the mixed strategy Nash equilibrium under the replicator dynamics forms an evolutionary non-stable separating point for the basins of attraction belonging to each of the pure-strategy Nash equilibria of the game, the other being the risk-dominant one. What this basically means is that if initial populations playing the stag hunt of Matrix 12 or 13 are formed at random, half of these will go to the payoff dominant state of universal cooperation, while the other half will be carried to the risk dominant state of universal defection.

On the one hand this is really good news. What have been provided by the evolutionary approach is the rudiments of an account of how contingent equilibrium behaviour can emerge and stabilise in the stag hunt game. That is, by analogy, the rudiments of an account for how conventions of cooperation may emerge and stabilise both for scenarios incorporating the existence of sanction systems and scenarios where such are absent, but where the shadow of the future is sufficiently large. On the other hand, it is crucial to notice that the basic assumption in the study of cooperation is that the initial social state is one of universal defection. Under this assumption, prospects of cooperation are still extremely poor. The dynamics of the stag hunt game of Matrix 12 and 13, for instance, reveals a strong pressure capturing the population in a state of universal defection: for cooperation to emerge in the first place, it is required that more than 50% of the population 'mutate' simultaneously by creativity or error from playing d to playing c (for a similar result see Skyrms 2004, pp. 11–12). This is the threshold problem.

Notice that up until this point it has been assumed that a game is played repeatedly between *randomly* correlated pairs of players within one or more large populations. However, this assumption is obviously at odds with the context of most social interaction. Individuals usually interact with certain other individuals with a higher frequency than with other individuals. One reason for this is that individuals are *spatially* located and hence tend to interact only or to a higher degree with those located nearby their location. This leads to the conjecture that some kind of spatial correlation may improve on the prospect for cooperation.

Brian Skyrms (2004) discusses the effects of local interaction in the stag hunt. Although Skyrms finds local interaction to improve somewhat the prospect for cooperation—a smaller fraction of c players than that prescribed in random correlation is stable or may spread when located next to each other—the threshold problem remains (see Skyrms 2004, Chapters 1 & 3). Skyrms' analysis is carried out in a local interaction model comprising a 100-by-100 square lattice where each square is occupied by a player playing the stag hunt with his *Moore (8) neighbourhood* (i.e., with her neighbours to the N, NE, E, SE, S, SW, W, NW) with the payoffs given in Matrix 12 except for mutual d yielding a payoff 2 to each agent (see Skyrms 2004, p. 32). The dynamics he chooses is the simplest case of an evolutionary imitation dynamics—imitate the best of your neighbours. In this model Skyrms finds that the population is carried to universal cooperation more often than under a best-reply dynamics as well as the replicator dynamics in a population with random correlation. To be specific, with the particular payoffs of the stag hunt chosen by Skyrms the dynamics carries the population to universal stag hunt when the fraction of cooperators in the initial population exceeds two-thirds, as compared to the three-fourths needed in random correlation under the same payoffs. Skyrms concludes that, "local interaction opens up possibilities of cooperation that do not exist in a more traditional setting, and that imitation dynamics is often more conducive to cooperation than best-response-dynamics."

However, according to Skyrms the 'imitate the best of your neighbours' dynamics is not the most realistic one. An 'imitate the strategy that performs best on average in your neighbourhood' is more realistic. Yet he chooses not to model this alternative. Hansen (2005) constructs a local interaction model based on this dynamics with the following two specifications: (1) an agent only imitates the strategy that performs best on average in his neighbourhood if this did better in the last round than he himself did,[6] and (2) the stag hunt game played has the payoffs of the game in Matrix 11. Each agent is then taken to occupy a cell in a $n \times n$ lattice, where n is large (technically infinite) and playing the game with his Moore (8) Neighbourhood. Figure 2 reveals the result: given that 6 agents in adjacent cells forming a 2×3 square mutate so as to play c in the stag hunt, the 'imitate the strategy that performs best on average in your neighbourhood' dynamics leads to universal cooperation.

To be sure, this result does not obtain for a similar local interaction model playing a stag hunt with the payoffs chosen by Skyrms (2004, Ch. 3). However, as evolutionary dynamics are payoff sensitive so is Skyrms'. Thus, if enhancing the payoffs resulting from joint stag hunt in Skyrms' version to $4\frac{1}{2}$ and mutating 9 agents in adjacent cells forming a 3×3 square so as to play c in the stag hunt the same result obtains: stag hunt spreads so as to invade the whole of the population (Hansen, 2005). One might question the credibility of such models by asking what the possibilities are for the necessary initial configurations of cooperative agents appearing in the model. Answering this on an analytical level is not as difficult as it may seem. If the lattice is infinite and the game repeated indefinitely with a low mutation rate it may be argued that it is quite likely that at some point these configurations will appear. How often depends on the chosen mutation rate. In particular, as pointed out earlier, the very low mutation rates usually assumed in evolutionary models may be questionable for populations of human agents.

7.9 Closure

What has been shown is the analytical success of evolutionary game theory in solving the Nash equilibrium selection problem in a variety of simple but paradigmatic games of convention. Such analytical success, however, need to be evaluated and qualified in terms of a credible interpretation of what goes on in complex real world social interaction, if only to restrain any premature enthusiasm similar to that with which von Neumann and Morgenstern's *Theory of Games*

[6] In case of ties it is assumed that the agent keeps his strategy. This assumption turns out to slow down the spread of conditional cooperation, why the result does not depend on it.

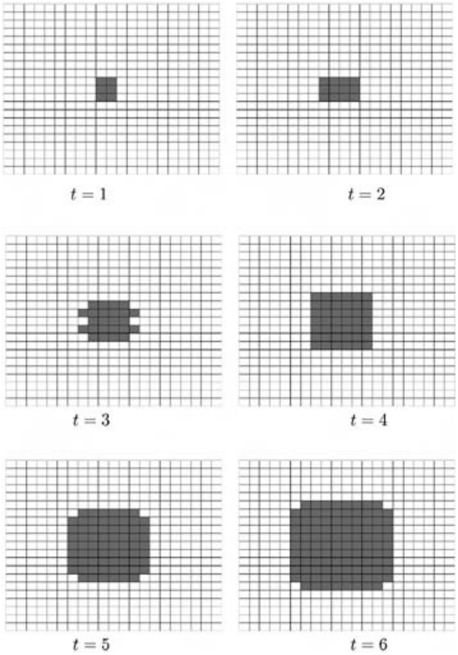


Figure 2: Cooperative strategies invading a 'defectionist' population state in a local interaction model

and Economic Behavior was met. Not that one should reject being enthusiastic about von Neumann and Morgenstern's great accomplishment through the provision of classical game theory. But as was the case with classical game theory in relation to strategic behaviour, there are significant reasons to assert that evolutionary game theory is very far from delivering a comprehensive theory of the nature and dynamics of social conventions and associated behaviour.

Most importantly, the features allowing the framework of evolutionary game theory to explicitly model the population dynamics of strategies poses a profound problem when applied to human social behaviour. To see this, one may begin by noticing the simple and basic intuition behind the evolutionary approach (Axelrod, 1984, p. 169): "whatever is successful is likely to appear more often in the future." In the biological application of evolutionary game theory the corresponding principle works through genetical heredity and differential reproductive success. However, the application of evolutionary game theory makes for asking how the concept of payoffs representing *reproductive success* in biological applications should be interpreted in the context of human social behaviour. In particular, contrary to the expected payoffs of classical game theory (von Neumann & Morgenstern subjective utilities), evolutionary game theory generally presumes that payoffs are interpersonally comparable. This is especially clear as the growth rate of a strategy is defined as a function of its *average payoff* such as in the replicator dynamics. In order to calculate this average there must be some natural way of comparing the payoffs to each agent following a strategy in some particular state of the population.

Besides posing serious theoretical problems of interpretation, this feature of the evolutionary approach yields intractable problems even for pragmatic inquires concerning social conventions. For some of the examples, the presumption that interpersonal comparisons are meaningful do not distort the relevant argument. For instance, in the telephone game the interest is in the possible effects of player positions rather than the particular dynamics. However, for other examples this is not necessarily true. For instance, if studying the phenomenon of the division of labour or property rights where the assignment of player-positions cease to be perfectly cross-cutting, it should be remembered that in many cases real world individuals belonging to disfavoured groups may devalue or even cease to desire what they perceive as unattainable due to the mechanism of *cognitive dissonance reduction* (Elster, 1989, p. 4), or, alternatively, may 'fall prey' to the phenomenon of *relative frustration* appearing if people quite reasonably come to reject the legitimacy of a convention assigning them less of some good relative to other groups merely due to some arbitrary feature.

On the practical level such phenomena seriously distort the perceived easiness by which payoffs may be thought to be attributed to agents of social interaction, whereby the postulated dynamics are rendered highly questionable. Ultimately, this pushes the theory of convention in the midst of a current battle between behavioural game-theorists on the one hand and analytic, classical and evolutionary game-theorists on the other. One crucial issue here is whether the payoff functions of human agents are fundamentally effected by such phenomena as relative frustration or other fairness considerations as claimed by the former, or whether such phenomena are just temporary responses of disequilibrium behaviour that gradually adapts to the real and overwhelming forces of the underlying dynamics (see e.g. Camerer 2003, Fehr et al. 2004 and Binmore 2005). Caution is recommended as the word is still out even on the most basic assumptions utilised in re-erecting a theory of convention within an evolutionary framework.

REFERENCES

Aumann, R. (1976). Agreeing to disagree. *Annals of Statistics*, **4**, 1236-1239.

Axelrod, R. (1986). *The Evolution of Cooperation*. Basic Books.

Bacharach, M. (1993). Variable universe games. In: *Frontiers of Game Theory* (K. Binmore, A. Kirman and P. Tani, eds.), MIT Press, Cambridge, Mass.

Binmore, K. (1990). *Essays on the Foundations of Game Theory*. Blackwell, Oxford.

Binmore, K. (1990). *Fun and Games: A Text on Game Theory*. D. C. Heath, Lexington.

Binmore, K. (1994). *Game Theory and the Social Contract: Playing Fair*. MIT Press, Cambridge, Mass.

Binmore, K. (1995). Foreword. In: *Evolutionary Game Theory* (J. W. Weibull). MIT Press, Cambridge, Mass.

Binmore, K. (2005). *Natural Justice.* Oxford University Press, New York.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press, Princeton.

Colman, A. M. (1997). Salience and focusing in pure coordination games. *Journal of Economic Methodology*, **4**, 61-81.

Colman, A. M. (1999). *Game Theory and Its Applications in the Social and Biological Sciences* (2nd ed.). Routledge, London.

Crawford, V. and H. Haller (1990). Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica*, **6**, 389-403.

Cubitt, R. P. and R. Sugden (2003). Common knowledge, salience and convention: A reconstruction of David Lewis' game theory. *Economics and Philosophy*, **19**, 175-210.

Elster, J. (1989). *Nots and Bolts for the Social Sciences.* Cambridge University Press, New York.

Fehr, E., et al. (2004). *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies.* Oxford University Press, Oxford.

Gauthier, D. (1975). Coordination. *Dialogue*, **14**, 195-221.

Gilbert, M. (1989). Rationality and salience. *Philosophical Studies*, **57**, 61-77.

Hansen, P. G. (2005). Exploring the nature and dynamics of social conventions. http://ruc.dk/~pgh.

Hansen, P. G. (2006). Towards a theory of convention. *Phinews*, **9**, 30-62.

Harsanyi, J. C. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games.* MIT Press, Cambridge, Mass.

Heal, J. (1978). Common knowledge. *Philosophical Quarterly*, **28**, 116-131.

Hobbes, T. (1968/1651). *Leviathan.* Penguin Books, London.

Janssen, M. W. (1995). *Rationalizing focal points.* Tinbergen Institute Discussion Paper, Rotterdam.

Janssen, M. W. (1998). Focal points. In: *The New Palgrave: A Dictionary of Economics*, **12**, pp. 150-155. Macmillan, London.

Jiborn, M. (1999). *Voluntary Coercion: Collective Action and the Social Contract.* Doctoral dissertation, Lund University.

Kavka, G. S. (1983). Hobbes's war of all against all. In: *The Social Contract Theorists: Critical Essays on Hobbes, Locke and Rousseau* (C. W. Morris, ed.), Rowan & Littlefield, Maryland.

Kincaid, P. (1986). *The Rule of the Road: An International Guide to History and Practice.* Greenwood Press, New York.

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Blackwell, Oxford.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.

Maynard Smith, J. and G. R. Price (1973). The logic of animal conflict. In: *Nature*, **246**, 15-18.

Metha, J., C. Starmer and R. Sugden (1994a). The nature of salience: An experimental investigation of pure-coordination games. *The American Economic Review*, **74**, 658-673.

Metha, J., C. Starmer and R. Sugden (1994b). Focal points in pure-coordination games: An experimental investigation. *Theory and Decision*, **36**, 163-185.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge, Mass.

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press, Cambridge, Mass.

Schuster, P. and K. Sigmund (1983). Replicator dynamics. *Journal of Theoretical Biology*, **100**, 533-538.

Skyrms, B. (1996). *Evolution and The Social Contract*. Cambridge University Press, Cambridge.

Skyrms, B. (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, Cambridge.

Sugden, R. (1986). *The Economics of Rights, Cooperation and Welfare*. London: Palgrave Macmillan.

Sugden, R. (1989). Spontaneous order. *Journal of Economic Perspectives*, **3**, 85-97.

Sugden, R. (1991). Rational choice: A survey of contributions from economics and philosophy. *Economic Journal*, **101**, 751-785.

Sugden, R. (1999). Conventions. In: *New Palgrave Dictionary of Economics and Law*, pp. 453-460. Macmillan, London.

Sugden, R. (2000). The motivating power of expectations. In: *Rationality, Rules, and Structure* (J. Nida-Rümelin and W. Spohn, eds.), Kluwer, Dordrecht.

Sugden, R. (2001). The evolutionary turn in game theory. *Journal of Economic Methodology*, **8**, 113-130.

Taylor, P. D. and L. B. Jonker (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, **40**, 145-156.

Ullman-Margalit (1977). *The Emergence of Norms*. Clarendon Press, Oxford.

Weibull, J. W. (1995). *Evolutionary Game Theory*. MIT Press, Cambridge, Mass.

Young, H. P. (1991). The economics of convention. *The Journal of Economic Perspectives*, **10**, 105-122.

Young, H. P. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton.

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 6

EVOLUTIONARY MODELS OF LANGUAGE

Cecilia Di Chio
University of Essex

Paolo Di Chio
University of L'Aquila

This paper deals with the evolutionary theory of games, and in particular the theory of evolutionary language games, a discipline which arose from the union of evolutionary game theory and language games. After giving an overview of the historical background, we will provide a review of some of the key works on evolutionary language games. We will then propose some simulation models for the evolution of language which aim (i) to verify previous results and (ii) to show how the presence of a topological structure influences the communication among individuals.

1 INTRODUCTION

The linguistic system appears to follow an evolutionary trajectory parallel to the genetic one (i.e., they co-evolve, see Cavalli-Sforza 2001). Isolation, either social or geographic, causes evolution and genetic differentiation to occur independently from one another. The same happens with languages: isolation reduces cultural exchanges and languages of isolated populations become more and more differentiated. The study of the emergence of these isolated clusters of languages has been the motivation for our research.

The subject of this paper is the theory of evolutionary language games, which is derived from two disciplines that were originally unrelated: evolutionary game theory and language games. We will give a brief insight into these two topics in Section 2.

In Section 4 we present two multi-agent simulation models to study the evolution of languages, based on (two player) evolutionary language games. The first model proposed (Section 4.1) is based on a mathematical model of Nowak & Krakauer (1999), Nowak et al. (1999) and Nowak (2000) and is designed to reproduce and verify (or refute) the results obtained in the simplest mathematical model. The second model (Section 4.2), again inspired by Nowak's work, extends the authors' first model with the introduction of a significant characteristic: a world where the languages live and evolve, and which influences interactions among individuals. The goal of this second simulation is to show how the presence of a topological structure influences the communication among individuals and contributes to the emergence of clusters of different languages.

Even though our models are largely based on Nowak's work, we should bear in mind that there have been many other models for the evolution of language (see the review papers by Kirby 2002b, Steels 1997a and Turner 2002). Section 3 summarises three such models.

2 HISTORICAL BACKGROUND

This research focusses on the evolutionary theory of games, and in particular on the theory of evolutionary language games, studied with the use of discrete simulation models. Evolutionary game theory was introduced by British biologist John Maynard Smith (1920–2004), whilst the idea of language games was developed by Austrian philosopher Ludwig Wittgenstein (1889–1951). The theory of evolutionary language games arose from the merger of these two disciplines.

2.1 MAYNARD SMITH'S EVOLUTIONARY GAME THEORY

According to Pinker & Bloom (1990), the ability for humans to learn languages is a product of natural selection. Therefore, genetic evolution can be considered to be the main reason for the origin and emergence of language in human beings.

Mathematical optimisation is the most appropriate tool when we want to understand why natural selection has preferred some features more than others. The theory of games is preferable when it is important to know the interactive behaviour of all the individuals in the population. The passage from classic to evolutionary game theory happens when the individuals learn, adapt and evolve over evolutionary time.

Evolutionary game theory, introduced in Maynard Smith (1982) and Maynard Smith & Price (1973), is a way to think about evolution from a phenotypic point of view, where the fitness (i.e., the ability to prevail) of certain phenotypes depends on how frequent those phenotypes are in the population. Evolutionary game theory is of fundamental importance when studying evolving individuals which can dynamically learn and adapt themselves to the environment (Hobauer & Sigmund, 1998).

The main differences between classical and evolutionary game theory are in the variations on the concepts of strategy, equilibrium and interactions among players or agents (Maynard Smith & Price, 1973):

Strategies: The set of strategies is replaced by the set of *genotypes* (in biology) or *cultural form* (human society); individuals "inherit" or "choose" variations from these sets;

Equilibrium: The Nash equilibrium of classic game theory is substituted with the concept of an *evolutionary stable strategy*: a strategy is evolutionary stable if the population/society that use it cannot be invaded by a different group with a different genotype/alternative cultural form;

Interactions: Players are coupled repeatedly and randomly, and play according to the strategies based on their genomes but typically not on the past history of the game.

2.2 Wittgenstein's Language Games

According to Ludwig Wittgenstein's *Philosophical Investigations* (1953, §7):

> We can also think of the whole process of using words in [an elementary language] as one of those games by means of which children learn their native language. I will call these games 'language games' and will sometimes speak of a primitive language as a language-game.
> And the process of naming the stones and of repeating words after someone might also be calling language-games. [...]
> I shall also call the whole, consisting of language and the actions into which it it woven, a 'language-game'.

Language games, as presented by Wittgenstein (1953), are regarded as involving both a language and the actions required to deal with it. They can be seen as the process of using words by which children learn their native language. Through (the use of) the game, the words of the language get their meaning, which is seen as the purpose of those words. The words are not held *to refer to objects*, but defined *through the ways they are used* in the context. This view of the meaning of a word as its 'use' contrasts with the classical interpretation of meaning as 'representation'.

When first presented the idea, Wittgenstein did not consider language games to have any evolutionary aspect. However, it is possible to assign an evolutionary interpretation to language games by defining a parallel between genetic and linguistic evolution: here we consider languages as species and the rules that characterise the languages as genes.

3 The State of the Art

As noted in Section 2, evolution (both genetic and linguistic) can be studied by means of game theory, where a game is an interaction either between players or between a player and the environment. In the literature on the evolution of communication systems, the combination of evolutionary game theory and language games (i.e., evolutionary language games) has been applied in quite different contexts. We present three of the most significant examples of these applications.

3.1 Kirby's Computer Simulations

The basic idea in Kirby (2000) and Kirby & Hurford (2002) is to consider a language as the result of the intersection of three different complex adaptive systems:

Individual learning: Children adapt their knowledge of a language in response to the environment;

Linguistic evolution: Languages change over time;

Biological evolution: The learning mechanism adapt in response to selection pressure from the environment.

Given the variety of systems involved in the emergence of a language, it is hard to understand the interactions among these three systems, and it is not clear which one is the best methodology to study the evolution of language. To solve this situation, Kirby proposes a model to study the process by which learning is transmitted across generations (Iterated Learning Model, ILM; see Kirby 2000 and Kirby & Hurford 2002).

According to these proposals, each language has two representations, internal (*I-language*) and external (*E-language*), and is transmitted from one generation to the other through *use* (from internal to external) and through *learning* (from external to internal). According to Kirby (2002a), these transformations act as a bottleneck on the transmission of a language over time. Kirby's model represents the structure of correspondences between meanings and signals and vice versa, which does not have to change despite the bottleneck.

The simulations are initialised without having any former linguistic systems, in other words the adult agents need not have any I-language and the population need not have any E-language.

At each iteration of the model, an adult agent emits some signals corresponding to a set of given meanings. The resulting pairs (meaning-signal) represents the pool of data from which the learning agents learn (i.e., their E-language). After a learning period, the learning agents create their own individual I-language (i.e., they become adults). Some new learning agents are introduced to the population, and some of the 'old' adults are removed to keep the population size constant.

The result of the simulation is the emergence of a linguistic system, which is stable and expressive. Stability (i.e., how much the language of the learner differs from the language of the adult) and expressivity (i.e., the proportion of the space of the meanings covered by the signals) vary according to the size of the learning pool.

3.2 Steels' Robotic Agents

Steels (1998, 1997b), using robotic agents together with software simulation ('Talking Heads'), analyses the process of the evolution of language through the theory of evolutionary language games.

The purpose of the experiments is to prove that the mechanisms that generate complexity in biological systems (i.e., evolution, co-evolution, self-organisation, and level formation) can also be used to explain the evolution of complexity in language. The hypotheses are that languages (i) are an emergent mass phenomenon that happen through the interaction among individuals, (ii) are not completely known or controlled by an individual, and (iii) emerge spontaneously once some physical, psychological and social conditions are satisfied.

Steels defines different kinds of language games according to different aspects of evolution:

Discrimination games: Discrimination games create the meanings of words;

Linguistic games: Linguistic games determine the formation of the lexicon;

Imitation games: Imitation games evolve phonology (i.e., the repertoire of phonemes that characterise the language).

The second mechanism here refers to co-evolution. Linguistic games require that the meanings created via discrimination games are distinct enough (not to be confused with one another) and that the lexicon is able to describe all of these meanings: the higher the number of meanings to be used in the game, the larger the size of the lexicon.

Self-organisation arises when a certain number of equally good linguistic structures exists, but only one of them is selected and adopted by the population (as we have already noted, there is no centralised control over the agents). Self-organisation can happen only when random mutation becomes predominant.

The last key point is the formation of level that emerges in biology when a certain number of independent entities develop symbiotic relationships. From a linguistic point of view, the formation of levels justifies the emergence of the syntax of a language.

3.3 NOWAK'S MATHEMATICAL MODELS

The mathematical model of Nowak & Krakauer (1999), Nowak et al. (1999) and Nowak (2000) is concerned with how *proto-languages* emerge in a non-linguistic society, and how a specific signal could be associated with a specific object. Assuming that languages evolve through communication, the basic evolutionary language game underlying this model consists of two individuals that emit a certain number of sounds to exchange information about a certain number of objects. In this model, an individual can be interpreted as the language it speaks and vice versa.

Suppose that there are n objects and m sounds (in the following, we will assume $n = m$). The set of pairs $\langle$*object, sound*$\rangle$ is the *vocabulary* on which the language is defined. The language $\mathcal{L}$ of each individual is defined by an *association matrix* $\mathcal{A}(n \times m)$ the entries $a_{i,j}$ of which are non-negative real values that represent the strength of association between the object i and the sound j (i.e., how often the individuals have referred to the object i by producing the sound j). Each individual acts both as a speaker and a listener:

Speakers: Speakers are characterised by the *active matrix* $\mathcal{P}(n \times m)$, the entries $p_{i,j}$ of which representing the probability that the object i is associated with the sound j;

Listeners: Listeners are characterised by the *passive matrix* $\mathcal{Q}(m \times n)$, the entries $q_{j,i}$ of which representing the probability that the sound j is associated with the object i.

Entries in $\mathcal{P}$ and $\mathcal{Q}$ are derived from those in $\mathcal{A}$ through normalisation on rows and columns, respectively:

$$p_{i,j} = \frac{a_{i,j}}{\sum_{j=1}^{m} a_{i,j}} \qquad q_{j,i} = \frac{a_{i,j}}{\sum_{i=1}^{n} a_{i,j}} . \tag{1}$$

Let us consider two individuals (i.e., *players*), l and l', who speak language $\mathcal{L}$ and $\mathcal{L}'$ (with $\mathcal{L} \neq \mathcal{L}'$). The language game is defined as follows. l sees an object i and emits a sound; the probability that l' infers the same object i is given by:

$$\sum_{j=1}^{m} p_{i,j} q'_{j,i} \qquad p_{i,j} \in \mathcal{P},\ q'_{j,i} \in \mathcal{Q}' . \tag{2}$$

The ability of l to communicate with l' is the sum, over each of the n objects, of all these probabilities.

The payoff of the game between the two individuals is the average between the ability of l to communicate with l', and the ability of l' to communicate with l. In other words:

$$F(\mathcal{L}, \mathcal{L}') = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} (p_{i,j} q'_{j,i} + p'_{i,j} q_{j,i}) . \tag{3}$$

Since the individuals are both speakers and listeners, the game is symmetric, that is, $F(\mathcal{L}, \mathcal{L}') = F(\mathcal{L}', \mathcal{L})$. Each player's total payoff will be the sum (over all possible player pairs) of that player's individual payoffs.

Based on one of the main assumptions of evolutionary game theory, the payoff of each individual represents its fitness. Therefore, each individual generates a number of offspring proportional to its payoff (as a fraction of the total payoff). The set of new individuals completely replaces the old generation. That is, the population size is constant over time.

Each individual in the new generation will *learn* its language by sampling its parent's active matrix $\mathcal{P}$ (observing the answers its parent gives to refer to specific objects), giving rise to the sampling process:

$$\mathcal{A}_0 \xrightarrow{(1)} \mathcal{P}_0 \xrightarrow{k} \mathcal{A}_1 \longrightarrow \mathcal{P}_1 \longrightarrow \cdots , \tag{4}$$

where k is the number of samples each offspring makes on its parent's active matrix, namely the number of answers that the new individual observes.

4 Simulation Models

Based on the assumptions of Nowak's mathematical model, we now propose two agent-based discrete simulation models (see also Di Chio & Di Chio 2006a,b).

The first of these models aims to reproduce and verify Nowak's results. The process simulated is the one in which, starting from a population of (individuals who speak) different languages, the result is a single language spoken by the whole population. This final language emerges and survives because its adaptive behaviour is superior to its rivals'. In other words, the individuals who speak that language have a higher fitness than the ones speaking different languages, and the latter will therefore disappear through evolutionary time.

The second model to be proposed is an extension of the first and is aimed to assist in studying the environment's influence on the evolution of languages and the interactions among individuals. Nowak's mathematical model describes quite accurately the emergence of a linguistic system but, at the same time, it is grounded on some simplifying assumptions. In particular, there is no environment able to influence the communication among the individuals. Since isolation is one of the main reasons for the differentiation of languages and the emergence of linguistic groups, we develop a simulation model adding to Nowak's *world*: an environment with a topological structure in which the (individuals who speak the) languages live and evolve.

In both models and also in Nowak's work, there is no real distinction between a language and the individual who speaks that language. Therefore, each agent in the simulation represents both an individual and its language. More details on the implementation of both models are given in Di Chio (2004).

4.1 The Basic Model

Following Nowak's notation, each language is defined by an association matrix $\mathcal{A}(n \times m)$ with non-negative entries, selected at random for the first generation and through a sampling function for the next generations (Equation 4). The size of the population is constant in time and each new generation completely replaces the old one.

The fitness is calculated as the sum of partial payoffs that each individual gains playing the language game (Equation 3). The number of offspring for each agent is proportional to its fitness.

Since this model has been designed according to the mathematical model, we expect a similar behaviour to that in Nowak's: the emergence of a single, (possibly) optimum language, that is a language in which each sound is associated with a single object and vice versa.

4.2 Adding the World

We now extend the previous model by placing the agents into an environment. The world where the agents live (in fixed *cell* locations) is a 2-D discrete grid topologically equivalent to a torus and the x and y dimensions of which are exogenous parameters. Agents produce offspring that will be generated and put into the environment according to a certain set of rules.

We rewrite the payoffs of the language game between agents a_h and a_k (see Equation 3) as

$$\pi(a_h, a_k) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(p_{i,j}q'_{j,i} + p'_{i,j}q_{j,i})\,. \tag{5}$$

The similarities with the mathematical model end here. The computation of fitness, the generation of offspring, and the positioning of the newborn agents now take into account the presence of the world. The fitness function is modified in such a way that the contribution to the fitness of the agent a_h is higher for closer individuals it plays with. This mirrors a real-world situation, where communication is more likely to happen between individuals that are closest to each other (using some suitable metric). The number of offspring that each agent generates is still proportional to the agent's fitness, but the factor of proportionality is no longer the global fitness (the fitness of the whole population) but a 'locally global' fitness (i.e., the fitness of a suitable neighbourhood of the generating agent). To avoid too abrupt a separation among agents, we adopt a parameter of fuzziness in the definition of the neighbourhood, weighting the fitness of pairs of agents with a smooth function.

We position the newly generated agents by putting the offspring either in *neighbouring* cells or in (a *list* in) the same cell as the parent. These strategies have been chosen to mirror a more or less strong isolation process (respectively).

More formally, let $d(a_h, a_k)$ be the Euclidean distance between the agents a_h and a_k, and let $\rho(a_h, a_k) = e^{-d(a_h, a_k)}$ be the function of d we will use to weight the payoffs. The fitness ϕ for a_h is given by:

$$\phi_{a_h} = \sum_{k \neq h} \pi(a_h, a_k)\rho(a_h, a_k)\,. \tag{6}$$

To compute the number of offspring, we have to take into account the 'locally global' fitness. If $\Phi(a_h)$ is the global fitness relevant to the individual a_h, and A_h is a suitable neighbourhood of a_h, we have:

$$\Phi(a_h) = \sum_{a_k \in A_h} \phi_{a_k}\,. \tag{7}$$

The number of offspring s_{a_h} for a_h is proportional to the ratio between the individual's fitness and the global fitness, that is:

$$s_{a_h} = n_{A_h}\frac{\phi_{a_h}}{\Phi(a_h)} \propto \frac{\phi_{a_h}}{\Phi(a_h)}, \tag{8}$$

where n_{A_h} is the number of agents in A_h. We do not know A_h, but we can 'fuzzify' it and write, for the global fitness:

$$\Phi(a_h) = \sum_k \phi_{a_k} \rho(a_h, a_k), \tag{9}$$

and for n_{A_h}:

$$\tilde{n}_{A_h} = \sum_{k \neq h} \rho(a_h, a_k). \tag{10}$$

In each generation, the population size N is constant. Therefore, we have:

$$\sum_h s_{a_h} = \sum_h n_{A_h} \frac{\phi_{a_h}}{\Phi(a_h)} = N \qquad \sum_h \tilde{n}_{A_h} \frac{\phi_{a_h}}{\Phi(a_h)} = M. \tag{11}$$

Thus, to retain the population size N per generation, the actual number of offspring for each individual is given by:

$$s_{a_h} = \frac{N}{M} \sum_{k \neq h} \rho(a_h, a_k), \tag{12}$$

where N/M is a normalisation factor.

At each generation, the offspring of the same language will be close to each other, their fitnesses will be higher, and they will leave more offspring. This is a phenomenon which happens locally and therefore we expect to observe the process of language clustering.

Starting from a population of many different languages (i.e., from many different populations, each one made of just one language), the simulation shows how these languages spontaneously move (closer or further away) until independent populations emerge. This happens without any form of 'artificial' constraint: it is just due to communication.

4.3 Results

We have implemented the models on the *Swarm*[1] platform with the OBJECTIVE-C programming language.[2]

The various settings for the parameters of the simulation runs are summarised in Table 1.

Parameter	Value
(objects, sounds)	(5, 5), (10, 10), (25, 25)
Population size	100 individuals
Sampling parameter k	1, 4, 7, 10, 25
Generations	100
Iterations	20

Table 1: Parameter settings

[1] More information can be found at www.swarm.org.

[2] Both simulations have been run on a 2.4GHz INTEL PENTIUM 4® CPU with 512MB RAM with the REDHAT® LINUX 9.0 operating system.

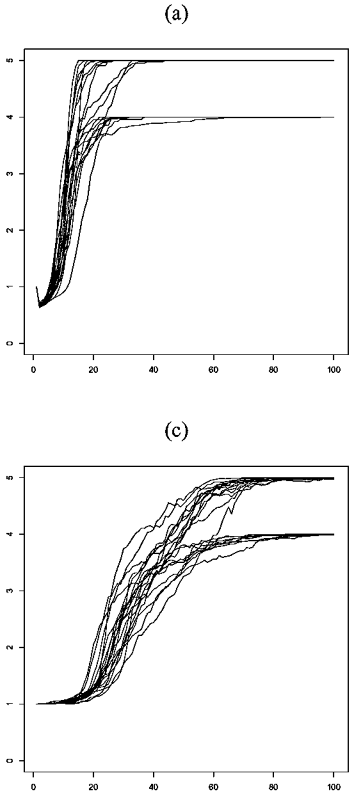


Figure 1: Fitness values. Simulation model with (a) (objects, sounds) $= (5,5)$ and $k = 1$. (b) (objects, sounds) $= (10,10)$ and $k = 1$. (c) (objects, sounds) $= (5,5)$ and $k = 25$. (d) (objects, sounds) $= (25,25)$ and $k = 25$

The graphs in Figure 1 show some results for the first model. A comparison of these fitness curves with those obtained in Nowak et al. (1999) shows a clear correspondence (especially qualitatively).

As one can observe from the figures, the behaviour of our simulation model is almost identical to that of Nowak's mathematical one.

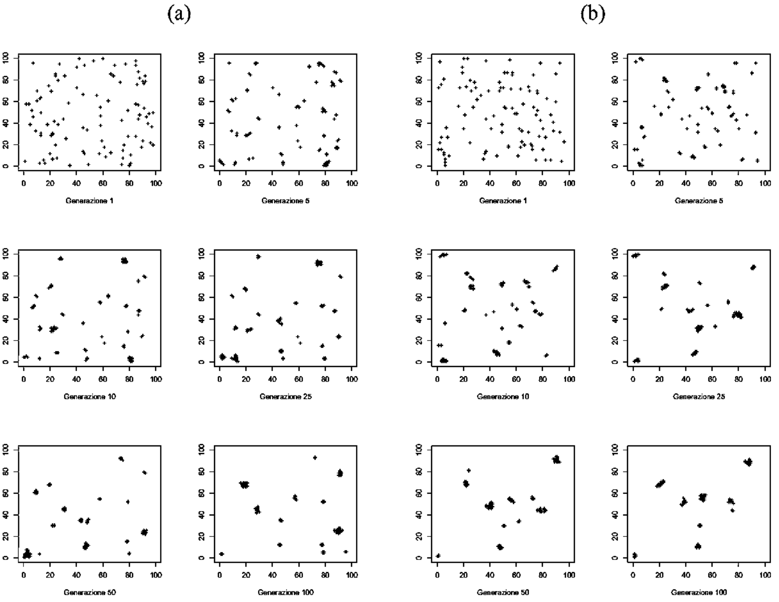


Figure 2: Simulation model with (objects, sounds) $= (5,5)$. (a) $k = 1$. (b) $k = 25$. Distance influences both ϕ and Φ. Population replaced with neighbourhood lookup

For the second model, we have run different experiments according to the positioning of the offspring in the world (neighbouring cells lookup or list) and according to the fitness weighted with the distance. The graphs in Figure 2 show the results of the simulations (with the smallest vocabulary and sampling parameter values 1 and 25) when the distance influences both the individual's fitness ϕ and the locally global fitness Φ, and the population is replaced with neighbourhood lookup.

Figure 3 shows the results of the simulations (with the same parameters as before) when the distance influences both the individual's fitness ϕ and the locally global fitness Φ, and the new population is positioned in a list.

The last two graphs (Figure 4) show the configuration of clusters in detail. In particular, we can observe that, if the replacement is with neighborhood lookup, it is possible to have clusters with more than one language, whilst if the population is positioned in a list, there is just one language in each cell.

As the simulation results show, it is obvious how important the presence of a topological structure is for the behaviour of the languages. We can, in fact, observe, by varying the parameters, the emergence of different clusters of different languages. The replacement with neighbourhood lookup causes the clusters to continually evolve. This happens because, by positioning the new individuals in the cells around their parents, the dimensions of the cluster are continuously

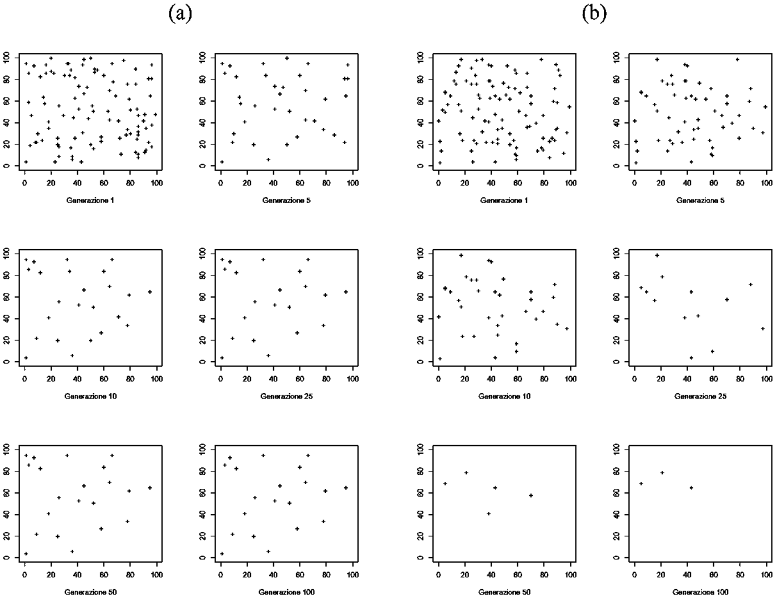


Figure 3: Simulation model with (objects, sounds) = (5, 5). (a) $k = 1$. (b) $k = 25$. Distance influences both ϕ and Φ. Population positioned in a list

varying, and therefore the distance among individuals in different clusters changes from one generation to the next. These variations help in the emergence of new languages in new positions (i.e., positions different from the starting ones). On the other hand, positioning the new population in lists provides a way to highlight the process of cluster creation. Since all the offspring of an individual are placed in the same cell, the spatial dimensions of each cluster are constant (and equal to one cell). Therefore, in this situation, we will not observe the emergence of new languages in new positions, but only the disappearance of isolated (weaker) languages.

For a more comprehensive set of graphs for both the simulation models as well as a complete list of clusters and their characteristics we refer to Di Chio (2004).

5 Conclusion

The goal of this research was to present a simulation model based on the theory of evolutionary linguistic games to study the emergence of languages.

First, we presented some historical background of this theory: evolutionary game theory and language games. Second, we described some applications of evolutionary language games in dif-

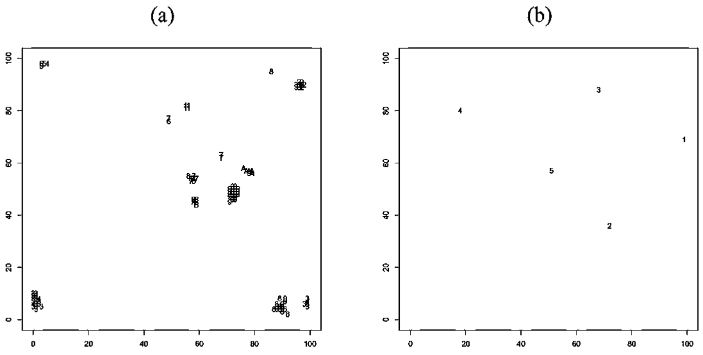


Figure 4: Simulation model with (objects, sounds) = $(5, 5)$ and $k = 7$. (a) Neighbourhood lookup, 11 languages and 14 clusters. (b) List, 5 languages and 5 clusters

ferent disciplines, namely computer simulations, robotic agents and mathematical models. Third, we proposed our agent-based simulation models for the study of the evolution of languages. The first of them is inspired by Nowak's mathematical model and it was seen to verify his results. We then used this simpler model to build a more refined simulation to study how clusters of different languages emerge and evolve in the world, due to the influence of the environment on the communication among individuals.

The results presented here have (i) confirmed Nowak's hypothesis: the emergence of a common linguistic system starting from a population of individuals speaking their own different and unrelated languages, and (ii) shown the emergence of different language configurations, according to the parameters acting on the system (e.g., the influence of the environment on the offspring generation and the way that the new languages are introduced to the world).

There are a number of interesting future directions we would like to explore, such as (i) to allow multiple parents and overlapping generations (population size no longer constant), (ii) to separate the individual from the language, allowing an individual to speak more than just one language, (iii) to study other linguistic phenomena such as dialects or Pidgin/Creole, and (iv) to expand our model to let agents move around (like in a particle swarm system).

References

Cavalli-Sforza, L. L. (2001). *People, Genes and Languages*. University of California Press, Berkeley.

Di Chio, C. (2004). *Modelli di Simulazione Evolutiva per lo Sviluppo del Linguaggio*. Tesi di Laurea, University of Roma "La Sapienza".

Di Chio, C. and P. Di Chio (2006a). A simple simulation model for the evolution of language. Manuscript.

Di Chio, C. and P. Di Chio (2006b). Simulation model for the evolution of language with spatial topology. In: *Proceedings of the 6th International Conference on the Evolution of Language*, pp. 51-58.

Hofbauer, J. and K. Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, Mass.

Kirby, S. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In: *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 303-323. Cambridge University Press, Cambridge, Mass.

Kirby, S. (2002a). Learning, bottlenecks and the evolution of recursive syntax. In: *Linguistic Evolution through Language Acquisition: Formal and Computational Models* (T. Briscoe, ed.). Cambridge University Press, Cambridge, Mass.

Kirby, S. (2002b). Natural language from artificial life. *Artificial Life*, **8**, 1185-1215.

Kirby, S. and J. Hurford (2002). The emergence of linguistic structure: An overview of the iterated learning model. In: *Simulating the Evolution of Language* (A. Cangelosi and D. Parisi, eds.), pp. 121-148. Springer, Berlin.

Maynard Smith, J. and G. R. Price (1973). The logic of animal conflict. *Nature*, **246**, 15-18.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, Mass.

Nowak, M. (2000). Evolutionary biology of language. *Philosophical Transactions Royal Society B: Biological Sciences*, **355**, 1615-1622.

Nowak, M. and D. Krakauer (1999). The evolution of language. In: *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 8028-8033.

Nowak, M., J. Plotkin and D. Krakauer (1999). The evolutionary language game. *Journal of Theoretical Biology*, **200**, 147-162.

Pinker, S. and P. Bloom (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, **13**, 707-784.

Steels, L. (1997a). The synthetic modeling of language origins. *Evolution of Communication*, **1**, 1-34.

Steels, L. (1997b). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In: *Approaches to the Evolution of Language: Social and Cognitive Bases* (J. Hurford, C. Knight and M. Studdert-Kennedy, eds.). Edinburgh University Press, Edinburgh.

Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, **103**, 133-156.

Turner, H. (2002). An introduction to methods for simulating the evolution of language. In: *Simulating the Evolution of Language* (A. Cangelosi and D. Parisi, eds.), pp. 29-50. Springer, Berlin.

Wittgenstein, L. (1953). *Philosophical Investigations*, Blackwell, Oxford.

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 7

GAME DYNAMICS CONNECTS SEMANTICS AND PRAGMATICS

Gerhard Jäger
University of Bielefeld

The chapter first gives an overview over the evolutionary interpretation of game theory, and compares two versions of it, the replicator dynamics and the best response dynamics. The ensuing notions of evolutionary stability are explored. In the second part, it is argued that the best response dynamics lends itself to an epistemic interpretation, and that this provides a suitable game-theoretic foundation for pragmatic reasoning in the Gricean tradition.

1 INTRODUCTION

Game theory has originally been conceived as a theory of strategic interaction among fully rational agents. Its applicability to real life phenomena like economic or political processes therefore depends on how realistic it is to assume that the acting individuals in these processes are fully rational. Rationality here means, among other things, full awareness of one's own beliefs and preferences and logical omniscience. Even stronger, for classical game theory to be applicable, every agent has to ascribe full rationality to each other agent.

These assumptions are of course unrealistic when applied to humans. However, this does not devaluate game-theoretic models. An apologist of the classical model might argue that to ride a bike one has to be able to act in accordance with the laws of physics, but one does not need to be able to solve differential equations. Likewise, to successfully embark upon a strategic interaction one need not be able to solve game-theoretic problems; all that is required is to *act* in accordance with the laws of game theory.[1]

This argument has a certain appeal. If game theory is to be applied as a descriptive rather than a normative theory, the question remains open of how not-fully-rational beings achieve the quasi-rationality that is required to apply game theory in the first place.

There are various answers to this problem. One line of research, going back to the work of Herbert Simon (1982) (see also Rubinstein 1998) explores the consequences of giving up the strong rationality assumptions of classical game theory. In other words, agents are assumed to be *boundedly rational.* The *evolutionary* interpretation of game theory (see, for instance,

[1] I owe this comparison to Helge Ritter (p.c.).

	A	B
A	1	0
B	0	1

Table 1: Utility matrix for a simple coordination game

Maynard Smith 1982) completely gives up any rationality assumptions. Instead, game theory is used to describe the dynamics of entire populations of agents. Strategies (in the game-theoretic sense) are identified with heritable traits of individuals, and utility with replicative success. Since replicative success of an individual may depend on other individuals of the same population, this involves a strategic component. Game theory can thus be used to model evolution via natural selection in the Darwinian sense.

It has repeatedly been noticed that Gricean pragmatics has a strong game-theoretic flavour (see, for instance, Stalnaker 2005). In particular, it makes the same strong rationality assumptions as classical game theory, and the mentioned objections apply as well. One would thus expect that the notion of bounded rationality has a role to play in laying the foundations of natural language pragmatics. The connection between evolutionary game theory and pragmatics is perhaps not so obvious, but research in economics has shown that evolutionary population dynamics is a useful tool to model cultural processes (see, for instance, Young 1998). Language use, as a cultural phenomenon, thus falls squarely within the realm of this interpretation of game theory.

In the present chapter, I will explore a particular version of an evolutionary game dynamics called *best response dynamics*. While its standard interpretation applies to learning processes in population, it can also receive a very natural epistemic interpretation involving boundedly rational agents. This model will be applied to the problem of relating (conventionalised) semantic and the (non-conventionalised) pragmatic aspects of natural language interpretation.

2 Evolutionary Game Theory

2.1 The Replicator Dynamics

Evolutionary game theory (EGT) was developed by theoretical biologists and especially by John Maynard Smith (1982), as a formalisation of the neo-Darwinian concept of evolution via natural selection. It builds on the insight that many interactions between living beings can be considered to be games in the sense of game theory—every participant has something to win or to lose in the interaction, and the payoff of each participant may depend on the actions of all other participants. In the context of evolutionary biology, the payoff is an increase in fitness, where fitness is basically the expected number of offspring. According to the neo-Darwinian view on evolution, the units of natural selection are not primarily organisms but heritable traits of organisms. If the behaviour of organisms, that is, interactors, in a game-like situation is genetically determined, the strategies can be identified with gene configurations.

For illustration, consider a simple coordination game from Lewis (1969). The utility matrix is given in Table 1. In the evolutionary setting, this is to be interpreted as follows. There is a large population. Each member of the population belongs to one of two sub-groups, A or B. Group membership is heritable. The individuals in the population reproduce via cloning (i.e., each newborn has exactly one parent). Reproductive success depends on the interaction with the other members of the population. The average number of offspring of an individual of type A equals the expected utility of a player of strategy A when playing against a random member of the population, and likewise for group B. For the given utility matrix, this means that the average number of offspring of an A-individual equals the proportion of A-players in the population, and the same for B.

If more than half of the population is of type A, A-players will thus on average have more children than B-players, and the proportion of A-players increases over time. The population as a whole will converge towards a state with only A-players. If the B-players have a majority in the initial state, the population converges towards a homogeneous B-state.[2] If the initial state is exactly 50:50, the population will remain in this state because A-players and B-players have exactly the same birth rate.

There are thus three stationary states of the population: 100% A-players, 100% B-players, and precisely fifty-fifty. Note that these are exactly the three Nash equilibria of this game. There is a difference though between the mixed equilibrium on the one hand, and the two pure equilibria on the other hand. Let the population be in the 50:50 state, but let us assume that the population dynamics is slightly noisy. This may be due to sampling effects if the population is finite after all, or replication may be unfaithful with a certain small probability (like mutations in genetic transmission). Then the population may leave the Nash equilibrium and develop a small A-bias or B-bias. However, as soon as the population has a bias, natural selection will enhance that bias until the population converges towards one of the two pure states. If, on the other hand, the population is in one of the two homogeneous states, a small group of invaders from the other strategy will die out soon because they receive a much lower utility against the incumbent population than the incumbents against themselves.

Hence, while Nash equilibria correspond to evolutionarily stationary states, some such states are resistant against mutations, while other states are not. Maynard Smith dubbed the resistant equilibria *evolutionarily stable states* (ESS).

It turns out that the notion of evolutionary stability is closely related to the rationalistic notion of a Nash equilibrium, but there are subtle differences. It can be shown that the following proper inclusions hold:

$$\textit{Strict Nash Equilibria} \subset \textit{Evolutionarily Stable Strategies} \subset \textit{Nash Equilibria}.$$

The mixed equilibrium from the example above demonstrates that there are Nash equilibria that are not an ESS. As an example for an ESS that is not a strict Nash equilibrium, consider the well-known game Rock-Paper-Scissor. The utility matrix is given in Table 2. This game has exactly one Nash equilibrium, the one where each of the three strategies is played with a probability of $\frac{1}{3}$. Suppose a population is in this equilibrium, that is, each of three sub-populations have exactly the same size. Then each sub-population has the same birth rate and the proportions are stationary. Now suppose Rock gets a small edge over the other two strategies due to unfaithful replication. Then in the next generation, Paper will thrive, one generation later Scissors, then again Rock etc., ad infinitum. Without another unfaithful replication, the population

[2] Standard EGT assumes that, for all practical purposes, populations are so large that they can be considered infinite. Random variation is disregarded.

	R	P	S
R	1	0	2
P	2	1	0
S	0	2	1

Table 2: Utility matrix for Rock-Paper-Scissor

will not return into equilibrium. This illustrates that the single Nash equilibrium of this game is not an ESS.

The population dynamics that ensues if the expected utility is identified with the expected number of offspring is called the *replicator dynamics*. Maynard Smith (1982) gives a static characterisation what it means for a strategy (of a symmetric game) to be evolutionarily stable:

- s is an Evolutionarily Stable Strategy in the replicator dynamics iff
    1. $u(s, s) \geq u(t, s)$ for all t, and
    2. if $u(s, s) = u(t, s)$ for some $t \neq s$, then $u(s, t) > u(t, t)$.

Evolutionary stability in this sense is a sufficient condition for a population state to be dynamically stable under the replicator dynamics. For a large class of games (including the cooperative signalling games that are frequently used to model linguistic communication), it is also a necessary condition.

2.2 The Best Response Dynamics

Many social scientists assume that cultural variables undergo an evolutionary process in a way more or less similar to genes in biology. How close this similarity actually is, is a matter of dispute (see, for instance, the discussion in Richerson & Boyd 2005). If cultural evolution exists, evolutionary game theory should be applicable in this domain as well. There is, in fact, a considerable body of literature on the subject from economics and other social sciences (see Skyrms 1996, Vega-Redondo 1996 and Young 1998, and the literature cited therein).

A strategy, in the social setting, can be considered a behavioral disposition, very much like in the original, rationalistic interpretation of game theory. However, to apply an evolutionary model to social phenomena, it has to be clarified how strategies reproduce. To pose the question in more general terms, which micro-dynamics underlies the macro-dynamics that is modelled by the evolutionary model? Obvious candidates are learning and imitation. If various strategies are differentially successful in being learned and imitated, we expect a process which resembles natural selection.

Such a learning or imitation dynamics disregards the rationality and creativity of human agents. The *best response dynamics* (introduced in Matsui 1992 and thoroughly investigated in Hofbauer 1995) takes these aspects into account, but without adopting the extreme assumptions of the classical model.

Suppose a population of individuals exists that plays certain strategies of a game, just as in the previous setting. In every time step, a new member enters the population. Unlike in the biological setting, the newcomer may freely choose her strategy. If we suppose that newcomers are rational (and well-informed) enough to maximize their expected utility, they will choose a (possibly mixed) strategy that is a best response to the average strategy of the population. Repeating this addition of new members indefinitely, an evolutionary dynamics ensues, but it is not a Darwinian one. New strategies may be invented with the purpose of maximizing utility, while in Darwinian evolution, new strategies only emerge due to undirected random mutation.

Despite this considerable conceptual difference, the replicator dynamics and the best response dynamics are mathematically both similar enough to subsume under the heading of 'Evolutionary Game Theory'.

The notion of evolutionary stability may be applied to the best response dynamics as well. It is easy to see that all ESSs are Nash equilibria—recall that by definition, a Nash equilibrium is a strategy that is a best response to itself. But what are the sufficient conditions for stability here? Reconsider the Rock-Paper-Scissor game. Suppose the population is close to equilibrium—there is the same number of Paper players and Scissor players, and a tiny excess of Rock players. Then the next newcomer will play Paper, and this will continue until Scissor becomes the best response to the population average, which will be followed by Rock, etc. This seems similar to the replicator dynamics scenario. However, here the difference between the state of the population (seen as a vector of fractions) and the Nash equilibrium actually converges to zero. In other words, the Nash equilibrium is, actually, evolutionarily stable.

This example illustrates that the best response dynamics induces a notion of ESS that is slightly more inclusive than Maynard Smith's notion. It can be defined in the following way (which is essentially identical to the formulation of Hofbauer & Sigmund 1998, p. 96):

- s is an Evolutionarily Stable Strategy in the best response dynamics iff
    1. $u(s, s) \geq u(t, s)$ for all t, and
    2. if $u(s, s) = u(t, s)$ for some $t \neq s$, then there is a t' with $u(t', s) = u(s, s)$ and $u(t', t) > u(t, t)$.

2.3 Epistemic Interpretation of the Best Response Dynamics

The best response dynamics can also be given an epistemic interpretation. Let us return to the classical picture of a strategic two-person game, where each player has a preference ordering over the set of profiles (including the mixed ones), that can be represented by a utility function in the standard way. Suppose, however, that the players are entirely irrational. They choose their strategy according to prejudice rather than rational deliberation. For the sake of simplicity, let us assume that both players have the same prejudices, and this prejudice is common knowledge.

It might occur though that player a is not entirely irrational but makes a rational choice with some probability $\epsilon > 0$ that may be arbitrarily small. Making a rational choice means to play a best response against the prejudicial strategy of the other player. Depending on the nature of the original strategy, a's choice may now differ from the original strategy by a small amount.

Now suppose that the other player, b, is also rational with probability ϵ, and furthermore he assumes that a acts as described in the previous paragraph. With probability ϵ, b will thus play the best response to a's strategy, which in turn is the initial strategy with probability $1 - \epsilon$ and a best response to it with probability ϵ.

This process may be iterated. In this way, we may define an infinite sequence of strategy profiles, starting with the initial prejudice and increasing the strategic depth at every step. If I take the action of my opponent into account in my decision, I have the strategic depth of at least 1. If I also take into account that my partner may take my actions into account, my strategic depth is 2, etc. In general, if I assign strategic depth n to my partner, my own strategic depth is $n+1$. However, boundedly rational agents have an upper limit for their strategic depth. Intuitively, a strategic depth of n is the ability to "think n steps ahead" in a sequential game, or to "think around n corners".

To make this notion formally precise, let a symmetric two-person A be given. x_i, the initial strategy, is a (possibly mixed) strategy for A, and ϵ is a real number with $0 < \epsilon \leq 1$. Let $\beta(x)$ be the set of best responses to the strategy x according to A. A *deliberation sequence (based on ϵ)* $x_0, x_1, x_2, \ldots$ is an infinite sequence of strategies with the property that:

$$\begin{aligned} x_0 &= x_i && (1) \\ x_{n+1} &\in \{\epsilon y + (1-\epsilon)x_n \mid y \in \beta(x_n)\} && (2) \end{aligned}$$

Let us assume that the agents are conservative, in other words ϵ is very small. A deliberation sequence may nevertheless lead to another strategy than x_i. In general, we say that *cautious deliberation leads from* x_i *to* x_f iff there is a positive $\epsilon_0 \leq 1$ such that for all $\epsilon < \epsilon_0$ and for all deliberation sequences $\vec{x}$ based on ϵ with $x_0 = x_i$, it holds that $\vec{x}$ converges to x_f. Intuitively, this means that boundedly rational players with a sufficiently small probability to act rationally and having the prejudice to play x_i will play, with arbitrary approximation, x_f, provided that they have, and assign to each other, a sufficiently large strategic depth.

Mathematically, cautious deliberation sequences are identical to time series in a discrete best response dynamics. One would thus expect that cautious deliberation may lead to some strategy x_f if and only if x_f is evolutionarily stable according to the best response dynamics. This equivalence (between end points of cautious deliberation and ESSs) does hold if we add the assumption that, with an arbitrarily small probability η, players choose their strategy completely at random, without any considerations of prejudice or rationality.

Let us reconsider the two games discussed so far. Suppose Rock-Paper-Scissor players have the prejudice to play the Nash equilibrium $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. A boundedly rational player may decide to play Rock instead with probability ϵ, because this is as good a response to the Nash equilibrium as the equilibrium itself. However, a second round of deliberation will reveal that, with the strategic depth of 2, Paper turns out to be the best response. Since ϵ can become arbitrarily small, the probability of playing Rock rather than the mixed equilibrium is thus arbitrarily small, and the same holds for all other strategies that deviate from the equilibrium. Cautious deliberation will thus always remain in the neighbourhood of the equilibrium, and this neighbourhood may be arbitrarily small.[3]

Compare this to the $(\frac{1}{2}, \frac{1}{2})$-equilibrium of the coordination game in Table 1. Suppose that, instead of the mixed equilibrium, a player decides to play, with probability ϵ, the pure strategy A. A second round of deliberation reveals that the best response to this mixed strategy is the pure A, etc. ad infinitum. Via iterated deliberation, the probability of A will thus converge to 1. The Nash equilibrium $(\frac{1}{2}, \frac{1}{2})$ is thus not an endpoint of cautious deliberation, while the two pure equilibria A and B are.

[3] It can actually be shown that, independendly of the initial state, cautious deliberation will always lead to the equilibrium in this game, but the proof of this fact is omitted for reasons of space. The interested reader is referred to Hofbauer & Sigmund (1998).

	S	H	F
S	3	0	–5
H	1	1	–6
F	–2	–3	–10

Table 3: The extended stag hunt game

2.4 Trajectories

The best response dynamics does not only define a notion of stability, but also sequences of strategies that may start at any point in the strategy space and lead, in most cases, to ESSs. This offers a partial solution to one of the central problems of game theory, namely equilibrium selection. If two players cannot communicate prior to the beginning of the game (either due to the lack of communication channels or the lack of mutual trust) and the game has multiple equilibria, there is no obvious way to predict the action of the other player and thus to make an informed strategy choice oneself, even if it were common knowledge that all players are perfectly rational. The situation may actually improve if the players are boundedly rational in the sense described above, provided they know each other's prejudices, in other words the strategy that the other will play if he does not apply rational and strategic deliberation.

Schelling (1960)'s observation about the role of *salience* in equilibrium selection is a case in point. Schelling assumes that in a symmetric game with several equilibria, it is advisable to choose one that is more salient than the other(s). The point can nicely be illustrated by an experiment that is reported in Camerer (2003). The test subjects were grouped in couples. Each person was asked to secretly write down a day of the year. If both members of a couple managed to write down the same date (without any previous communication), they both scored a point. It turned out that a majority wrote down salient days like December 25.

This can be seen as a coordination game with 366 different strategies. The utility is 1 at the diagonal and 0 at all other profiles. The expected probability of a subject who is not thinking strategically will be distributed across all 366 dates, with probability peaks at salient days like the Christmas day. The tendency to choose December 25 will be enhanced by each round of strategic thinking, since the best response to a strategy with a probability peak at December 25 is to choose that very date with probability 1. The best response dynamics is thus bound to converge to this pure strategy, and this is how most test subjects were actually seen to behave.

Lewis (1969) points out that precedent is a good heuristic for equilibrium selection as well. If you had played such a coordination game against the same partner before and managed to meet at an equilibrium, it is prudent to stick to that equilibrium. This is unsurprising given the presumption that people are more likely to repeat themselves than to change their strategy without clear reason to do so.

In the previous examples, if played against itself, each strategy is also a Nash equilibrium. Table 3 contains a more complicated example that includes a strictly dominated strategy. If the game is restricted to the strategies S and H, we obtain the well-known stag hunt game. As a reminder: two hunters have the choice to try to catch a stag (S) or a hare (H). A stag is preferable over a hare, but stag hunt requires the two to collaborate, while each hunter can hunt a hare by himself. The worst outcome is to rely on the cooperation of the other hunter and to try to hunt a stag while the other one is, in fact, not collaborating. The game has two Nash equilibria—both hunting stag or both hunting hare.

In the extended stag hunt game, each player has a third option, namely to coerce the other hunter to go stag hunting by force (F). Such a conflict reduces the utility for both participants, but the one applying force is better off than the one being coerced. Retaliating with force is the worst outcome for both because they will embark upon a costly fight. The best reaction to force is to comply and play S.

The strategy F is strictly dominated and thus should play no role in the considerations of rational players. However, let us suppose that, applying force is the first choice that an irrational player would choose, if he does not think about the consequences of this action. Then the best response is to play S. S is also the best response to any convex combination of the strategies S and F (i.e., any mixed strategy that assigns some probability p to S and $1 - p$ to F). Cautious deliberation thus necessarily leads from F to S.

In real life terms, this illustrates an effective threat to break a deadlock. If one player conveys the impression that he might be willing to apply force—even though this is irrational—then this possibility is reasonable enough for a sufficiently rational player to comply, as long as compliance is in his own enlightened self interest. (Of course in reality threats also work if the victim is forced to act against his own good interests, but this only works if the threat either does no harm to the bully or else the bully is not arbitrarily rational.)

2.5 Best Response Dynamics in Asymmetric Games

So far I have restricted the discussion to symmetric games, namely games where both players have the same strategy set and the same utility matrix. A symmetric Nash equilibrium is a strategy in a symmetric game that is a best response to itself. Asymmetric games, on the other hand, are two-person games where the two players play different roles (or, in the population dynamic interpretation, belong to different populations). An asymmetric Nash equilibrium is a pair of strategies $\langle s_i^A, s_j^B \rangle$ (of player A and B, respectively), such that s_i^A is the best response to s_j^B and vice versa.

Asymmetric games can be transformed into symmetric ones in a straightforward way. Suppose the players are A and B, their strategy sets are $S^A = s_1^A, \dots, s_n^A$ and $S^B = s_1^B, \dots, s_m^B$, and their utility matrices are u_A and u_B, respectively. The new symmetric game has the strategy set $S^A \times S^B$. In other words, each strategy of the meta-game is a pair consisting of an A-strategy and a B-strategy. The utility is calculated as

$$u(\langle s_i^A, s_j^B \rangle, \langle s_k^A, s_l^B \rangle) = u_A(s_i^A, s_l^B) + u_B(s_j^B, s_k^A)\,. \tag{3}$$

The notions defined so far can straightforwardly be applied to asymmetric games, simply by first symmetrising the game. It turns out that the characterisation of evolutionary stability in asymmetric games is actually much simpler to characterise than in the general case:

Theorem 1. A strategy pair $\langle s_i^A, s_j^B \rangle$ of an asymmetric game is evolutionarily stable according to the best response dynamics if and only if it is a strict Nash equilibrium.

Proof. (See Appendix.) □

Recall that a pair of strategies is a *strict* Nash equilibrium iff each component is the *unique* best response. The proof of the theorem is given in the appendix; the same result also holds for evolutionary stability in the Maynard Smith sense, as shown in Selten (1980).

3 FROM SEMANTICS TO PRAGMATICS

The main point of this paper is that cautious deliberation, in the sense defined above, leads from conventionalised semantics, that is, what is said, to the pragmatic content, that is, what is meant. To illustrate this on an informal level with a standard case of scalar implicature, consider the sentence (4).

Some boys came to the party. (4)

The conventionalised semantic strategy is that this sentence serves to convey the proposition that the set of boys coming to the party is non-empty, and this is how a non-rational hearer will interpret it. If we grant that the Gricean maxims (Grice, 1975) are somehow part of the utility function of the speaker, the following statement will be preferred:

All boys came to the party. (5)

This is so if the speaker beliefs that both sentences are true, and the usage of (4) is confined to situations in which some but not all boys came. A listener with strategic depth of 1 will anticipate this and conclude from (4) exactly this—that some but not all boys came.[4] The scalar implicature of (4) that not all boys came is thus a part of the unique ESS that is reachable from the semantic convention via the best response dynamics. A similar story can be told for other cases of conversational implicatures, provided that communication is appropriately modelled as a game with a utility function that formalises the Gricean maxims.

Following much of the work in game-theoretic pragmatics (see for instance Rooij 2004), I will model communication as a version of a *signalling game* in the sense of Lewis (1969). In this setup, a game can be identified with a single utterance situation. Speaker and hearer are the players. Their actions are the production and interpretation of an utterance, respectively, and their payoff preferences correspond to the economy of the speaker and the economy of the hearer.

To be more precise, let us assume that a fixed set of possible worlds W is given. The set of meanings M is a set of propositions over W, that is, $\mathsf{M} \subseteq \mathrm{POW}(W)$. Furthermore, a set F of forms is given. A speaker strategy is any function s from W to F, that is, a production grammar. Likewise, a hearer strategy is a comprehension grammar, that is, a function h from F to M.

Let us thus assume that in each game, some random device, called *nature*, presents the speaker with a possible world $w \in W$ which is not revealed to the hearer. The speaker then has to choose a form that is shown to the hearer and reveals as much information about w as possible. Nature's choice of w is probabilistic; w is drawn from W according to the probability distribution P_n, which is mutually known by the speaker and the hearer.[5]

[4] [See the chapter by Ian Ross in this volume for related discussion.]

[5] For the sake of simplicity, I will assume that W is finite. If P_n is modelled as a probability density function, the model can straightforwardly be extended to an infinite set of possible worlds. For expository reasons, I will refrain from doing this here.

I will choose a utility function for this general signalling game setup that formalises, at least partially, the Gricean maxims. The overarching cooperativity principle translates into the assumption that communication is a game of cooperation. This means that utilities for speaker and hearer are always identical.

Next, let us assume that the hearer has a prior probability function P_H over W (which is also mutually known by the speaker and the hearer). Since P_n is known by the hearer, for a rational hearer it should hold that $P_n = P_H$. Since communication is pointless if it is mutually known that the hearer would not believe what the speaker is trying to say, I will assume that $P_H(m) > 0$ for all $m \in M$. The information state of the hearer is his posterior probability distribution, after incorporating his interpretation of the signal that the speaker emits. This captures part of the maxim of quality: the hearer completely trusts what he is told. For a given speaker strategy s, hearer strategy h, and possible world w, the posterior distribution is $P(\cdot|h(s(w)))$. The number of bits of information that the hearer is still missing to achieve complete information is thus $(-\log_2 P(w|h(s(w))))$. ($\log_2$ is the binary logarithm. The informativity is the number of yes/no-questions that the hearer has to ask in order to figure out with certainty what the real world is like.)

If the hearer's interpretation $h(s(w))$ is false, namely $w \not\in h(s(w))$, this is infinite; otherwise it is the lower the more information $h(s(w))$ contains. Thus, according to the maxims of quality and quantity, the speaker should strive to maximize $\log_2 P(w|h(s(w)))$ in each possible world w.[6] (Since utilities have to be finite, we assume that the hearer mistrusts the speaker with some sufficiently small amount η, which ensures that lying leads to an extremely low yet finite utility.)

The maxim of manner refers to the complexity of the form. I thus assume that there is a cost function c from F to the positive real numbers. The players have an interest in keeping $c(f)$ low.

The game takes the form of a Bayesian game. This means that the utility does not depend just on the strategies of the players, but also on nature's choice. The cooperativity principle together with the maxims of quality, quantity and manner thus lead to the following utility function:

$$u(w, s, h) = \log_2 P_H((w|h(s(w))) - c(s(w))) . \tag{6}$$

By relativising utility further to certain decision problem, it is possible to incorporate the maxim of relevance here (see, for instance, Rooij 2004). For the sake of simplicity, I will assume any information is relevant to the hearer.

A Bayesian game can now be transformed into a strategic game in normal form by averaging over nature's choice:

$$u(s, h) = \sum_{w \in W} P_n(w) \log_2(P_H(w|h(s(w))) - c(s(w))) . \tag{7}$$

4 Implicatures

In this section I will explore the behaviour of the best response dynamics, given the kind of game that was defined in previous section. Let us first take up the example (4) again. To formalise the situation, let us assume that there are just three different possible worlds that can be characterised by first-order formulas (where B stands for *boy* and C for *came to the party*).

[6] According to Robert van Rooij (see for instance Rooij 2004), this is the utility of a proposition provided every piece of information is equally relevant.

To avoid complications relating to the existential presuppositions of the determiner *all*, I assume that there exist boys in all possible worlds.

- $w_1 : \exists x.Bx \wedge \forall y.By \rightarrow Cy$
- $w_2 : \exists x.Bx \wedge Cx \wedge \exists y.By \wedge \neg Cy$
- $w_3 : \exists x.Bx \wedge \neg\exists y.By \wedge Cy$

All three worlds are equally likely both for nature and for the hearer:

$$P_{\{n,H\}}(w_i) = \frac{1}{3}\,. \tag{8}$$

There are four possible forms that the speaker can choose from:

- f_1: *Some boys came to the party.*
- f_2: *All boys came to the party.*
- f_3: *No boys came to the party.*
- f_4: *Some, but not all boys came to the party.*

f_4 is more complex than the others, which all have roughly the same complexity. Let us say that:

$$c(f_1) = c(f_2) = c(f_3) \;=\; 1 \tag{9}$$

$$c(f_4) \;=\; 3\,. \tag{10}$$

The semantic conventions of English relate forms to possible worlds. This corresponds to a pair of strategies: in each possible world, the speaker chooses a conventionally true form at random, and the hearer fully believes the conventional meaning of the form that he perceives. In the current game, this can be depicted as follows:

$$(s_0, h_0) = \left[\begin{tikzcd}
w_1 \arrow[r] \arrow[dr] & f_1 \arrow[r] & \{w_1, w_2\} \\
w_2 \arrow[ur] \arrow[ddr] & f_2 \arrow[r] & \{w_1\} \\
w_3 \arrow[r] & f_3 \arrow[r] & \{w_3\} \\
 & f_4 \arrow[r] & \{w_2\}
\end{tikzcd}\right]$$

If symmetrised, the utility of this strategy pair against itself is about −3.22.

The best response of the speaker against the conventional hearer strategy is to map w_1 to f_2 (because then the hearer's posterior probability is 1 while the costs of f_1 and f_2 are identical), and to map w_2 to f_1 (which means a loss of .5 in informativity but a gain of 2 in costs). The best response of the hearer to the conventional speaker strategy is still the conventional hearer strategy. Hence, the best response to the conventional strategy pair is:

$$(s_1, h_1) = \left[\begin{tikzcd}
w_1 \arrow[dr] & f_1 \arrow[r] & \{w_1, w_2\} \\
w_2 \arrow[ur] & f_2 \arrow[r] & \{w_1\} \\
w_3 \arrow[r] & f_3 \arrow[r] & \{w_3\} \\
 & f_4 \arrow[r] & \{w_2\}
\end{tikzcd}\right]$$

The utility of (s_1, h_1) against (s_0, h_0) is about -2.94, and the utility of (s_1, h_1) against itself is about -2.67. If a cautiously rational player decides to play (s_1, h_1) with a sufficiently small probability ϵ, and otherwise the convention (s_0, h_0), the best response to this mixed strategy is still (s_1, h_1). However, after some finite number n of iterations,[7] the probability of (s_1, h_1) is large enough such that (s_2, h_2) becomes the best response. Nothing changes with regard to the speaker strategy, but in h_2 the hearer has figured out that the speaker utters f_1 if and only if w_2 is true; hence $\{w_2\}$ is the pragmatically informed interpretation of f_2.

$$(s_2, h_2) = \left[\begin{tikzcd} w_1 \arrow[dr] & f_1 \arrow[r] & \{w_2\} \\ w_2 \arrow[ur] & f_2 \arrow[r] & \{w_1\} \\ w_3 \arrow[r] & f_3 \arrow[r] & \{w_3\} \\ & f_4 \arrow[r] & \{w_2\} \end{tikzcd}\right]$$

The utility of (s_2, h_2) against (s_1, h_1) is -2.333, while the utility of (s_2, h_2) against (s_0, h_0) is -3.0.

Notwithstanding the fact that (s_2, h_2) is the best response to a mixed strategy consisting predominantly of (s_1, h_1) and (s_2, h_2), it is still not a stable state. We have to take into account that the speaker believes, with a small probability η, that the speaker picks out a signal at random. If the probability of s_0 and s_1 (the speaker strategies in which f_4 may be used to express $\{w_2\}$) drops below a certain threshold, the best response for the hearer is to ignore f_4 altogether. This leads to

$$(s_4, h_4) = \left[\begin{tikzcd} w_1 \arrow[dr] & f_1 \arrow[r] & \{w_2\} \\ w_2 \arrow[ur] & f_2 \arrow[r] & \{w_1\} \\ w_3 \arrow[r] & f_3 \arrow[r] & \{w_3\} \\ & f_4 \arrow[r] & \{w_1, w_2, w_3\} \end{tikzcd}\right]$$

This strategy pair is a strict Nash equilibrium and thus evolutionarily stable. Using the terminology from the previous section, we have shown that cautious deliberation leads from the conventional, semantic strategy pair (s_0, h_0) to the pragmatic equilibrium (s_4, h_4).

The transition from the semantic convention (s_0, h_0) to the pragmatically usable strategy pair (s_4, h_4) via the best response dynamics illustrates two important pragmatic phenomena. In s_2, the speaker expresses the fact that all boys come to the party with 100% certainty as f_2, *All boys came to the party*, because this is, according to h_0, more specific than the equally true f_1, *Some boys came to the party*. This is a consequence of the maxim of quantity. Anticipating this, the hearer pragmatically strengthens the interpretation of *Some boys came to the party* in h_2 to the interpretation $\{w_2\}$, *some but not all boys came to the party*. This is a scalar implicature, and the best response dynamics captures the intuitive reasoning used in Gricean pragmatics to explain this effect. Furthermore, the hearer figures out in h_4 that the signal f_4 is pragmatically sub-optimal for the speaker in all conceivable situations, because its conventional meaning can be conveyed via f_1 in a more economical way. Therefore, this signal ceases to carry a pragmatic meaning and is ignored in the stable state. In the literature, this phenomenon is called *total blocking* (as opposed to *partial blocking*, where only part of the conventional meaning of an expression is pragmatically blocked by a competing expression). A good example for this phenomenon are

[7] Where $n > \frac{\log(u(1,0)-u(1,1))-\log(u(1,0)-u(1,1)+u(2,1)-u(1,1))}{\log(1-\epsilon)} \approx \frac{-2.783}{\log(1-\epsilon)}$ for the particular numbers chosen here; $u(i,j)$ being the symmetrised utility of (s_i, h_i) against (s_j, h_j).

regular derivations that compete with underived words, like *pig* that cannot be used to refer to meat from pigs (while *chicken* or *lamb* can be used to refer to meat from the respective animals) because there is a special term, *pork* with precisely this meaning.

In the example at hand (*Some but not all boys came to the party*), this effect does not actually occur. The wrong prediction is related to the fact that I assumed it to be common knowledge that the speaker has perfect knowledge. In most situations, the hearer cannot be sure about this. If the speaker has only partial knowledge, the scalar implicature only arises with a certain probability, not with certainty. This is sufficient to preempt total blocking.

It should be noted that not much hinges on the particular numbers chosen here. What is relevant is just the following inequality:

$$-\log_2(P_H(w_1|\{w_1,w_2\})) > c(f_4) - c(f_1)\,. \qquad (11)$$

In some sense, this inequality compares incommensurable quantities, namely the differential informativity of two strategies versus that of differential complexity of two expressions. The relative weight of informativity and complexity depends on various situational factors, and thus the inequality may actually be true or be false depending on the context. A more complete model could show the relative importance of these two factors by some random parameter that is itself a component of strategic reasoning.

5 Conclusion

Space does not permit to spell out the consequences of this approach to other pragmatic phenomena in detail. I will thus conclude by briefly setting the approach that was developed in this chapter into a broader context.

The basic idea to connect semantics and pragmatics via an iterated process of computing the best response in a signalling game setting is due to Stalnaker (2005). In the dynamics that Stalnaker uses, a strategy is entirely replaced by the best response to it. (In terms of my formalisation, that means $\epsilon = 1$ in Stalnaker's model.) The two models are not equivalent, but they coincide in many applications, including the example discussed above. In either version, best response dynamics shares with most of the alternative game-theoretic solution concepts the notion of a Nash equilibrium as a stable configuration in a strategic interaction. However, and again in either version, the attractive feature of iterated best response is that equilibria can be grounded in strategy profiles that need not be in equilibrium and may even be strictly dominated (like the F-strategy in the extended stag hunt game above). This appears to be an important asset in many situations of strategic interaction, including communication, in which salience and precedent single out a profile that may or may not be rationally justifiable. I hasten to add that I do believe that natural languages are in an evolutionarily stable state, at least with a very high probability. This applies to languages (in the sense of populations of utterances) as a whole. A strategy profile that may be optimal on average may be non-rationalisable in a particular utterance situation. Best response dynamics thus serves to establish a link between the macro-structure and the micro-structure of linguistic communication.

REFERENCES

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press, Princeton.

Grice, H. P. (1975). Logic and conversation. In: *Syntax and Semantics 3: Speech Acts* (P. Cole and J. Morgan, eds.), pp. 41-58. Academic Press, New York.

Hofbauer, J. (1995). Stability for the best response dynamics. Preprint.

Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics.* Cambridge University Press, Cambridge.

Lewis, D. (1969). *Convention.* Harvard University Press, Cambridge.

Matsui, A. (1992). Best response dynamics and socially stable strategies. *Journal of Economic Theory*, **57**, 343-362.

Maynard Smith, J. (1982). *Evolution and the Theory of Games.* Cambridge University Press, Cambridge.

Richerson, P. J. and Boyd, R. (2005). *Not by Genes Alone. How Culture Transformed Human Evolution.* University of Chicago Press, Chicago and London.

Rubinstein, A. (1998). *Modeling Bounded Rationality.* MIT Press, Cambridge, Mass.

Schelling, T. C. (1960). *The Strategy of Conflict.* Harvard University Press, Cambridge, Mass.

Selten, R. (1980). A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology*, **84**, 93-101.

Simon, H. (1982). *Models of Bounded Rationality* (Vol. 2). MIT Press, Cambridge, Mass.

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

Stalnaker, R. (2005). Saying and meaning, cheap talk and credibility. In: *Game Theory and Pragmatics* (A. Benz, G. Jäger and R. van Rooij, eds.). pp. 83-100. Palgrave MacMillan.

van Rooij, R. (2004). Signalling games select Horn strategies. *Linguistics and Philosophy*, **27**, 493-527.

van Rooij, R. (2004). Utility, informativity, and protocols. *Journal of Philosophical Logic*, **33**, 389-419.

Vega-Redondo, F. (1996). *Evolution, Games, and Economic Behaviour.* Oxford University Press, Oxford.

Young, H. P. (1998). *Individual Strategy and Social Structure. An Evolutionary Theory of Institutions.* Princeton University Press, Princeton.

APPENDIX

Proof of Theorem 1: It is immediate from the definition that each strict Nash equilibrium is an ESS in the best response dynamics. So suppose $\langle s_i^A, s_j^B \rangle$ is a non-strict Nash equilibrium that is an ESS in the best response dynamics. This means that there is a strategy pair $\langle s_k^A, s_l^B \rangle \neq \langle s_i^A, s_j^B \rangle$ such that $u_A(s_k^A, s_j^B) = u_A(s_i^A, s_j^B)$ and $u_B(s_l^B, s_i^A) = u_B(s_j^B, s_i^A)$. Therefore either $i \neq k$ or $j \neq l$. Let us assume, without loss of generality, that $i \neq k$. Since $\langle s_i^A, s_j^B \rangle$ is an ESS, there must be a pair $\langle s_l^A, s_m^B \rangle$ with $u_A(s_l^A, s_j^B) = u_A(s_i^A, s_j^B)$ and $u_B(s_m^B, s_i^A) = u_B(s_j^B, s_i^A)$ such that $u_A(s_l^A, s_j^B) + u_B(s_m^B, s_k^A) > u_A(s_k^A, s_j^B) + u_B(s_j^B, s_k^A)$. Since both $\langle s_l^A, s_m^B \rangle$ and $\langle s_k^A, s_j^B \rangle$ are best responses to $\langle s_i^A, s_j^B \rangle$, $u_A(s_l^A, s_j^B) = u_A(s_k^A, s_j^B) = u_A(s_i^A, s_j^B)$. Hence $u_B(s_m^B, s_k^A) > u_B(s_j^B, s_k^A)$, and thus $m \neq j$.

We restrict the game to the sub-game G that results if all strategies are eliminated that are not best responses to $\langle s_i^A, s_j^B \rangle$. It follows from the previous paragraph that in the resulting sub-game, there are at least two A-strategies (including s_i^A) and at least two B-strategies (including s_j^B). The definition of ESS entails that $\langle s_i^A, s_j^B \rangle$ is the only Nash equilibrium in this sub-game.

We define an accessibility relation R between profiles in the following way: $R(\langle s_a^A, s_b^B \rangle, \langle s_c^A, s_d^B \rangle)$ iff either $s_a^A = s_c^A$ and s_d^B is a best response to s_b^B, or $s_b^B = s_d^B$ and s_c^A is a best response to s_a^A. Let R^* be the reflexive and transitive closure of R. Now there are two options:

1. There is a profile x such that not $xR^*\langle s_i^A, s_j^B \rangle$. Then we can form the sub-game G' consisting of all the strategies in G that are components of profiles reachable from x via R^*. Neither s_i^A nor s_B^j belong to this G', because either strategy in G is a best response either to s_i^A or to s_j^B by construction. According to Nash's theorem, G' has a Nash equilibrium. This equilibrium must simultaneously be a Nash equilibrium of G, since none of the excluded strategies is a best response to any of the strategies in G'. So G has a second Nash equilibrium besides $\langle s_i^A, s_j^B \rangle$, which is in contradiction with the assumption that $\langle s_i^A, s_j^B \rangle$ is an ESS.

2. $\langle s_i^A, s_j^B \rangle$ is reachable from any profile in G via R^*. This entails that there are strategies $s_o^A \neq s_i^A$ and $s_p^B \neq s_j^B$ such that either $R(\langle s_o^A, s_p^B \rangle, \langle s_i^A, s_p^B \rangle)$ and $R(\langle s_i^A, s_p^B \rangle, \langle s_i^A, s_j^B \rangle)$, or $R(\langle s_o^A, s_p^B \rangle, \langle s_o^A, s_j^B \rangle)$ and $R(\langle s_o^A, s_j^B \rangle, \langle s_i^A, s_j^B \rangle)$. In the former case, $\langle s_i^A, s_p^B \rangle$ must be a Nash equilibrium of G, and likewise in the latter case $\langle s_o^A, s_j^B \rangle$.

Hence, both scenarios lead to the conclusion that G has a Nash equilibrium besides $\langle s_i^A, s_j^B \rangle$, which is in contradiction with the assumption that $\langle s_i^A, s_j^B \rangle$ is an ESS. We have thus proved that any asymmetric non-strict Nash equilibrium is not evolutionarily stable, or equivalently, that every asymmetric ESS is a strict Nash equilibrium. □

This page intentionally left blank

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 8

BUILDING GAME-THEORETIC MODELS OF CONVERSATIONS

Jun Miyoshi
Kanto-Gakuin University

The purpose of this paper is to build game-theoretic models of conversations. First, a simple example of conversation is analysed, and how to formulate it into a game is discussed. Second, a family of games with perfect and complete information is presented as a general model of conversations, and some theorems in game theory are applied to it. Third, a family of games with incomplete information is presented as a more realistic model of conversations, and it is suggested that techniques in game programming are applicable to it. Finally, the models are examined for their strengths and weaknesses.

1 INTRODUCTION

The purpose of this paper is to build game-theoretic models of conversations. While conversations can be approached in many ways, the rational viewpoint should not be underemphasised. Almost all conversations are carried out by human beings. They are rational agents who try to maximize their interests. Accordingly, a conversation can be understood as the collective activity of intelligent subjects maximizing utility for themselves, cooperatively or competitively. Hence, game theory, which rigorously analyses rational players' mutual behaviour, will provide promising methods of investigating conversations. This is why models for which the theory is available are worth constructing.

There are many game-theoretic studies of conversations, for example, Asher et al. (2001), Hashida (1996) and Parikh (2001, 2006). I think, however, that studies of this kind often have some of the following inadequacies. First, they determine players' utility functions arbitrarily. In a typical case, the speaker's and the hearer's utility functions are supposed to give the maximum values when she[1] communicates her intention and when he successfully understands it, respectively.[2] Unfortunately, this is not justifiable because it is possible that she tells a lie and he does

[1] I refer to a speaker by "she" and to a hearer by "he".

[2] For instance, Hashida (1996, p. 531) says, "This restricted sense of nonnatural meaning implies that communication is inherently collaborative, because both S [the speaker] and R [the receiver] want that R should recognize c [content] and I [the proposition that S intends to communicate c to R]."

not want to listen to her.[3] Actually, theorists could prove any behaviour to be perfectly rational if they could determine relevant utility functions as they like. Such a proof would have little significance. Thus, in principle, utility functions should be open in a model and decided from reliable observations for application. Second, they deal with only a small part of a conversation; in most of the cases, just a pair of the speaker's one-shot utterance and the hearer's understanding of it. It seems that they miss the structure of a whole conversation. Third, each of these studies is only for one particular purpose such as to show the derivability of some conversational implicature. Those models of conversations that are general enough and application-independent will be more desirable. Finally, they do not make the best use of game theory. The theory has many ideas and theorems which can be effective tools of analysing a conversation. Nevertheless, many game-theoretic studies of pragmatics utilise only Nash equilibrium.

Game programming is another discipline concerning games. Though it has been developed quite independently from game theory, I believe that it is applicable to games dealt with in game theory. If conversation is a game, then it should be possible for it to be studied in game programming as well as in game theory.

The remainder of this chapter is organised as follows. In Section 2, I present a simple example of conversation and discuss how to formulate it into a game. In Section 3, I propose a general model of conversations, to which some theorems in game theory are applied. In Section 4, a more realistic model is formulated, which concerns a family of incomplete-information games. I also suggest that techniques in game programming will be useful to the study of conversation. In Section 5, the models are examined for their strengths and weaknesses. In Section 6, some open questions are posed.

2 Action, Tree and Path in Conversation

In this section, I discuss how to describe a conversation and to formulate it into a game. First, in presenting an example of conversation, I argue that it should be described not as a sequence of utterances but as that of speech acts. Second, I show that the conversation example can be formulated into an extensive game by viewing speech acts as actions in terms of game theory. Third, I suggest that the conversation develops along its subgame perfect equilibrium path.

Let us consider the following natural and simple example of conversation.

(1) A and B are in a train. They are sitting side by side; A on the aisle seat and B on the window seat. A begins to feel that it is hot and stuffy.

A: Excuse me.

B: Yes?

A: Would you open the window?

B: Sure.

(While A waits, B opens the window.)

A: Thank you very much.

B: You're welcome.

[3] This sort of inadequacy is shared by other formal studies of conversation. For example, in Cohen & Levesque (1985, p. 55), their Theorem 2 seems to say that a hearer *automatically* obeys a speaker's request.

Though the most natural way of describing a conversation might be describing it as a sequence of the utterances or sentences uttered in the conversation, it has two undesirable features. One is that it does not include non-linguistic behaviour, which is an element of conversational interaction. B's opening the window in the example, which is not a linguistic utterance, does not have its place in the sequence of the utterances. The other is that it does not show the connections between utterances and actions. In the above example, obviously, A's asking to open the window and B's opening the window have a close connection. However, it does not appear in the sequence of utterances. Because of these two points, describing a conversation as a sequence of utterances is not expressive enough to provide models of conversations.

I propose, hence, to describe a conversation as a sequence of speech acts, or illocutionary acts, and physical acts. By applying speech act theory (Austin 1975, Chapter 8; Searle 1969, Chapter 3), the conversation example is converted to the following.

1. A addresses B in saying, "Excuse me."
2. B replies to A in saying, "Yes?"
3. A asks B to open the window in saying, "Would you open the window?"
4. B accepts A's asking in saying, "Sure."
5. A waits for a while.
6. B opens the window.
7. A thanks B in saying, "Thank you."
8. B replies to A in saying, "You're welcome."

In this sort of representation, physical acts can be located correctly, and the connections between them and speech acts are clear. In the example above, A's asking B to open the window is related to B's opening the window. This relation needs the speech-act description of a conversation in order to appear palpably.[4]

The above description suggests a straightforward way to set the form of extensive game for the conversation. Let us consider that the speakers are the players, their speech acts and related physical acts are actions,[5] and the turn-taking constitutes the game tree of the conversation as a game. In this way, the conversation is represented by the game tree in Figure 1 considering many alternatives A and B did not choose at each turn.

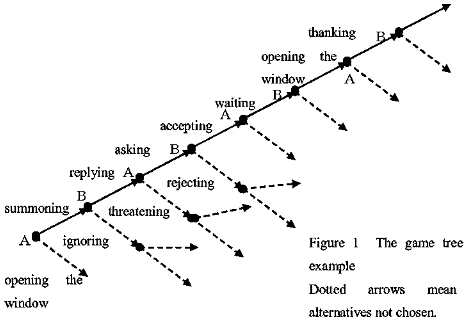


Figure 1 The game tree of the example
Dotted arrows mean some alternatives not chosen.

[4] The view that a conversation is a combination of speech acts is not new. See e.g. Holdcroft (1979, p. 125).

[5] Hereafter I use "action" in the game-theoretic sense.

I suggest that the conversation develops along its subgame perfect equilibrium path. A subgame perfect equilibrium is a Nash equilibrium of a game such that it induces a Nash equilibrium in every subgame of the game (Selten, 1973, 1975). A subgame perfect equilibrium path is found by a backwards induction. Backwards induction is to choose the best action for the player at each node beginning from the terminal nodes and proceeding backwards to the root of the game tree.

Let us exercise the backwards induction on the above example of conversation. Unfortunately, it should be informal using common sense and natural language because the utility or payoff functions of A and B are not known and because not all the alternatives at each turn can actually be enumerated. However, its result will be adequately persuasive since it is very close to our ordinary judgments in social activity. Beginning from the last node, at turn 8, replying will be the best action for B.[6] If he ignores A's thanking, it may cause a trouble to him by making her angry or by letting her think mistakenly that he missed her thanking and repeat the same utterance. At turn 7, then, the best action for A will be to thank B for his opening the window. If he ignores her thanking, she does not have to say "Thank you." But he will not, as we have seen. Moreover, if A does not thank B, he will not be pleased and it will lead to a worse situation for her than when she thanks him. Then, at turn 6, it will be the best for B to open the window. By his doing it, the conversation will peacefully close and the cost of opening the window is small. Additionally, B's not performing the action may produce tension between A and B and thus a worse result for B. Because B will, as we have seen, open the window, at turn 5, A had better wait for a while instead of pressing him to do it. At turn 4, accepting A's asking will be the best action for B. Otherwise his action will cause a worse result in the same way as at turn 6. At turn 3, this is the reason why A will happily ask B to open the window. At turn 2, B should reply to A's request for almost the same reason as at turn 8. Finally, at turn 1, presupposing these reactions of B, the best action for A to do is to address B. Now, the informal backwards induction is complete and it has been reasonably suggested that the conversation develops along the subgame perfect equilibrium path.

3 A General Model

In this section, I formulate a general model of conversations by the method described in Section 2. The model is a family of games with perfect and complete information. By applying theorems of game theory some corollaries are derived.

3.1 The Definition of Model C

Restating some results of Section 1, a conversation is a game if the speaker and the hearer are assumed to be the players, speech acts and related physical acts to be the actions, and turn-taking to be a component of the game tree. Generally, a speech act can be represented by the ordered pair of illocutionary force and propositional content (Austin 1975, p. 102, Searle 1969, Section 2.4).[7] Thus, we can define a family of games, C, which is a model of conversations, as follows.

[6] I refer to A by "she" and to B by "he".

[7] Searle's notation "F(p)", where "F" is a device indicating illocutionary force and "p" is an expression for a proposition (Searle, 1969, p. 31), will have to be understood as an ordered pair (F, p). For "F" seems neither a function nor a modal operator.

Definition 1. C is a family of games which satisfy the following conditions:[8]

- The set of players: $N = \{1, 2\}$;
- The set of actions: $A = F \times S$ (F is the set of illocutionary forces and S is the set of declarative sentences. An example of action is (asking, "Player 2 opens the window")), that is, asking player 2 to open the window.;
- The set of strategies: $A_i = A^m$ ($\forall i \in N$. m is the number of i's moves in the game);
- Utility functions: $U_i\colon A_i \times A_j \to \mathbb{R}$ ($i, j \in N, i \neq j$);
- The game tree is such that

  1. each player has her or his move one after the other;[9]
  2. any action in A is possible at every turn;
  3. there is no chance move;
  4. every information set contains only one node (perfect information).[10]

Some remarks about the definition of C are in order.

1. **Information:** Perfect and complete information is assumed in Definition 1. Perfect information was mentioned above in the statement about the game tree. Complete information means that every player knows each other's utility function and the rules of the game, mutually.

2. **The set of players** N**:** The members of N above are just 1 and 2 because only two-person conversations are under consideration. In principle, the cardinal number of N can be any natural number, though within the practical limit of computability.[11]

3. **The set of illocutionary forces** F**:** F includes "executing", which indicates doing a physical act. For example, player 2's action (executing, "Player 2 opens the window") means that player 2 opens the window but not, for instance, that player 2 promises that he opens the window, which is (promising, "Player 2 opens the window"). In terms of philosophy of action, (executing, s) will be interpreted as an intentional action under the description s (Davidson, 1980).

4. **The set of declarative sentences** S**:**

   (a) S should be an adequately large but finite set of sentences since the participants of an actual conversation have the limits of both short term memory and physical ability to speak.

[8] C produces a game when the utility functions are determined.

[9] I refer to player 1 by "she" and to player 2 by "he".

[10] This is the simplest one. I think that the game tree of a conversation should be refined following the empirical studies of turn-taking system in conversation analysis.

[11] The more players there are, the more complicated the game tree becomes and so the more computation is needed to solve the game. Furthermore, multi-person conversations would begin to show aspects of cooperative games.

(b) S includes the empty set ∅. While it is used to mean a speech act with no content like addressing, "executing" may also have it. The interpretation of (executing, ∅) is "doing nothing", "passing" or "dummy move". (See Remark 5 below.)

(c) A sentence in S is to indicate the propositional content of a speech act. When a propositional content is expressed not in a subordinate clause but with an infinitive, a gerund, or any other non-sentential form according to the natural usage of the language used to describe the conversation under consideration, an appropriate sentence should be selected for the content. For instance, asking player 2 to open the window corresponds to (asking, "Player 2 opens the window") by being given the sentence "Player 2 opens the window". In addition, in English, S includes incomplete or open sentences that contain interrogative pronouns or "whether",[12] in case of questioning. For example, questioning what player 2 does is equivalent to (questioning, "Player 2 does what").[13]

(d) S involves not only true sentences but also false sentences. When s is false and the speaker knows it, (stating, s) is to tell a lie.

(e) The meaning of a sentence in S is given by a semantics, say, truth-conditional semantics, independently of the model C. In other words, C is neutral in semantic theories. Problems about reference, anaphor or intensionality are not addressed here.

(f) S contains the description of a speech act. For example, "Player 1 asks player 2 to open the window" is in S. Thus, (executing, "Player 1 asks player 2 to open the window") is a member of A. However, it should be replaced with (asking, "Player 2 opens the window") to make its force explicit. Generally, supposing that s and s′ are in S, f in F, and i in N and that s′ is such that "Player i does f that s", then (f, s) is to be chosen in place of (executing, s′).[14] Because they are equivalent in behaviour as well as in their effects, this prescription is harmless for the generality of the model.

(g) If s in S is "Player i says that c" or something similar, "say" and any other general locutionary verb should be exchanged for a specific illocutionary verb such as stating and commanding, and be modified by item (f) above.

(h) If s in S is a description of a physical act that has the illocutionary force, for example nodding, (executing, s) should be replaced with (f, ∅), provided with some appropriate f. For example, when player 1 states "It rains" and player 2 nods, their actions are reported as ((stating, "It rains"), (agreeing, ∅)).

Rules (f)–(h) might appear ad hoc, but they are necessary to make explicit, at the same time, the illocutionary force of each utterance, to include relevant physical acts in a conversation, and to admit all the meaningful declarative sentences, according to semantics, into S.

[12] "Whether" seems necessary to distinguish asking A to do X and asking whether A does X. The former is (asking, "A does X") and the latter (asking, "Whether A does X").

[13] Questionings have various types of content, which should be treated carefully. For a classification of the cores of interrogative sentences, see Ludwig (1997, pp. 42–45). The core of an interrogative sentence is similar to the content of a questioning speech act.

[14] This is applied to the second-order directives, too. For instance, in a three-person game, (executing, "Player 1 asks player 2 to ask player 3 to open the window") should be player 1's (asking, "player 2 asks player 3 to open the window"). If player 2 obeys it, he will do (asking, "Player 3 opens the window") rather than (executing, "player 2 asks player 3 to open the window").

5. The number of all the moves m**:** The number of the moves of a game can be made constant by a large enough natural number being assigned to it. Until the number is reached, both players are supposed to make dummy moves after the game is substantially closed (von Neumann and Morgenstern, 1953, p. 60).

Using the notation of C, the sequence of actions in the example of conversation in Section 1 is redescribed as follows:

1. A (addressing, $\emptyset$)
2. B (replying, $\emptyset$)
3. A (asking, "B opens the window")
4. B (accepting, "B opens the window")
5. A (executing, "A waits")
6. B (executing, "B opens the window")
7. A (thanking, $\emptyset$)
8. B (replying, $\emptyset$).

The fact that the conversation develops along the equilibrium path is expressed, though a little loosely, by the formula:

$$\begin{aligned} u_i \quad & ((\text{summoning}, \emptyset), (\text{asking}, \text{“B opens the window”}), \\ & \qquad (\text{waiting}, \emptyset), (\text{thanking}, \emptyset); \\ & (\text{replying}, \emptyset), (\text{accepting}, \text{“B opens the window”}), \\ & (\text{executing}, \text{“B opens the window”}), (\text{replying}, \emptyset)) \\ \geq \quad & \qquad u_i(a_i; a_j) \\ & \quad \forall a_i \in A_i, \forall a_j \in A_j; i, j \in N, i \neq j. \end{aligned}$$

Clearly, C covers all types of conversation as far as they have two participants and are composed of speech acts and physical acts. Therefore, C is a general model of conversations.

The explication of C is now complete. Next, I will discuss some questions it poses.

1. Some might think that, since some actions are logically impossible for a player—for example, player 1's (executing, "Player 2 opens the window")—they should be excluded from her or his set of actions. However, on the one hand, it is plausible that impossible actions never affect the utility value so that they are not chosen. Thus, they are innocuous in the set of actions. On the other hand, if every impossible action should be out of the set of actions, we will need complicated rules to decide whether an action should be in the set or not. The reason is that a complex action involving an impossible action with a logical connective can be a possible action. For instance, while (executing, "Player 2 opens the window") is impossible for player 1, but (executing, "Player 1 opens the window or player 2 opens the window") is possible for her. Taking these into consideration, I prefer the simplicity of the definition to the restriction of the set.

2. Some might ask how the negation of a force (Searle, 1969, pp. 32–34) is treated. I would point out that the force negation is not the truth-functional negation, because illocutionary

force is neither a sentence nor a proposition. Therefore, if there is any negative force, it should be thought to be a kind of force, say, non-command or anti-command, if they make sense, which should be included in F. Otherwise, non-f should be described as (executing, "Player i does not f that s"), which is included in $A = F \times S$.

3. A question about the set of actions would be why it is not just the set of sentences to be uttered. I have three reasons. First, a mere sentence uttered or utterance is often not rich enough to identify its force. In fact, explicit performatives are rarely used in real conversations. When the illocutionary force of an utterance is implicit, it will be expressed by various non-linguistic means such as a facial expression and the tone of voice. However, installing them in the model makes it too complicated. Second, the set of utterances does not include physical acts mentioned in Section 2. Finally, though the set of sentences to be uttered must contain imperatives and interrogatives, their semantics is controversial.

4. A question might be asked why an action in C is the pair of force and content but not just one description of a speech act (or physical act) because the latter can play the role of the former as shown in point (f). This is a very reasonable view. Actually, there is a one-to-one correspondence such that $N \times F \times S \rightarrow S$ (here, S is infinite). In other words, player i's (f, s) is translated into the sentence "Player i does f that s". It is possible, therefore, to make the set of actions just S or the set of declarative sentences. My reply is pragmatic: the models which indicate forces explicitly are more serviceable than others for studying conversation. They show the relationship between commanding s and doing s, and the distinction between commanding s and requesting s, for example. These will usually affect the values of utility functions. In addition, representing the performance of a speech act as an ordered pair of force and content will avoid semantic difficulties about force and meaning, that-clauses and quotations, which are not subjects of conversation studies.

5. Adverbs modifying illocutionary verbs raise another problem. How can we deal with "Player 1 urgently commands player 2 to open the window", for example? I have two solutions. One is to decide that "urgently commanding" is another illocutionary force. In this case, F is defined as $F = Adv \times V$, where Adv is the set of adverbs which can modify illocutionary verbs and V is the set of illocutionary verbs. The other solution is to make an action a triplet of adverb, force and sentence, for instance, (urgently, commanding, "Player 2 opens the window"). Since this problem seems not very important, I will ignore it for the sake of simplicity.

6. Some might feel that the notion of a dummy move is arbitrary. This notion nevertheless has some support based on observation. Schegloff & Sacks (1973, p. 324), citing a recorded example, state that, "[T]here are possibilities throughout a closing, including the moments after a 'final' good-bye, for reopening the conversation". The dummy moves in C could be interpreted as filling these possibilities after a good-bye.

3.2 Some Applications of Game Theory

Some well known theorems in game theory can be applied to C. I state three of them, without proof, in order to derive some corollaries about C. They are provided with interpretations, which should be of some philosophical interest.

Theorem 1 (The existence of a Nash Equilibrium, Nash 1950, 1951). Every finite strategic game has a mixed-strategy Nash Equilibrium. □

Corollary 1. Every game which belongs to C has a mixed-strategy Nash Equilibrium.

Proof. By definition, every game which belongs to C is a finite strategic game. (Remark 4a states that S is finite. Obviously, F is so, too. An extensive game can be formulated into a strategic game.) Therefore, by Theorem 1, it has a mixed-strategy Nash equilibrium. □

An interpretation of Corollary 1 is that every conversation has a solution or is feasible. (Note that Corollary 1 holds for games with imperfect information.)

Theorem 2 (The existence of a Subgame Perfect Equilibrium, Kuhn 1953). Every finite extensive game with perfect information has a pure-strategy subgame perfect equilibrium. □

Corollary 2. Every game which belongs to C has a pure-strategy subgame perfect equilibrium.

Proof. By definition, every game which belongs to C is a finite extensive game with perfect information. Therefore, by Theorem 2, it has a pure-strategy subgame perfect equilibrium. □

An interpretation of Corollary 2 is that every conversation is realised as a sequence of speech acts and physical acts, a sequence which is predictable, definite, and stable for all the players. (By "definite" I mean that the choices of actions are not probabilistic, and by "stable" that no player has any good reason to deviate from the sequence or path because it necessarily causes a loss to the deviant player.)

Theorem 3 (Truncation and Subgame Perfect Equilibrium, Kuhn 1953, Selten 1973). Suppose that a game has subgames and the remaining part of the game. Then, the game has a subgame perfect equilibrium if and only if (1) it induces a subgame perfect equilibrium for each of the subgames and (2) it induces a subgame perfect equilibrium for the remaining part of the game. □

Corollary 3. A game of model C has a subgame perfect equilibrium such that it induces a subgame perfect equilibrium for every part of the game. (Proof is omitted.) □

An interpretation of Corollary 3 is that parts of a conversation can be studied separately. For example, the opening section (turns 1 and 2 in the example in Section 2), the middle section (turns 3 through 6), and the closing section (turns 7 and 8), or the opening, the middle game, and the end game. (The corollary assumes perfect information, which makes its actual applicability very limited.)

Game theory is pertinent to analysing conversations. The corollaries above might be felt to be too general and abstract. That is because the model C is simple and only the basic theorems are employed for it. However, C can be made richer by specific conditions being added to it. This will make the application of game theory to conversations wider and more productive.

4 A More Realistic Model

In this section, I propose a more realistic model of conversations. It pertains to a family of games with incomplete information. In order to construct it, I apply findings from game programming.

4.1 The Definition of Model C′

Obviously, our model C is not very realistic, because it presupposes both perfect and complete information. Actually, a participant in an ordinary conversation does not know other speakers' utility functions. Further, future actions are typically not foreseeable at an earlier turn. A more realistic model, hence, should be in terms of games with incomplete information.

I define model C′ by introducing horizon and restriction of information into C.

Definition 2. C′ is a family of games which satisfies the following conditions:

- The set of players, the set of actions, utility functions, and the game tree are the same as Definition 1.
- Horizon: each player is given a horizon at each turn. She or he can see just the part of the game tree within the horizon from the node where she or he is (Figure 2).
- Restriction of information: neither player knows one other's utility function.[15]

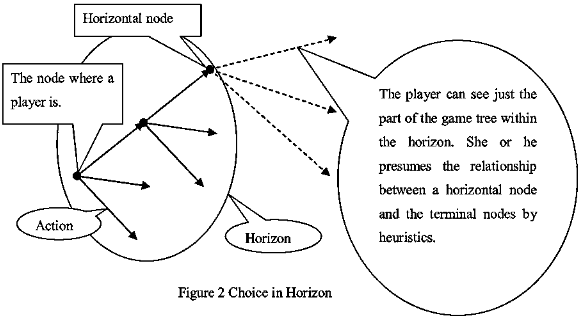

Horizontal node
The node where a player is.
Action
Horizon
The player can see just the part of the game tree within the horizon. She or he presumes the relationship between a horizontal node and the terminal nodes by heuristics.

Figure 2 Choice in Horizon

In a game of the model C′, a player cannot choose the subgame perfect equilibrium path as in C. Backwards induction is impossible for those whose foresight is limited to the range of the given horizons; besides, each player does not know the other player's utility function.

However, a player is able to find, in C′, the counterpart of a subgame perfect equilibrium path in C. More specifically, a player can do the following:

1. Estimate the utility value for each player at each horizontal node. One method will be calculating the expected utility, that is, summing all the products of the probability that the horizontal node leads to a terminal node and the utility value at the terminal node, both of the probability and the utility value being determined by the player's heuristics (Figure 3);[16]

[15] The games of C′ will be called perfect but incomplete information. See Harsanyi (1967, Section 1).

[16] Other methods will be possible, but how a horizontal node should be estimated in C ′ is not relevant to my other arguments.

2. Choose an action on the horizontal subgame perfect equilibrium path, which is based on the estimated utility values at the horizontal nodes, in that part of the game tree which is limited by the horizon.

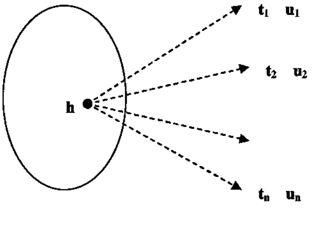


Figure 3 Evaluation of a Horizontal Node

Let t_1, ..., t_n be the terminal nodes accessible from horizontal node **h**.

The utility at **h** is estimated to be an expected utility, that is, the sum of the products of the probability of reaching a terminal node and the utility value at the node.

$$U(h) = \sum_{k=1}^{n} Pr(t_k / h)u_k$$

In other words, a player selects the best action at a turn only via the use of heuristics and the information limited by the horizon. Since a further horizon is given at a new turn, each player has new information there, and, taking advantage of it, decides the new path, which may be different from the one decided at the previous turn.

How much will a conversation of C' deviate from the subgame perfect equilibrium path (as in C)? It depends on the players' heuristics. The method of choosing an action in C' is not founded by formal theories but by heuristics, which may or may not be efficient. If the players' heuristics are efficient, their conversation path will be close to the equilibrium path. If not, their conversation will go far away from it, and surprise them with many unexpected windings.

4.2 Conversation and Game Programming

The situation for a participant in a conversation of the model C' is very close to that for a player of a game not in the sense used in game theory, but in the sense of, for example, chess. Apart from certain endgames, a player cannot consider all the positions in a play of chess. Moreover, the evaluation of a position which is not close to checkmate is bound to be uncertain. In fact, using the terminology of game programming, implementing 1 above is static evaluation and 2 is almost the same as minimax search.[17] Consequently, it is very plausible to think that techniques in game programming will be useful to study human conversations through C'.

One of the differences between game programming and game theory is the domain of a utility function. In the former, the domain is the set of the positions in a play of game, for instance,

[17] For classical explications of basic techniques in chess programming, see Shannon (1950a,b) and Turing et al. (1953), though the terms "static evaluation" and "minimax search" are not used in the literature yet. For a more recent one, see Levy & Newborn (1991). Minimax search is not exactly the same as to find the horizontal subgame perfect equilibrium. Minimax search is applied to zero-sum games, in which a player's positive payoff means the other player's negative one. However, conversation is a non-zero-sum game, in which win-win situations are possible. To this type of game, the notion of apparent subgame equilibrium is applicable, but minimax search is not without slight modification.

the arrangement of chess pieces on the chessboard. In the latter, it is the set of the sequences of actions, as in Definition 1 above.

Yet, they correspond to each other. Generally, an action can be seen as an operation on a given position. Thus, if the initial position is defined, a sequence of actions is equivalent to a sequence of positions, and each node of a game tree in game theory indicates a position in the game.[18] Thus, it is clear that game programming is related to the game-theoretic formulation of a game.

There is a difficulty for straightforwardly applying game programming to conversation, however. Conversation has nothing like the chessboard for chess. In conversation, the counterpart of the chessboard is the world with its history.[19] For example, A in a train can not only talk with B about things in the train, but also blame him for what he did in a foreign country 10 years ago or promise to vote for him in the next presidential election some years later if he is a candidate. This problem will be a kind of frame problem in artificial intelligence (McCarthy & Hayes, 1969).[20]

I think that, at least pragmatically, the problem can be solved. The reason is that, while human speakers also have the same problem, they, being of finite intelligence, solve it competently as shown in their fluent plays of conversation. It suggests that the difficulty above is not insurmountable. Finding human heuristics for solving conversations and devising the artificial method of their solution will be identical to the study of conversation.

Game theory will also be applicable to the games of the model C′. Schemes exist that analyse incomplete-information games. For instance, Bayesian games and, more specifically, signalling games will be useful, though how to install the notion of horizon in them is not yet clear.

5 Evaluation: the Strengths and Weaknesses of the Models

First, models C and C′ have the following strengths.

1. Values of utility functions are left open. As Definitions 1 and 2 show, the models themselves do not have specified utility functions. They can deal with a lying speaker and a disobedient hearer.

2. The structure of a whole conversation is apparent. It is expressed in the equilibrium path and the game tree. In addition, the models cover all parts of a conversation, from the opening section to the closing.

3. The models are general and application-independent. They are not found on any special hypothesis and do not appeal to any arbitrary postulates. Annexing appropriate conditions makes them richer and more suited to a study with a particular purpose.

[18] A difference remains, however. In many games like chess, the so-called game tree is, mathematically, not a tree but a directed graph, because it has different paths to the same node (Plaat et al., 1996). Especially chess has perpetual check and other loops. Game-theoretically, the same positions in this sense should be differentiated.

[19] Positions in a conversation game will also involve aspects of the social reality and institutional facts as well as natural or brute facts (Searle, 1994).

[20] How the problem of mutual knowledge occurs in the model C ′ is yet to be pursued. It does not seem to occur in chess and similar games.

4. Game theory can be fully applied. Models are completely constructed in the game-theoretic way. Theorems are applicable as shown in Section 3.2.

5. Models are independent from semantics (see Remark 4e). To put it another way, they can be adapted to any semantic theory.

6. Collaborations of various disciplines are possible, including game theory, microeconomics, game programming, artificial intelligence and pragmatics such as the theory of speech acts and discourse analysis.

Second, the models have the following weaknesses.

1. The models do not deal with utterance understanding. They have only actions of which forces are explicit as Definition 1 and Remarks 3 and 4f–h state. It might not be satisfactory for researchers on pragmatics after all, who focus on how the hearer interprets an utterance. Nevertheless, I argue that the models are useful also for the pragmatic study of utterance understanding. First, a sequence of actions is observationally more approachable than an utterance-interpretation pair. Second, the pair should be located in a total conversation to be studied precisely. Finally, an utterance understanding costs more or less. How much the hearer will pay for an utterance understanding depends on his utility function and other elements of the conversation game, or the whole conversation.

2. The models presuppose discrete time. Whereas they have one player's turn after the other's, actual conversations are carried on in continuous time. Overlapping and intervening are always possible in reality. Refining the game tree is one of the remaining problems.

6 Conclusion

Using game theory, two models of conversations were proposed. One is a family of games with perfect and complete information. The other is that of incomplete information. I argued that game theory and game programming are applicable, and stated the strengths and weaknesses of the models.

Some further research topics arise on conversation studies.

1. Analysing the distributive function of a conversation. In the example of Section 2, B, who was closer to the window, opened it for A, who benefited more from the action. The conversation produced a desirable result of the maximum benefit gained by the minimum cost. (Here is the impossibility of interpersonal utility comparison, though.) What mechanism does a conversation have and how does it work? Does it always bring out optimal results? In other words, is the situation after a conversation always better than before?

2. Studying human heuristics for conversation. A conversation is a complicated game, whereas the abilities of and resources for human beings are limited. How do they solve a conversation game? What heuristics do they have?

3. Building artificial intelligence for conversation. Computers can play games and beat humans. Can they play conversation games with humans? Since the aim of a conversation game is not to beat the opponent as in chess, what constitutes the condition of success for artificial intelligence in playing conversation?

4. Finding the rational conditions of speech acts. In the example in Section 2, A asked B to open the window, and B opened it. This combination of two acts is (part of) an equilibrium in terms of game theory. This means that a successful condition for performing a directive act (Searle, 1979, pp. 13–14) is for it to be an equilibrium together with the hearer's obeying act. This rational condition for successful performance of a speech act is radically different from the speech act theorists' concept of a successful condition. What is the rational condition of a speech act in concrete terms? In the case of directives, rational success conditions can be extremely perplexing. Suppose that the president of a company commands the vice-president to command an employee to order goods from a supplier. What conditions does her commanding need to satisfy in order to be successful? If they are too complex, what is the trick a hierarchical organisation should perform so that her command becomes manageable?

ACKNOWLEDGMENTS

The preparation of this paper was supported by the Japan Society for the Promotion of Science. I thank Syun Tutiya (Chiba University) and Tomoyuki Yamada (Hokkaido University) for their helpful comments of the draft versions of this chapter.

REFERENCES

Asher, N., I. Sher and M. Williams (2001). Game-theoretical foundations for Gricean constraints. In: *Proceedings of the 13th Amsterdam Colloquium* (R. van Rooy and M. Stokhof, eds.), pp. 31-37. Amsterdam.

Austin, J. L. (1975). *How to Do Things with Words* (2nd edition). Clarendon Press, Oxford.

Cohen, P. R. and H. J. Levesque (1985). Speech acts and rationality. In: *Proceedings of the 23rd Annual Meeting on Association for Computational Linguistics*, pp. 49-60. Chicago.

Davidson, D. (1980). Actions, reasons, and causes. In: *Essays on Actions and Events* (D. Davidson), pp. 3-19. Oxford University Press, Oxford.

Harsanyi, J. C. (1967). Games with incomplete information played by 'Bayesian' players. Part I: Basic model. *Management Science*, **14**, 159-182. (Reprinted in Kuhn 1997, pp. 216-246.)

Hashida, K. (1996). Issues in communication game. In: *Proceedings of the 16th International Conference on Computational Linguistics*, pp. 531-536. Copenhagen.

Holdcroft, D. (1979). Speech acts and conversation I. *The Philosophical Quarterly*, **25**, 125-141.

Kuhn, H. W. (1953). Extensive games and the problem of information. In: *Contributions to the Theory of Games II* (H. W. Kuhn and A. W. Tucker, eds.), pp. 193-216. Princeton University Press, Princeton. (Reprinted in Kuhn 1997, pp. 46-68.)

Kuhn, H. W. ed. (1997). *Classics in Game Theory*. Princeton University Press, Princeton.

Levy, D. and M. Newborn (1991). *How Computers Play Chess*. Freeman, New York.

Ludwig, K. (1997). The truth about moods. *Proto Sociology*, **10**, 19-66.

McCarthy, J. and P. J. Hayes (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, **4**, 463-502.

Nash Jr, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, **36**, 48-49. (Reprinted in Kuhn 1997, pp. 3-4.)

Nash Jr, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, **54**, 286-295. (Reprinted in Kuhn 1997, pp. 14-26.)

Parikh, P. (2001). *The Use of Language*. CSLI Publications, Stanford.

Parikh, P. (2006). Radical semantics: A new theory of meaning. *Journal of Philosophical Logic*, **35**, 349-391.

Plaat, A., J. Schaeffer, W. Pijls and A. de Bruin (1996). Exploiting graph properties of game trees. *13th National Conference on Artificial Intelligence* (Vol. 1), pp. 234-239. Portland.

Schegloff, E. A. and H. Sacks (1973). Opening up closings. *Semiotica*, **8**, 289-327.

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

Searle, J. R. (1979). A taxonomy of illocutionary acts. In: *Language, Mind, and Knowledge* (K. Gunderson, ed.), pp. 344-369. Minnesota Studies in the Philosophy of Science, **7**, University of Minnesota Press. (Reprinted in Searle, J. R. 1979. *Expression and Meaning: Studies in the Theory of Speech Acts*, pp. 1-29. Cambridge University Press, Cambridge.)

Searle, J. R. (1994). *The Construction of Social Reality*. Simon and Shuster, New York.

Selten, R. (1973). A simple model of imperfect competition, where 4 are few and 6 are many. *International Journal of Game Theory*, **2**, 141-201.

Selten, R. (1975). Reexamination of perfect equilibrium points in extensive games. *International Journal of Game Theory*, **4**, 25-55. (Reprinted in Kuhn 1997, pp. 317-354.)

Shannon, C. E. (1950). A chess-playing machine. *Scientific American*, **182**, 48-51.

Shannon, C. E. (1950). Programming a computer for playing chess. *Philosophical Magazine*, **41**, 256-275.

Turing, A. M., C. Strachey, M. A. Bates and B. V. Bowden (1953). Digital computers applied to games. In: *Faster Than Thought* (B. V. Bowden, ed.), pp. 286-310. Pitman, London.

von Neumann, J. and O. Morgenstern (1953). *Theory of Games and Economic Behavior* (3rd edition). Princeton University Press, Princeton.

This page intentionally left blank

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 9

SITUATIONS AND SOLUTION CONCEPTS IN GAME-THEORETIC APPROACHES TO PRAGMATICS

Ian Ross
University of Pennsylvania

If Bidirectional Optimality Theory (BiOT) is restricted to operating over lexical items (or even simple clauses), it is unable to account for certain scalar implicatures that are determined by larger contextual factors. In order to accommodate such cases in this framework, the relevant units of optimization must be multiclause sentences. If this step is taken, the predictions of BiOT and Games of Partial Information (GPIs) converge in the case we examine, although they remain distinct in the general case. Such robust context-dependent examples of scalar implicature show that adequate models cannot reduce such phenomena to a localized, lexical account.

1 Introduction

When scalar implicature triggers interact with each other in the greater context of an utterance, the standard accounts of scalar implicature can fail (for such accounts, see Horn 1972; Gazdar 1979; Hirschberg 1985). A representative example of the kind of utterances that scalar implicature theorists aim to explain is shown in (1)-(2).

(1) Some people like kale.

(2) Not all people like kale.

(2) is said to be implicated by the utterance of (1). The reasoning goes as follows: *all* is more informative than *some* (when the domain is nonempty, *all* entails *some*), and since the utterer chose *some*, he was not in a position to assert that all people like kale. These scalar implicature triggers are on the Horn scale ⟨all, some⟩ (Horn, 1972). While such reasoning works for simple examples like (1), matters are more complicated for utterances with interacting scalar implicature triggers. In these cases, what is most informative on a local level may not be what is most informative on a global level. The issue of locality in implicature theory (in the form of the effect that negative or downward entailing contexts have on implicature) has recently been discussed by Chierchia (2004) and Sauerland (2004) (who come to rather different conclusions), but neither

author's proposal addresses the cases of non-locality investigated here. In particular, we will examine cases in which sentences in the wider context affect the (non-)occurrence of implicature in sometimes startling ways. Cases which are quite removed from standard theories of implicature projection are along the lines of what Levinson (2000) has called "Gazdar's bucket".

BiOT (Dekker and van Rooy 2000, van Rooy 2004) and GPIs (Parikh, 2001) are two game-theoretic frameworks that build upon earlier work in implicature (Grice 1975, Levinson 1983) with intuitive new formalisms. They both apply the notion of utility maximization to implicature computation (among other topics), a notion which has been recognized in one form or another as important for decades but has long resisted the relatively transparent and precise treatments offered by these authors. BiOT extends traditional Optimality Theory (Kager, 1999) by adding and optimizing along a second dimension. BiOT's two dimensions are form and meaning (ranked by something like brevity or ease of utterance and informativeness respectively, rankings which will doubtless be improved upon as these phenomena are further examined), and winning candidates must be optimal with respect to each of them (unlike in OT syntax or semantics, in which only one dimension is optimized). The notion of form here corresponds to phonology as well as syntax and content, or meaning, corresponds to semantics. Note that the candidiates in the "form" column of BiOT differ from those in the corresponding column of classical OT. In classical OT, the highest-ranked form wins the prize of well-formedness, and is not ranked according to anything like brevity. In BiOT, the candidates are all well-formed and the decision to be made is how they should be mapped to meanings; a decision that does not seem to turn on classical OT constraints like *VOICEDCODA. Dekker and van Rooy (2000) cast BiOT in terms of strategic games.

GPIs are a different framework that did not grow out of Optimality Theory. In GPIs, one starts with a set of possible intended meanings (with associated probabilities). For each such meaning the utterer must choose an utterance (possibly ambiguous) to verbalize it and for each utterance the addressee must choose an interpretation for it (payoffs are assigned based on a number of efficiency-related factors including successful communication, markedness of forms, informativeness of utterance, and processing/production costs). This is represented as an extensive, rather than strategic, game. The solution concept used is simply that of Nash equilibrium. In GPIs, a similar form-meaning opposition manifests itself. For the speaker to use a less ambiguous utterance (to increase his payoff by constraining the choices of the addressee), he must usually utter a less brief expression (which will decrease his payoff since brevity is preferred). Solving for Nash equilibria is one way to find a balance between these opposing forces.

Both formalisms use Nash equilibria, but they are applied to different forms of strategies and games. Further work in the area may reveal the usefulness of different solution concepts (and forms of games). Perhaps BiOT may be fruitfully generalized to cover strategic games with imperfect information. Also, the probabilistic nature of communication might better be captured with the notions of correlated and mixed strategy Nash equilibrium in either of these formalisms. Other possibilities include eliminating actions that are not rationalizable or pursuing risk minimization strategies (assuming an opponent whose goal is to minimize their opponent's payoff). Lastly, evolutionarily stable strategies might be able to model phenomena in language change and acquisition.[1]

Dekker and van Rooy (2000, p. 240) state that, "It remains an open question how Parikh's approach relates to the one discussed in this paper". We will examine how they handle difficult cases of scalar implicature and shed light on this question in the process of doing so.

[1] [See the papers in this volume by Cecilia & Paolo Di Chio, Pelle Guldborg Hansen as well as Gerhard Jäger on the evolutionary accounts of language change and acquisition.]

2 DATA

BiOT has been used to explain phenomena in lexical pragmatics (Blutner 1998, 2004) by recasting Horn's *division of pragmatic labor* (Horn 1984; McCawley 1978) in a game-theoretic framework, but this lexically-based approach is unable to explain the full range of scalar implicature data. When applied to the scalar implicature trigger *some*, it predicts that the *pep* (form-meaning pair) ⟨"some," *some but not all*⟩ is optimal and the *pep* ⟨"some if not all," *some (possibly all)*⟩ is superoptimal (a weaker notion than optimality, see Dekker and van Rooy 2000), yielding the typical prediction for scalar implicature.

Such predictions, however, are not always accurate. Consider (3)-(4):

(3) Some of the girls and some if not all of the boys like kale.

(4) Some of the boys and some but not all of the girls like kale.

How the first instance of *some* is interpreted is dependent upon material modifying the second instance in both sentences. Intuitively, in (3) the *possibly all* reading is explicitly disambiguated, so the leftover *but not all* reading is assigned to the plain (i.e., without implicature reinforcement or cancelation) *some*. In (4) the *but not all* reading is explicitly disambiguated, so the leftover *possibly all* reading is assigned to the plain *some*. While (3) and (4) differ in what is said, they do not differ in what is communicated (i.e., what is said and implicated).

Uttering (3) yields the implicature in (5), which is predicted by classical treatments of scalar implicature. Uttering (4), however, does not yield the implicature in (6), which is what classical treatments would predict. The classical treatments are unable to take the wider context of the scalar implicature trigger in (4) into account.

(5) Not all of the girls like kale.

(6) Not all of the boys like kale.

Chierchia (2004) and Sauerland (2004) have made progress in developing more robust predictions for scalar implicature triggers in downward entailing (DE) environments and within the scope of other triggers, but (4) does not fall into either of these categories. The explicit reinforcement of the expected implicature of the second *some* implicitly *cancels* (or at least severely weakens) the expected implicature of the first *some*. If the first *some* in (4) was meant as *some but not all*, why was it not explicitly put that way, given that the utterer or (4) has demonstrated that he is willing to explicitly reinforce implicatures? Intending to communicate *some but not all* with the first *some* in (4) is unnecessarily confusing, since the reinforcement on the later *some* sets up a contrast in form that is not manifested in meaning.

GPIs are able to flexibly accommodate examples like (4) because they can take reinforcements and cancellations in the context into account. There are two equally efficient strategies (or action profiles) to exploit the semantic meaning of *some*, shown in (7). One could use plain *some* as *some but not all* and cancel this meaning with *if not all* when needed (as is done in 7a) or one could use plain *some* as *some (possibly all)* and cancel this meaning with *but not all* (as is done in 7b).

To fully specify a strategy, the utterer must choose an utterance for every initial situation and the addressee must choose an interpretation (meaning) for every utterance (note that the one-many mapping from utterances to interpretations makes this a game of imperfect information). Payoffs are assigned by tallying up the costs (functions of brevity and ease of utterance and interpretation) and benefits (functions of informativeness) as is also done in BiOT (see Parikh

2001 for initial thoughts on payoffs). Asymmetries in the costs of processing and production of utterances could lead to games besides those of pure coordination.

(7) a. Utterer: $s_{\underline{\text{not all}}}$ → “some”, $s_{\underline{\text{possibly all}}}$ → “some if not all”;
addressee: “some” → *some but not all*, “some if not all” → *some (possibly all)*.

b. Utterer: $s_{\underline{\text{not all}}}$ → “some but not all”, $s_{\underline{\text{possibly all}}}$ → “some”;
addressee: “some but not all” → *some but not all*, “some” → *some (possibly all)*.

Both strategies convey the same meanings with expressions that are equally prolix. Classical implicature theory designates (7a) as the only available strategy. In practice, cancelations (e.g., *or all*) greatly outnumber reinforcements (e.g., *but not all*), so (7a) is the strategy we are given more evidence for in natural language (Ross, 2004). With (7a) as a strategy, *some but not all* would not be uttered since plain *some* already conveys that meaning. The same goes for the strategy (7b) and the absence of *some if not all*. So using implicature reinforcements or cancelations is a way to signal which of the two strategies one is following.

If reinforcements happened to outnumber cancelations, we would expect a different pattern of implicatures (namely that of 7b), but classical accounts have nothing to contribute on this point. These accounts maximize communicative efficiency locally: plain *some* gets the standard *but not all* implicature. Such accounts do not, however, globally maximize communicative efficiency. This can be observed by examining (8).

(8) Some if not all of the boys and some but not all of the girls like kale.

According to classical accounts, in order to communicate what uttering (4) does in actuality, one would need to utter (8), which is unnecessarily more prolix than (4), or utter (3), which differs syntactically from (4) even more than (8) does. By allowing both (7a) and (7b) as strategies, we can correctly predict the meaning and efficiency of (4).

The GPI that yields the correct implicatures for (3) and (4) is shown in Figure 1, with descriptions of the variables in Figure 2. This is an extension of the game in Parikh (2001, p. 94) (with the same payoff scheme and initial probabilities). Instead of operating on one scalar implicature trigger, we operate on two (specifically, the Cartesian product of two of them). Each situation/intended meaning has a corresponding probability. Given a situation, the utterer must then choose an utterance. In each of the four possible situations, the utterer may choose from among five (different) utterances whose conventional meaning is consistent with the situation. Given one of these utterances, the addressee chooses from among one to four interpretations. Payoffs are then assigned to each such sequence. Solution candidates are probability-weighted sums of such payoffs. The four Nash equilibria (all with the same payoff, shown below) of this game are shown in (9). Payoffs from interpretations that contribute to at least one Nash equilibrium are boxed. Note that this example demonstrates the collective nature of the process of arriving at a solution. In each of the solutions, it is not the case that for each situation an utterance and interpretation are chosen such that the payoffs are maximized. Rather, what is maximized is the weighted sum of the payoffs from each situation. For example, in (9b) the constituent payoffs for (S,N) and (N,S) are each (11,12) instead of (13,14) because while (13,14) would maximize the payoffs for (S,N) and (N,S), it would not lead to a global Nash equilibrium.

(9) a. $((S,S) \rightarrow \delta, (S,N) \rightarrow \beta, (N,S) \rightarrow \epsilon, (N,N) \rightarrow \alpha; \delta \rightarrow d, \beta \rightarrow c, \epsilon \rightarrow b, \alpha \rightarrow a)$

b. $((S,S) \rightarrow \delta, (S,N) \rightarrow \eta, (N,S) \rightarrow \epsilon, (N,N) \rightarrow \alpha; \delta \rightarrow d, \eta \rightarrow c, \epsilon \rightarrow b, \alpha \rightarrow a)$

c. $((S,S) \rightarrow \delta, (S,N) \rightarrow \beta, (N,S) \rightarrow \gamma, (N,N) \rightarrow \alpha; \delta \rightarrow d, \beta \rightarrow c, \gamma \rightarrow b, \alpha \rightarrow a)$

d. $((S,S) \rightarrow \delta, (S,N) \rightarrow \eta, (N,S) \rightarrow \gamma, (N,N) \rightarrow \alpha; \delta \rightarrow d, \eta \rightarrow c, \gamma \rightarrow b, \alpha \rightarrow a)$

$$.49(2,3) + .21(11,12) + .21(11,12) + .09(28,29) = (8.12, 9.12).$$

An example of a BiOT analysis of scalar implicature is shown in Figure 3. With two possible meanings and three possible forms, we have six combinations. One of them is unavailable (shaded in the figure) because the conventional meaning of the form (i.e., what is said) is not compatible (and is in fact logically contradictory in this case) with the pragmatic meaning (i.e., what is communicated) it is matched with. The leftward-pointing arrows show that, given a form, the more informative *but not all* meaning is preferred to the *if not all* meaning. The downward-pointing arrows show that the briefer *some* is preferred to the more prolix *some but not all* and *some if not all*. Since there are no outgoing arrows from the *pep* ⟨ "some", *some but not all*⟩, it is optimal. Since no outgoing arrows of the *pep* ⟨ "some if not all", *some (possibly all)*⟩ point to an optimal (or superoptimal) *pep* (see Dekker and van Rooy 2000 for details), it is superoptimal.

The predictions made in Figure 3 coincide with those of classical implicature theory. However, if one takes a more expansive view of *peps*, different results can be derived. Specifically, rather than candidate forms being composed of individual scalar implicature triggers (and their attendant reinforcements/cancelations), one might consider making entire sentences with multiple scalar implicature triggers candidate forms. If we take sentences like (3) and (4) as our forms, we can derive our theoretical indifference between them (they will both be classified as superoptimal with no ranking between them). However, expanding our forms along these lines will not result in the interpretational indifference of sentences like (10)—*but not all* readings will still be preferred on grounds of informativeness. This is in contrast to the treatment within GPIs, in which the interpretation of such sentences crucially depends on the initial probabilities of the situations/intended meanings. In setting these probabilities, we are making assumptions analogous to those made in computational linguistics about the prior probabilities of words or conditional probabilities of a part of speech given a word. In practice, probabilities like these are in the eye of the beholder. If they happen to be divergent enough, a Bayesian game could be employed to model the player's different beliefs about the underlying probabilities.

(10) Some of the boys and some of the girls like kale.

The fact that changing the "level" of *peps* affects theories' predictions raises an important question. At exactly what level should pragmatic games (e.g., BiOT or GPIs) be played? We have seen examples where they need to be played on complex sentences—anything smaller would miss contextual cues needed to compute the solution. The problem is similar to the one that faces "Gazdar's bucket". In this procedure, different types of implicatures (clausal, scalar, etc.) are added to the context in order according to their kind, and implicatures that conflict with the context are discarded. If implicatures are added to the context as soon as they are encountered, Gazdar's ordering mechanism will not have an effect, and if implicatures are added only after a discourse is over (to make sure the ordering is done properly), then there is no possible way for human memory to store all the information needed for the procedure. In (4) one is likely to assume that the first *some* has the standard implicature, and only later rescind this judgment (upon seeing the second, reinforced *some*). In this particular domain and many others, non-monotonic reasoning appears inevitable (Wainer, 1991).

When multiple successive games are being played, there is also the issue of how they are related. It is conceivable that one may switch from one solution to another during a discourse.

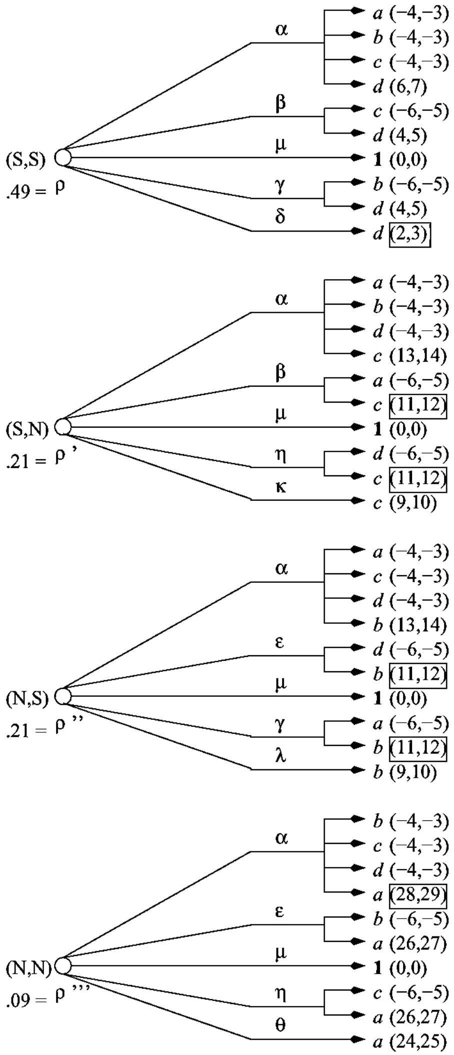


Figure 1: A GPI to describe parallel scalar implicature

situations (speaker)
(S,S) : some (possibly all) of the girls and some (possibly all) of the boys like kale
(S,N) : some (possibly all) of the girls and some but not all of the boys like kale
(N,S) : some but not all of the girls and some (possibly all) of the boys like kale
(N,N) : some but not all of the girls and some but not all of the boys like kale

sentences
α : "some of the girls and some of the boys like kale"
β : "some if not all of the girls and some of the boys like kale"
μ : —
γ : "some of the girls and some if not all of the boys like kale"
δ : "some if not all of the girls and some if not all of the boys like kale"
η : "some of the girls and some but not all of the boys like kale"
κ : "some if not all of the girls and some but not all of the boys like kale"
ε : "some but not all of the girls and some of the boys like kale"
λ : "some but not all of the girls and some if not all of the boys like kale"
θ : "some but not all of the girls and some but not all of the boys like kale"

induced information sets (adressee)
α : {(S,S),(S,N),(N,S),(N,N)}
β : {(S,S),(S,N)}
μ : {(S,S),(S,N),(N,S),(N,N)}
γ : {(S,S),(N,S)}
δ : {(S,S)}
η : {(S.N),(N,N)}
κ : {(S,N)}
ε : {(N,S),(N,N)}
λ : {(N,S)}
θ : {(N,N)}

propositions
a: that some but not all of the girls and some but not all of the boys like kale
b: that some but not all of the girls and some (possibly all) of the boys like kale
c: that some (possibly all) of the girls and some but not all of the boys like kale
d: that some (possibly all) of the girls and some (possibly all) of the boys like kale
1: *true*

Figure 2: Key to game in Figure 1

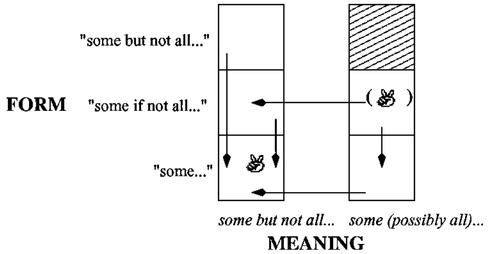


Figure 3: Implicature in Bidirectional OT

This can be accounted for at least partially by state. One can estimate the probabilities of situations/intended meanings by tabulating unambiguous declarations of such (an utterance of "some but not all" presumably increases the estimated probability for the intended meaning *some but not all* in the future) and noting the most salient or recent instances. Another approach would be to treat these successive games as an evolutionary game.

3 Differences between BiOT and GPIs

One chief difference between BiOT and GPIs is their solution concepts. Although both formalisms use Nash equilibria, they play a different role in each. A solution to a BiOT tableau is a *pep*. If multiple *peps* are (super)optimal, then there are multiple solutions. A solution to a GPI is a strategy profile over multiple situations for speaker and addressee. The speaker chooses an utterance for each situation and the hearer chooses a hypothesized meaning for each utterance (only the speaker knows the actual situation). This strategy profile results in a *set* of *peps*. So while it is individually determined for each *pep* whether it is a solution in the BiOT tableau, for GPIs the solution set of *peps* is collectively determined. Rather than optimizing individual *peps*, the aggregate value of a set of related *peps* is optimized in GPIs, which is a more global form of optimization. The latter type of optimization predicts a language more efficient than the former, and linguistic data confirm that this additional efficiency is present. Under the BiOT prediction, the plain *some* in (6) should be interpreted as *some but not all* (and the *some but not all* should have never been explicitly disambiguated as such) and the only way to express *some (possibly all)* is to explicitly disambiguate it. The fact remains that *some but not all* is explicitly disambiguated in (6) and this has consequences for the interpretation of the plain *some* in (6).

Another difference between BiOT and GPIs is their representation of different situations. In BiOT, situations are only implicitly represented (as entries in the meaning dimension) and never quantified over—no distinction is made between intended meanings (of the utterer) and assigned meanings or interpretations (of the addressee). As a consequence, the temporal element of communication is removed.

In GPIs, situations are explicitly represented, assigned probabilities, and solutions are over a weighted combination of *all* situations. BiOT is only sensitive to markedness of forms and

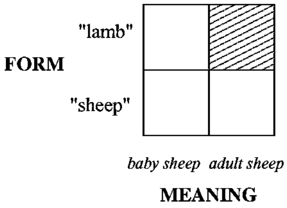


Figure 4: BiOT tableau for lamb/sheep

meanings, remaining blind to usage frequencies and their implications for *peps*. GPIs are able to capture patterns of usage by assigning probabilities to situations. If explicit instances of *if not all* outnumber those of *but not all* (or outweigh them in prominence) then plain *some* will presumably be interpreted as *some but not all*, otherwise it will be interpreted as *some (possibly all)*. This sensitivity to usage and flexible predictions that follow from it can also provide an account of various diachronic pragmatic phenomena. For example, *lamb* is the term for immature sheep and since there is no comparable term for adult sheep, *sheep* is used for adult sheep. What began as implicature has since been lexicalized and the term is now considered polysemous. GPIs predict this straightforwardly but BiOT lacks such an account. In Figure 4 we see that there are no preferences between any of the *peps*. "Lamb" and "sheep" are equally brief and the concepts *baby sheep* and *adult sheep* are equally informative. We want to be able to say that uttering "sheep" implicates adult sheep, but BiOT gives us no reason to do so.

One could go a step further within GPIs, which distinguish between intended and interpreted meaning, and allow for their arbitrary intersection. This would then split communicative meaning into four categories: that which was intended and interpreted, that which was intended but not interpreted, that which was not intended but interpreted nonetheless, and that which was neither intended nor interpreted. This intersective view of communication in some sense transcends the utter-centered treatments of Grice and the addresser-centered treatments of Relevance theorists (Sperber & Wilson, 1995).

On the other hand, the GPI in Figure 5 yields (11a) as a solution with a payoff shown in (12a), beating out the rival strategy (11b), which has a lower payoff shown in (12b). Payoffs and probabilities for the GPI in Figure 5 are adjusted from Parikh's GPI for scalar implicature. The probabilities are set equal to each other since there is no strong asymmetry in informativeness between the concepts *baby sheep* and *adult sheep* like there was for *some but not all* and *some (possibly all)* (although on a detailed level one could claim that *baby* is more informative since infancy is only a small portion of an animal's life, but this would only help our case so we will ignore it here). Also, there is no informational gap between the propositions, so payoffs are adjusted accordingly. The asymmetry that leads to (11a) as a solution is the difference in brevity between unambiguous utterances expressing the concepts of *baby sheep* (lamb) and *adult sheep* (adult sheep), which is intuitively why *sheep* picks up the meaning it does.

(11) a. $(s \rightarrow \mu, s' \rightarrow \varphi; \mu \rightarrow p, \varphi \rightarrow l)$

b. $(s \rightarrow \varphi, s' \rightarrow \gamma; \varphi \rightarrow p, \gamma \rightarrow l)$

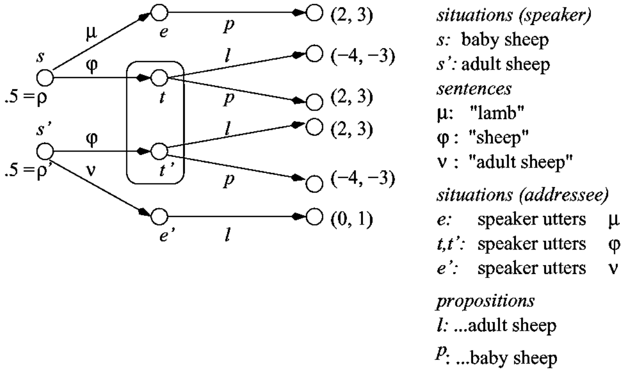


Figure 5: GPI for lamb/sheep

(12) a. $.5(2,3) + .5(2,3) = (2,3)$

b. $.5(2,3) + .5(0,1) = (1,2)$.

4 Mapping Between BiOT Tableaus and GPIs

We can demonstrate a key difference in the solution concepts of BiOT and GPIs by mapping a BiOT tableau into a GPI and vice versa. Examine Figure 6. Here, assume a is always preferred to b and α is always preferred to β and that the conventional meanings of α and β are consistent with both a and b. In this tableau there is one optimal *pep* ($\langle\alpha, a\rangle$) and one superoptimal *pep* ($\langle\beta, b\rangle$). Each column of this tableau can be mapped to an extensive game that comprises part of the total GPI. The meaning a is mapped to the situation a. Since a is part of the *pep* $\langle\alpha, a\rangle$, the mapping results in the utterer in this part of the GPI choosing to utter α. As a consequence, α is interpreted as a. The same goes for b and β. Since meanings are not weighted in the BiOT tableau, we attach equal weights to the subgames of the GPI anchored by these meanings (i.e., .5 to each). This is still not enough information to fix a unique solution to the corresponding GPI. Why? Because the BiOT tableau does not tell us *how* preferred the chosen *peps* are—we are only given an ordering. Let us attach a numerical value to each *pep* in the tableau (and assume that the values of $\langle\alpha, b\rangle$ and $\langle\beta, a\rangle$ are equivalent). If half the sum of the values of $\langle\alpha, a\rangle$ and $\langle\beta, b\rangle$ is greater than half the sum of the values of $\langle\alpha, b\rangle$ and $\langle\beta, a\rangle$, then not only do we have enough information to determine a unique solution to the corresponding GPI ((13) is the solution of the GPI in Figure 7), but said solution also corresponds to the solution of the BiOT tableau.

(13) $(a \rightarrow \alpha, b \rightarrow \beta; \alpha \rightarrow a, \beta \rightarrow b)$.

Now consider the case in which half the sum of the values of $\langle\alpha, a\rangle$ and $\langle\beta, b\rangle$ is less than half the sum of the values of $\langle\alpha, b\rangle$ and $\langle\beta, a\rangle$ (if they are equal, then both solutions are available). A

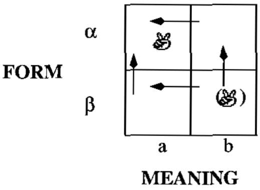


Figure 6: Bidirectional OT tableau

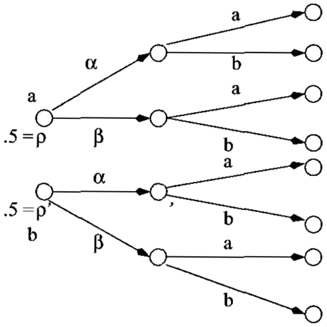


Figure 7: GPI corresponding to BiOT tableau in Figure 6

concrete example is shown in Figure 8. What is the corresponding GPI? It is the same one shown in Figure 7 with one important change: the solution is different.

Recall that the solution is calculated as a weighted sum of payoffs, one for each intended meaning. The substrategy $(a \rightarrow \alpha; \alpha \rightarrow a)$ yields a payoff of 10 (for both players), substrategies $(a \rightarrow \beta; \beta \rightarrow a)$ and $(b \rightarrow \alpha; \alpha \rightarrow b)$ yield payoffs of 8 each, and substrategy $(b \rightarrow \beta; \beta \rightarrow b)$ yields a payoff of 0. The two competing strategies are shown in (14) and their payoffs are shown in (15) respectively. (14b) is the only Nash equilibrium because the utterer could profitably unilaterally deviate from (14a) to (14b) by switching α and β.

(14) a. $(a \rightarrow \alpha, b \rightarrow \beta; \alpha \rightarrow a, \beta \rightarrow b)$

b. $(a \rightarrow \beta, b \rightarrow \alpha; \alpha \rightarrow b, \beta \rightarrow a)$

(15) a. $.5(10,10) + .5(0,0) = (5,5)$

b. $.5(8,8) + .5(8,8) = (8,8)$.

In this case (with the values in Figure 8), the solutions to the corresponding BiOT tableau and GPI are not in correspondence. The solution to the GPI is (14b), but the solution to the BiOT tableau is equivalent to (14a). While the solution to the GPI crucially depends on the values

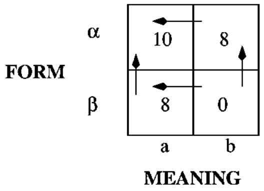


Figure 8: Bidirectional OT tableau with values

assigned to *peps* in the tableau, the solution to the BiOT tableau only depends on the *ordering* of the *peps*, and no amount of changing the values without changing the ordering of the preference relation (under the constraint that the weighted sum of the values of ⟨α, *b*⟩ and ⟨β, *a*⟩ is greater than the weighted sum of the values of ⟨α, *a*⟩ and ⟨β, *b*⟩) will modify the solution of the BiOT tableau to the equivalent of (14b).

5 Conclusion

While both BiOT and GPIs are improvements upon classical accounts, the local character of optimization in BiOT (optimizing *peps* instead of sets of them) results in its inability to explain different possible solutions to a class of games examined here. Namely, those involving scalar implicature triggers that are not only equally optimal from a global perspective but also empirically attested for in sentences like (6) if we restrict *peps* to lexical items or simple phrases. The key difference that gives GPIs the flexibility to account for examples like (4) and Figure 5 is that Nash equilibria are calculated collectively over a set of peps, which allows contextual cues that materially affect implicature, be it in the form of an implicature reinforcement or markedness asymmetry between unambiguous candidate forms ("lamb" vs. "adult sheep"), to be taken into account.

Acknowledgements

I would like to thank the audience of the 2004 Prague International Colloquium for their insightful questions and other constructive feedback, as well as the anonymous reviewer who provided detailed comments which made this revision much clearer. I would also like to thank Prashant Parikh and my advisor Robin Clark for getting me thinking about game-theoretic applications to natural language and answering innumerable questions along the way. All remaining errors are my own.

References

Blutner, R. (1998). Lexical pragmatics. *Journal of Semantics*, **15**, 115-162.

Blutner, R. (2004). Pragmatics and the lexicon. In: *Handbook of Pragmatics* (L. Horn and G. Ward, eds.), pp. 488-514. Blackwell, Oxford.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In: *Structures and Beyond* (A. Belletti, ed.), pp. 39-103. Oxford University Press, Oxford.

Dekker, P. and R. van Rooy (2000). Bi-directional optimality theory: An application of game theory. *Journal of Semantics*, **17**, 217-242.

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form.* Academic Press, New York.

Grice, H. P. (1975). Logic and conversation. In: *Syntax and Semantics 3: Speech Acts* (P. Cole and J. Morgan, eds.), pp. 41-58. Academic Press, New York.

Hirschberg, J. (1985). *A Theory of Scalar Implicature.* Dissertation, University of Pennsylvania.

Horn, L. (1972). *On the Semantic Properties of the Logical Operators in English*, Dissertation, Indiana University Linguistics Club, Bloomington.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In: *Meaning, Form, and Use in Context: Linguistic Applications* (D. Schiffrin, ed.), pp. 11-42. Georgetown University Press, Washington, D.C.

Jäger, G. (2000). Some notes on the formal properties of bidirectional optimality theory. *Journal of Logic, Language and Computation*, **11**, 427-451.

Kager, R. (1999). *Optimality Theory.* Cambridge University Press, Cambridge.

Levinson, S. (1983). *Pragmatics.* Cambridge University Press, Cambridge.

Levinson, S. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature.* MIT Press, Cambridge, Mass.

McCawley, J. (1978). Conversational implicature and the lexicon. In: *Syntax and Semantics 9: Pragmatics* (P. Cole, ed.), pp. 245-259. Academic Press, New York.

Parikh, P. (2001). *The Use of Language.* CSLI Publications, Stanford.

van Rooy, R. (2004). Relevance and bidirectional optimality theory. In: *Optimality Theory and Pragmatics* (R. Blutner and H. Zeevat, eds.), pp. 173-210. Palgrave Macmillan, New York.

Ross, I. (2004). Pragmatic tagging for scalar implicatures. In: *Proceedings of KONVENS 7 Advanced Topics in Modeling Natural Language Dialog Workshop* (H. Horacek and T. Kruijff-Korbayova, eds.), pp. 15-22. Vienna.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, **27**, 367-391.

Sperber, D. and D. Wilson (1995). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, **27**, 367-391.

Wainer, J. (1991). *Uses of Nonmonotonic Logic in Natural Language Understanding: Generalized Implicatures.* Dissertation, The Pennsylvania State University.

This page intentionally left blank

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 10

AN INTRODUCTION TO EQUILIBRIUM SEMANTICS FOR NATURAL LANGUAGE

Prashant Parikh
University of Pennsylvania

Robin Clark
University of Pennsylvania

This paper presents an application of game theory and situation theory to linguistics and the philosophy of language. In it, we introduce equilibrium semantics, a framework for the study of meaning that combines semantics and pragmatics into a single discipline or, alternatively, builds *use* into *reference* at the "ground" level so that there may be no need for a separate discipline of pragmatics. This paper derives most proximately from Parikh & Clark (2007), where we first described equilibrium semantics in the context of a theory of definite descriptions. That paper came out of Parikh (2006) and earlier work of Parikh's on game-theoretic semantics, most notably Parikh (2001). We start with a brief discussion of semantics and pragmatics.

1 Semantics and Pragmatics

The study of meaning has been schizophrenic for much of the twentieth century, initially in the philosophy of language and subsequently in linguistics. So-called ideal language philosophy, first developed by Frege, Russell, and the early Wittgenstein, focused largely on the relation of reference between language and world by seeing natural language via the lens of formal language, almost completely ignoring the dimension of use.[1] Reacting to this, the so-called ordinary language philosophers, the later Wittgenstein, Austin, and Grice, focused exclusively on the relation of use, abstracting almost completely from the details of reference.

This pernicious split in the approach to meaning was carried over into modern semantics and pragmatics, the former discipline concerned largely with its referential aspect and largely formal and conventional, the other largely with its use-related or communicative aspect and largely informal and inferential. Montague Grammar and its variants represent perhaps the most notable successes of the former within linguistics; there is less consensus on the pragmatics front.

[1] The often awkward facts of use were treated as a kind of defect that would be removed by *idealizing* language.

This clean separation has been questioned by a variety of researchers, especially in recent times, starting with Grice (1989) himself, but also by Barwise & Perry (1983), Recanati (2004) and the Relevance Theorists (Wilson & Sperber, 1986), and also of course by the present authors. However, all these researchers, except earlier work by Parikh, have generally accepted the need for this bicameralism, but have argued for ways of "intertwining" the two dimensions by, for example, considering so-called "primary" and "secondary" pragmatic processes (see Recanati 2004).

We have argued earlier in Parikh (2006) and Parikh & Clark (2007) and will show conclusively in this paper that the split is unnecessary and that a unified theory of meaning is possible, and we will also show how this unified theory makes it feasible in principle to literally *compute* the meaning of any utterance from first principles, given access to the ambient data which serve as inputs to the theory!

It should perhaps be pointed out that to the best of our knowledge no other existing framework, either in semantics or pragmatics, comes at all close to the level of mathematical detail we will provide in equilibrium semantics. Since the proof of the pudding is in the eating, we now turn to the building blocks of our theory in the context of a non-trivial example.

2 An Example

Assume there is a heated discussion at the Global Astronomical Society one day in April 2006 about whether Pluto is a planet. In this context, consider the following sentence:

(1) John saw a planet. (φ)

Assume $\mathcal{A}$ utters φ to $\mathcal{B}$ in the utterance situation u.

The framework we will describe will be used to give a detailed derivation of the literal meaning or content of this utterance from first principles.

3 Situation Theory

Situation theory is a theory of information originally developed by Barwise (1989). Its key insight is that much information is always available and representable only partially. We present our version of it here—just the parts we need.

The world, itself a situation, consists of smaller parts that are situations, collections of individuals standing in relations. These form the basis for more abstract objects called variables, parameters, and types. The collection of all these entities is called $\mathcal{O}$.

For example, it may be that John, denoted by a_1, saw, designated by P^{see}, a planet, denoted by a_2, at a location l and a time t preceding the time of utterance t_u; this item of information would be written as a tuple $\langle\langle\, P^{see};\ a_1;\ a_2;\ l;\ (t \mid t \prec t_u)\,\rangle\rangle$. It is possible to represent this information partially by omitting one or more arguments, all the way down to the empty tuple. Each such tuple is called an infon; $\mathcal{I}$ denotes the subset of $\mathcal{O}$ containing all the infons.

Situations, the next type of entity in $\mathcal{O}$, are just collections of infons. The relation between a situation and an infon that holds in it is written $s \models \sigma$ or $\sigma \in s$, and is described by saying that s supports σ or σ holds in s.

Parameters are indeterminates or variable-like placeholders for any of the entities above and are denoted by $\dot{a}$, $\dot{R}$, $\dot{\sigma}$, and $\dot{s}$ for parameters involving individuals, relations, infons, and situations respectively. Types, like the type of object that is, say, a planet, are intuitively what is obtained when some individual or property is considered generically. Types are always relative to some situation and so are written $[\, \dot{x} \mid s \models \sigma(\dot{x}) \,]$ where s is the relevant *grounding* situation.

"Predicative individuals" are one or more individuals satisfying a property or relation, written $(\, x \mid s \models \sigma(x) \,)$. A predicative parameter is formed with a parameter instead of a variable, which secures an indeterminate reference to an object satisfying a property or relation. This is written $|\, \dot{x} \mid s \models \sigma(\dot{x}) \,|$.

Finally, an object can be formed by lambda abstraction and is written $\hat{x}(\, y \mid \sigma(\dot{x}, y) \,)$, where the variables can refer to any appropriate object.

While infons may be operated on using $\vee$ and $\wedge$ or quantified over with $\exists$ and $\forall$ to form compound infons in the usual way, it is also possible to define a special associative operation $\oplus$ of unification or merging that makes $\mathcal{I}$ a monoid when the empty infon is included as the identity.

The binary operation involves at least two parts which are best presented informally by example:

1. Merging: $\langle\!\langle\, a_1 \,\rangle\!\rangle \oplus \langle\!\langle\, P^{see} \,\rangle\!\rangle \oplus \langle\!\langle\, l;\ t \,\rangle\!\rangle = \langle\!\langle\, P^{see};\ a_1;\ l;\ t \,\rangle\!\rangle$
2. Hat conversion: $\langle\!\langle\, \hat{P}\langle\!\langle\, \dot{P};\ x;\ t \,\rangle\!\rangle \,\rangle\!\rangle \oplus \langle\!\langle\, P_1^{planet} \,\rangle\!\rangle = \langle\!\langle\, P_1^{planet};\ x;\ t \,\rangle\!\rangle$.

When the context is clear, it is possible to write simply $\sigma\tau$ instead of $\sigma \oplus \tau$ for the sum of two infons.

4 THE FOUR CONSTRAINTS

4.1 THE SYNTACTIC CONSTRAINT

The sentence φ is made up of individual words and may be represented as $\omega_1 \circ \omega_2 \circ \omega_3 \circ \omega_4$ (where $\circ$ is an associative concatenation operation on the set $\mathcal{L}$, which contains all the words, phrases, and sentences of the language) or more simply as $\omega_1\omega_2\omega_3\omega_4$ (where the operation symbol is left implicit). Clearly, $(\mathcal{L}, \circ)$ is also a monoid.

Our example φ has the following single parse tree:

$$[_{S}\, [_{NP}\, \omega_1\,] \circ [_{VP}\, [_{V}\, \omega_2\,] \circ [_{NP}\, [_{Det}\, \omega_3\,] \circ [_{N}\, \omega_4\,]\,]\,]\,]\,. \tag{2}$$

We call the set of syntactic structures for the sentences of $\mathcal{L}$ the *syntactic constraint* and label it **S**.

4.2 THE CONVENTIONAL CONSTRAINT

We assume that every word in $\mathcal{L}$ is associated with a *conventional meaning* which is either a property or relation. The conventional meaning of JOHN is denoted by P^{John} and is simply the property "is named John". The verb SAW has many conventional meanings, but we will restrict our attention to just one (the relation "to discern visually") to keep things simple and denote it by P^{see}. We assume A also has just one conventional meaning. We will soon see what it is. In the current astronomical context, PLANET has two conventional meanings that might be denoted P_1^{planet} and P_2^{planet}, the first including Pluto, the second excluding Pluto. We shall often refer to these conventional associations of words as a *conventional map.*

As the term suggests, conventional meanings are conventional, that is, they are the sorts of meanings that can largely be found in a dictionary.[2] They are also independent of the utterance situation.

We call the set of conventional meanings or "senses" for the words in $\mathcal{L}$ the *conventional constraint* and label it **C**.

4.3 The Informational Constraint

We assume next that given a set of conventional maps for a word in $\mathcal{L}$ and given an utterance of the word (typically in the context of a sentence), we can map the word's conventional meanings into one or more infons that represent the *possible* contents or meanings of the word uttered. We call this the *informational map* and all of them together the *informational constraint* and label it **I**.

We can string the conventional and informational maps together for the words in φ as follows.

1. JOHN:

   (a) Referential Use: $\omega_1 \longrightarrow P^{\omega_1} \xrightarrow{u} \langle\langle\, (\, x \mid (\, r_1 \models \langle\langle P^{\omega_1};\, x \rangle\rangle\,) \;\wedge\; (\, \forall y (r_1 \models \langle\langle P^{\omega_1};\, y \rangle\rangle) \Longrightarrow y = x\,)\,)\,\rangle\rangle = \sigma_1$.

   Note: Here the resource situation r_1 picks out the individual John. If there is more than one John, we would have corresponding resource situations for each John.

2. SAW:

   (a) Predicative Use: $\omega_2 \longrightarrow P^{\omega_2} \xrightarrow{u} \langle\langle P^{\omega_2};\; (t \mid t \prec t_u) \rangle\rangle = \sigma_2$.

   Note: Here the relation P^{ω_2} for SAW is more or less mapped into itself. The tense gets mapped into a temporal argument.

3. A:

   (a) Referential Use: $\omega_3 \longrightarrow P^{\omega_3} \xrightarrow{u} \langle\langle \hat{P}(\, x \mid (r_3 \models \langle\langle \dot{P};\, x \rangle\rangle\,) \;\wedge\; (\, \forall y (r_3 \models \langle\langle \dot{P};\, y \rangle\rangle) \Longrightarrow y = x\,)\,)\,\rangle\rangle = \sigma_3$.

   Note: Like the definite article, A has several uses. We give examples of four of the five uses of A as they would occur when concatenated with a noun in a sentence.

   Example: "John saw a planet" (used to indicate a particular planet but without access to a resource situation—the resource situation r_3 is not available to the addressee, unlike the case of definite descriptions, and so the planet remains unidentifiable).

   (b) Generic Use: $\omega_3 \longrightarrow P^{\omega_3} \xrightarrow{u} \langle\langle \hat{P}[\, \dot{x} \mid r_3' \models \langle\langle \dot{P};\, \dot{x} \rangle\rangle\,]\,\rangle\rangle = \sigma_3'$.

   Note: Here r_3' is the grounding situation for the type.

   Example: "A planet is an astronomical body" (where the description is used to refer generically to a type of object).

   (c) Predicative Use: $\omega_3 \longrightarrow P^{\omega_3} \xrightarrow{u} \langle\langle \hat{P}(\, \dot{P}\,)\,\rangle\rangle = \sigma_3''$.

   Note: The difference between the predicative uses of THE and A does not appear to lie in the contents expressed but possibly in their presuppositions.

   Example: "Pluto is a planet" (where it picks out just the property of being a planet).

[2] Largely, but not entirely. See Parikh & Clark (2007) for further discussion.

   (d) Indeterminate Use: $\omega_3 \longrightarrow \mathsf{P}^{\omega_3} \xrightarrow{u} \langle\!\langle\, \hat{\mathsf{P}} |\, \dot{x} \,|\, \langle\!\langle\, \dot{\mathsf{P}};\ \dot{x} \,\rangle\!\rangle \,|\, \rangle\!\rangle = \sigma_3'''$.
   Example: "Choose a planet you will write a paper on" (where there is an indeterminate reference to a planet).

4. PLANET:

   (a) Predicative Use: $\omega_4 \longrightarrow \mathsf{P}_1^{\omega_4} \xrightarrow{u} \langle\!\langle\, \mathsf{P}_1^{\omega_4} \,\rangle\!\rangle = \sigma_4$.
   Note: This refers to the inclusive sense of PLANET, that is, the sense that includes Pluto.

   (b) Predicative Use: $\omega_4 \longrightarrow \mathsf{P}_2^{\omega_4} \xrightarrow{u} \langle\!\langle\, \mathsf{P}_2^{\omega_4} \,\rangle\!\rangle = \sigma_4'$.
   Note: This refers to the exclusive sense of PLANET, that is, the sense that excludes Pluto.

Several observations are in order:

1. Implicit in the instances of **C** and **I** listed above is a *theory* of names and descriptions and of verbs. Our purpose here is not to defend this implicit theory; we simply take it as a postulate that this is how these categories of words work. Clearly, we believe it to be plausible, but we would be willing to replace the infons we have identified as the possible contents with others should someone persuade us otherwise. The important thing for equilibrium semantics is that the two maps give us appropriate infons as the possible contents of the words in an uttered sentence.

2. We deliberately do not discuss the similarities and differences between THE and A here. It would take us too far afield. The conventional meaning of A can be read off from its predicative use and is $\hat{\mathsf{P}}(\ \dot{\mathsf{P}}\ )$. We have also deliberately omitted a fifth quantificational use for A here.

3. We call the value of the conventional map the conventional meaning and the value of the informational map the referential meaning. When the latter is clear from the context, we will refer to it as just the meaning (or content). What the two maps together give us is the possible (referential) meanings of the words in an utterance.

4. Conventional meaning and referential meaning are intended as generalizations or refinements of the traditional distinctions between intension and extension, or connotation and denotation, or sense and reference.

5. Some contents make a reference to resource situations. When accessible, these situations enable the identification of the object that is being referred to.

6. We have already simplified these maps by assuming that there is just one John and that only PLANET is (conventionally) ambiguous. Likewise, to avoid a combinatorial explosion in our exposition, we will assume that A involves just the referential and generic uses. This reduces the number of possibilities we have to consider to a manageable size.

7. It should be easy to see that lexical ambiguity can involve either the conventional map or the informational map, an important fact that is often obscured.

8. The foregoing should give one some confidence that all words in $\mathcal{L}$ can be dealt with similarly with respect to some utterance situation u.

4.4 The Flow Constraint

Consider just the ambiguous word PLANET = ω_4 in the sentence φ uttered by $\mathcal{A}$ in u. We assume $\mathcal{A}$ is referring to the first inclusive sense of PLANET in this situation. Consider the game in Figure 1.

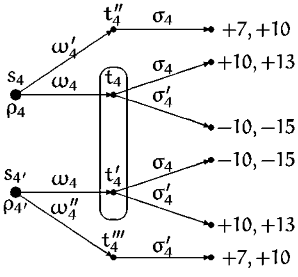


Figure 1: Game of Partial Information $g_4 = g_u(\omega_4)$

Here ω'_4 stands for an alternative, unambiguous, and therefore typically more complex expression that the speaker might have uttered but chose not to. This could be, for example, "planet, the sense that includes Pluto". Similarly, ω''_4 could be just "planet, the sense excluding Pluto". The symbol s_4 stands for the first situation where $\mathcal{A}$'s intention is to convey σ_4, and $s_{4'}$ is the counterfactual situation where $\mathcal{A}$ intends to convey σ'_4. The rest of the context in our example is shared by s_4 and $s_{4'}$, that is, it is contained in $\bar{s}_4 = s_4 \cap s_{4'}$. ρ_4 and $\rho_{4'}$ are the probabilities that $\mathcal{A}$ is referring to σ_4 or σ'_4 . Assume they are identical for $\mathcal{A}$ and $\mathcal{B}$ and are 0.9 and 0.1 respectively. The payoffs result from the costs and benefits of the various utterance parts and interpretations. The key thing to observe here is that a certain ordering amongst them must prevail, where successful interpretations are valued more highly than unsuccessful ones, and more complex expressions are penalized more highly than less complex ones. Benefits and costs result from a variety of situational and linguistic factors.

If we solve this game by most of the standard solution concepts (see Parikh 2001 for a discussion), we would find that σ_4 is the solution, which means that PLANET gets disambiguated in the intended way.

Note that a similar game can be formed for each word in the sentence φ and is denoted by $g(\omega_i, u) = g_u(\omega_i) = g_i$. Let the class of such games over all words and phrases and sentences be designated $\mathcal{G}$. For example, the game $g_u(\omega_3)$ is shown in Figure 2.[3] It is reasonable to assume that the payoffs for all the words in a sentence follow the same "pattern". When the number of initial nodes is the same, the pattern will be identical.

We can also define a product $\otimes$ on $\mathcal{G}$ for which we simply show an example here. The game in Figure 3 is the product of the games in Figures 1 and 2.[4]

[3] Recall that we are considering only two possible contents for A.

[4] Notice that the initial nodes of the product game are in the center of the diagram rather than to the left—where the situations and probabilities are placed beside slightly darker nodes—and the utterances go off to the right and left. This way of drawing it just reduces the tangle of nodes and branches.

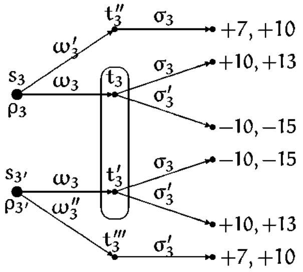


Figure 2: Game of Partial Information $g_3 = g_u(\omega_3)$

First, there are $2 \times 2 = 4$ initial nodes, labeled with the unions of the corresponding situations. The tree takes its shape from the "product" of the input trees in a natural way. The utterances in the product involve the concatenations of the corresponding utterances in the multiplier and multiplicand, the contents involve the sum of the corresponding addends, and the payoffs involve the corresponding arithmetical sums.

Note that several branches of the tree that occur from conjoining, for example, ω_3 with ω'_4 at the node $s_3 \cup s_4$, and likewise at other initial nodes, along with their corresponding contents and payoffs, have *not* been shown. This would needlessly clutter the tree diagram. Essentially, all possible combinations need to be accounted for for the tree itself, the utterances, the contents, and the payoffs and probabilities.

The probability of an initial node in a game of partial information is really the probability of the speaker's intention in the corresponding situation to refer to a particular content. For example, in Figure 1, ρ_4 would be the probability that $\mathcal{A}$ intends to refer to σ_4 in s_4 and $\rho_{4'}$ would be the probability that $\mathcal{A}$ intends to refer to σ'_4 in $s_{4'}$.

Let x_i stand for the possible contents σ_i^y of ω_i, where y stands for zero or more primes. Then the probability of an initial node for a lexical game is representable as $\rho_i(x_i \mid x_{-i}, \bar{s}_i)$, that is, the meanings of the words in a sentence are interdependent and influence each other.[5]

For example, $\rho_1 = \rho_1(\sigma_1 \mid x_2, x_3, x_4, \bar{s}_1) = 1$ since there is just one John; $\rho_2 = \rho_2(\sigma_2 \mid x_1, x_3, x_4, \bar{s}_2) = 1$ since there is just one infon corresponding to the verb; $\rho_3 = \rho_3(\sigma_3 \mid x_1, x_2, x_4, \bar{s}_3)$ and $\rho_{3'} = \rho_{3'}(\sigma'_3 \mid x_1, x_2, x_4, \bar{s}_3) = 1 - \rho_3$, since this game has just two initial nodes; and $\rho_4 = \rho_4(\sigma_4 \mid x_1, x_2, x_3, \bar{s}_4)$ and $\rho_{4'} = \rho_{4'}(\sigma'_4 \mid x_1, x_2, x_3, \bar{s}_4) = 1 - \rho_4$.

Now, the probabilities in the product would be $\rho_{34} = \rho_{34}(\sigma_3, \sigma_4 \mid x_1, x_2, s)$, $\rho_{34'} = \rho_{34'}(\sigma_3, \sigma'_4 \mid x_1, x_2, s)$, $\rho_{3'4} = \rho_{3'4}(\sigma'_3, \sigma_4 \mid x_1, x_2, s)$, and $\rho_{3'4'} = \rho_{3'4'}(\sigma'_3, \sigma'_4 \mid x_1, x_2, s)$. Here $s = [s_3 \cup s_4] \cap [s_3 \cup s'_4] \cap [s'_3 \cup s_4] \cap [s'_3 \cup s'_4]$. All of these sum to 1 of course.

These probabilities can be more compactly expressed by leveraging (and abusing) the notation: write $\rho_i(x \mid \bar{s}) = \rho_i(x_i \mid x_{-i}, \bar{s})$, where x is the vector (x_i) and $\bar{s}$ is simply the intersection of the situations at all the initial nodes. It is always possible to use the subscript of the ρ's to figure out which variables are the relevant random variables and which ones are the conditioning

[5] $\bar{s}_i = s_i \cap s_{i'} \cap s_{i''} \cap \ldots$ as before.

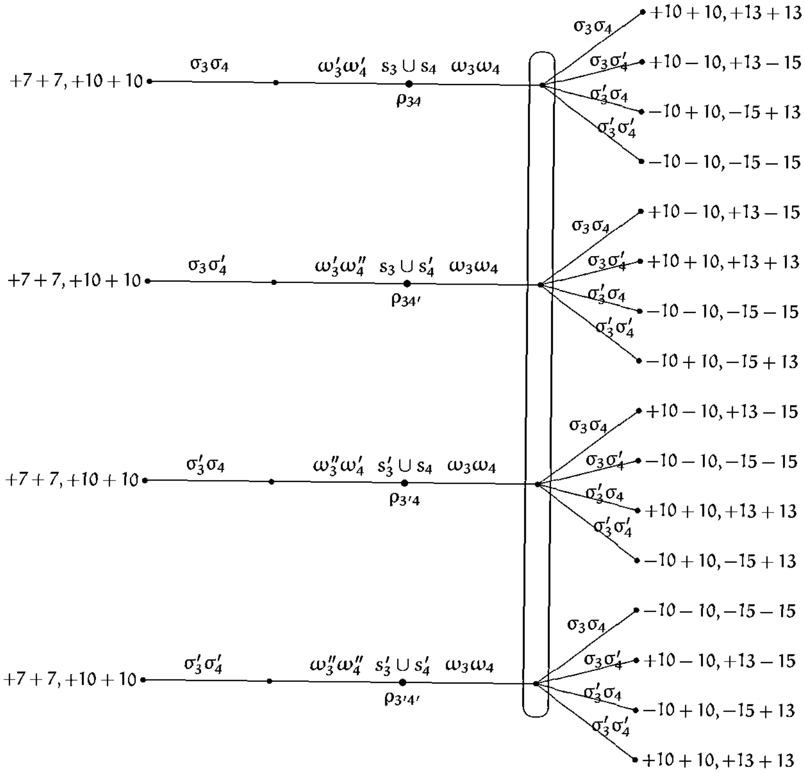


Figure 3: The product $g_{34} = g_3 \otimes g_4$

random variables.[6] For example, $\rho_{34}(x \mid \bar{s})$ would be identical to $\rho_{34}(x_3, x_4 \mid x_{-34}, \bigcap[s_{3^y} \cup s_{4^{y'}}])$, where x_{-34} is just the contents left out, namely, x_1 and x_2 and the y's represent zero or one primes. The key is the subscript of ρ which allows anyone to unambiguously determine the position of the arguments. This notation enables a compact representation of the fundamental equation in the next section.

This completes the description of the product. The solution to the product game is just $\sigma_3 \oplus \sigma_4$ as would be expected.

This shows how games of partial information can be set up to model the context of utterance u in order to derive the actual content of each word, phrase, and the entire sentence from the various possible contents as the solution to the relevant game. We label this fourth *flow constraint* **F**.

5 Equilibrium Semantics: SCIF

Equilibrium semantics is a generalization of model theory and draws upon four central ideas: reference, use, indeterminacy, and equilibrium.[7] It involves combining the four constraints we have introduced into a single framework.

The framework consists essentially of two *partial* homomorphic maps $(\mathcal{L}, \circ) \xrightarrow{g_u} (\mathcal{G}, \otimes) \xrightarrow{i} (\mathcal{I}, \oplus)$ connecting three monoids that take us from words and phrases via their embedding situations and corresponding games to their contents. We can then form the content $\mathcal{C}_u = i \circ g_u$ where $\circ$ now stands for function composition. This content function plays the same role in equilibrium semantics as the interpretation function plays in conventional model theory.

These maps essentially involve maps from the parse trees of the sentence to corresponding "trees of games" and further to trees of contents (or infons). That is, g_u maps a word or expression into a corresponding game, relative to the context of utterance u, as explained in Section 4.4. These games are embedded in an isomorphic tree of games. And i maps each game into its solution, which is its corresponding content.[8]

We have already seen the parse tree. The isomorphic tree of games is as follows:

$$[_S\, [_{NP}\, g_1\,] \otimes [_{VP}\, [_V\, g_2\,] \otimes [_{NP}\, [_{Det}\, g_3\,] \otimes [_N\, g_4\,]]]]\,. \tag{3}$$

The third corresponding tree is the contents tree which is isomorphic to the first two.

$$[_S\, [_{NP}\, \sigma_1\,] \oplus [_{VP}\, [_V\, \sigma_2\,] \oplus [_{NP}\, [_{Det}\, \sigma_3\,] \oplus [_N\, \sigma_4\,]]]]\,. \tag{4}$$

These three trees give us the full solution to the problem of deriving the literal meaning of φ from first principles. The second tree is obtained from the first via g_u and the third tree is obtained from the second via i.

If we make explicit the dependence of the map $g_u(\omega_i)$ on the probabilities that come from u, we can write this as $g_u(\omega_i, \rho_i(x \mid \bar{s}))$, where x is the vector of infons. This then allows us to write $\mathcal{C}_u(\omega_i) = i[g_u(\omega_i, \rho_i(x \mid \bar{s}))]$ which of course is equal to x_i by definition. So if we express

[6] Of course, ordinarily $\rho_i(x \mid \bar{s})$ would equal $\rho_i(x_1, \ldots, x_k \mid \bigcap s_i^y)$ where the subscript k is the dimensionality of the vector. But in the present context, the notation is used in a non-standard way.

[7] In this paper, we will have nothing to say about indeterminacy, but see Parikh (2006). It should be obvious how the other three ideas are being used.

[8] The map is not to the whole solution or strategy profile, just to that part corresponding to the addressee's interpretation.

this in vector notation by including all the $\mathfrak{i}$'s, we get the compact and elegant vector fixed point equation:

$$\mathcal{C}_u(\omega, \rho(x \mid \bar{s})) = \mathfrak{i}[g_u(\omega, \rho(x \mid \bar{s}))] = x\,. \tag{5}$$

This is the equation we have to solve to compute the meaning of an utterance. The reader can verify that the solution to this equation is precisely the sum of the infons in Equation 4.

Thus, equilibrium semantics combines **SCIF** via two homomorphic maps g_u and $\mathfrak{i}$ into a new theory and framework for meaning.

6 Conclusion

We have compressed in these pages the framework of equilibrium semantics which allows us to compute the meaning or content of any utterance in principle. We have shown how to handle simple utterances involving simple noun and verb phrases, but to the best of our knowledge, it is the only framework of its kind that can actually derive the meanings of even such simple expressions from first principles, assuming little more than a limited rationality. It is not difficult to extend this to more complex utterances and we do so in Parikh & Clark (2007).

In any case, it should be reasonably clear how we build *use* into *reference* at the ground level via *equilibrium* by inserting the tree of games between the parse tree on the one side and the contents tree on the other. This allows us to combine semantics and pragmatics into a single discipline by enabling both the computation and the representation of the content of any utterance.

References

Barwise, J. (1989). *The Situation in Logic*. CSLI Publications, Stanford.

Barwise, J. and J. Perry 1983. *Situations and Attitudes*. MIT Press, Cambridge, Mass.

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

Parikh, P. (2001). *The Use of Language*. CSLI Publications, Stanford.

Parikh, P. (2006). Radical semantics: A new theory of meaning. *Journal of Philosophical Logic*, **35**, 349-391.

Parikh, P. and R. Clark (2007). *Language and Interaction*. In progress.

Parikh, P. and R. Clark (2006). The meaning of THE: A new account of definite descriptions. In progress.

Recanati, F. (2004). *Literal Meaning*. Cambridge University Press, Cambridge, Mass.

Wilson, D. and D. Sperber (1986). On defining relevance. In: *Philosophical Grounds of Rationality* (R. Grandy and R. Warner, eds.), pp. 243-258. Clarendon Press, Oxford.

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 11

RULE ORDERING: A LOOK AT QUANTIFIER SCOPE AND COORDINATION IN GTS

Tatjana Scheffler
University of Pennsylvania

This article investigates a topic in game-theoretical semantics (GTS) that has received relatively little attention in the previous literature: ordering principles that guide the application of game rules. Sentences with quantifier scope ambiguities show that ordering principles cannot impose a fixed hierarchy on game rules. It is proposed that the principles allow game rules to be played in different orders, which yields two or more different games for some input sentences. These distinct games correspond to distinct semantic interpretations. Based on data involving complex quantifier scope ambiguities, including inverse linking examples, a new ordering principle for quantifiers is proposed. It is argued that a hierarchy is needed that determines the relative precedence of ordering principles, and a partial hierarchy is presented towards that end. The approach is then tested with respect to coordination and quantifier scope.

1 INTRODUCTION

In the earlier literature, game-theoretical semantics (GTS) has been applied to the explanation of anaphora and of complex quantifiers, and it seems especially well suited for these tasks. Nevertheless, quantifier scope is an area that traditional approaches to formal semantics have been concerned with in a large scale. However, this topic has been cut short in the literature on GTS.

Yet GTS is a formalism that claims to be largely syntax-independent. Game rules are deliberately formulated to operate on input strings. The only structure that has been brought into the system is in terms of a couple of ordering principles that determine the sequences in which game rules are applied.

This article investigates data from quantifier scope ambiguity in order to illustrate the important role ordering principles play in the formalism. After a brief introduction to GTS (Section 2), Section 3 presents an attempt to derive multiple readings for ambiguous sentences. This amounts to considerable complexity in the ordering rules. Section 4 shows how a new ordering rule for quantifiers can handle inverse linking that initially poses some problems for this framework. The discussion elaborates on what the structure is that the ordering principles derive, which leads to the introduction of a hierarchy for ordering principles that determines their relative precedence (Section 5).

The remainder of the chapter presents some problems related to scope and coordination. In Section 6, a sketch of how such examples can be handled using the GTS-type ordering principles is provided. Section 7 concludes the chapter.

2 Game-Theoretical Semantics

As a formalism for semantics, game theory originated with Lorenzen (1955) and was significantly developed in Hintikka (1973) and subsequent works. In GTS, the truth (or falsity) of each sentence is determined by a non-cooperative game between two agents, sometimes called Eloïse and Abelard. The moves of the two players are determined by game rules which replace some part of the sentence; when the sentence is reduced to an atomic formula, the game stops. Eloïse is the initial verifier, Abelard the initial falsifier. This means that a sentence is true iff Eloïse has a *winning strategy* for the game (i.e., she can win no matter how Abelard moves), and it is false iff Abelard has a winning strategy.

2.1 Game Rules

In the previous literature, game rules have been proposed for a wide range of items in natural language, including quantifiers (Hintikka and Kulas 1985, Clark 2004), anaphora (Janasik et al., 2002), possessives, intensional verbs (Hintikka and Kulas, 1985), adverbs and eventualities (Pietarinen, 2001), among others. Throughout this paper, suitable game rules for the quantifiers *a* (= *some*) and *every* will be assumed. Two natural rules are the following:

(G.a): If the game G(S; M) has reached an expression of the form:

Z – a X who Y – W

then the current verifier chooses an individual c from the appropriate domain. The game continues as G(Z – c – W, c is an X and cY; M).

(G.every): If the game G(S; M) has reached an expression of the form:

Z – every X who Y – W

then the current falsifier chooses an individual c from the appropriate domain. The game continues as G(Z – c – W, if c is an X and cY; M).

These game rules formalise the insight that for existential quantification, the verifier has to find an example that makes the sentence true, whereas for universal quantification, the falsifier attempts to find a counterexample.

2.2 Rule Ordering

GTS, as it is presented for example in Hintikka (1996), is not independent of syntactic parsing but rather presupposes it. Although game-theoretic rules are formulated to operate on strings, syntactic structure must be given. For formal languages, this is trivial, since the syntax is always explicit in the formula: bracketings show which element is available for play and thus which rule can be applied. For example, the rule **(G.$\forall$)** is played in (1), whereas only **(G.$\exists$)** is available in (2).

(1) $\forall x[\text{linguist}(x) \rightarrow \exists y[\text{party}(y) \wedge \text{attends}(x, y)]]$

(2) $\exists y[\text{party}(y) \wedge \forall x[\text{linguist}(x) \rightarrow \text{attends}(x, y)]]$

As noted in Hintikka (1982), the scope of natural-language quantifiers is, in contrast, not as easily obtained. In GTS, rule ordering is used to derive many of the same effects that explicit scope marking has for the quantifiers in formal languages. The order in which game rules are applied to natural-language sentences is thus subject to certain principles, which can be either general or item-specific. **(O.comm)**, **(O.LR)** and **(O.LR.subgames)** are general ordering principles, whereas **(O.any)** is a specific principle; all are mentioned in the previous literature (Hintikka and Sandu, 1991, pp. 27f):

(O.comm): A game rule must not be applied to an ingredient of a lower clause if a game rule applies to an ingredient of a higher one.

(O.LR): In one and the same clause, game rules are applied from left to right.

(O.LR.subgames): Subgames are played in left-to-right order.

(O.any): (G.any) has priority over **(G.not)**, **(G.or)**, and **(G.cond)**.

These principles take over the role that syntactic structure (bracketing) had in the semantics of formal languages. In this capacity, they are also the counterpart of syntactic structure in traditional approaches to natural-language semantics: Issues that have been analysed by appeal to *Logical Form*, a syntactic structure of sentences on which semantic interpretation is based, will have an impact on the ordering rules of GTS. Some of these issues, among which quantifier raising and reconstruction effects are the most notable, are discussed in this chapter. The underlying question is whether rule ordering can be used to obtain the available interpretations for natural-language sentences. To do that, what kinds of ordering principles are needed?

3 QUANTIFIER SCOPE AMBIGUITY

So far, the ordering principles uniquely determine the proceedings of a game, just like the syntactic structure (bracketing) does for formal languages. At each point, only one game rule is available for play. Following these principles derives exactly one interpretation for each sentence no matter how many quantifiers it contains. Because of **(O.LR)**, this interpretation will be the one corresponding to the surface scope of all quantifiers.[1]

Consider an English sentence (3) with two quantifiers. The two possible interpretations are in (4).

(3) Every linguist attends a party.

(4) a. $\forall x[\text{linguist}(x) \rightarrow \exists y[\text{party}(y) \wedge \text{attends}(x, y)]]$

[1] An exception is *any*, since the specific ordering principle **(O.any)** determines that *any* always takes wide scope.

b. $\exists y[\text{party}(y) \wedge \forall x[\text{linguist}(x) \rightarrow \text{attends}(x, y)]]$

With the existing rules, only the surface-scope reading, according to which *every* scopes over the indefinite *a* as in (4a), can be obtained. In a discourse, the inverse reading (4b) might, however, be the appropriate one. Therefore, the game must be allowed to optionally proceed in a different way, so that *a* can be processed before *every*.

Thus, there must be some variability concerning the application of the ordering principles. For a first approximation, one can assume that within a clause, quantifiers can be chosen freely (hence violating **(O.LR)**):[2]

(O.quant): Within a clause, any quantifier can be chosen at any point before it actually appears in the string. This takes precedence over **(O.LR)**.

Application of this non-deterministic principle allows for two alternative proceedings of the game. Some theoretical implications of this are discussed in Section 5. In practice, the new principle allows us to derive the two interpretations for the above example sentence. The games are shown in (5).[3]

(5) a. *Every linguist attends a party.*

John attends a party, if John is a linguist.	**(G.every)**
John attends $\mathcal{A}$, if John is a linguist and $\mathcal{A}$ is a party.	**(G.a)**
$\Rightarrow \forall > \exists$	

b. *Every linguist attends a party.*

Every linguist attends $\mathcal{A}$, and $\mathcal{A}$ is a party.	**(G.a)**
John attends $\mathcal{A}$ and $\mathcal{A}$ is a party, if John is a linguist.	**(G.every)**
$\Rightarrow \exists > \forall$	

4 INVERSE LINKING

An interesting problem concerning quantifier scope refers to inverse linking, as in (6):

(6) Some sailor on every ship in some harbour is drunk.[4]

The sentence contains a series of nested quantifiers. In one reading, these quantifiers are interpreted in their surface order, so that each quantifier scopes over the ones that are embedded into it. A different reading where a (lower) quantifier scopes over its embedding (higher) quantifier is called inverse linking.

Hintikka and Sandu (1991, p. 74) claim that the inverse-linking reading is available here because the surface reading is nonsensical and thus, in the course of the game, yields uninterpretable sentences. The ungrammaticality of the surface reading is, according to Hintikka and Sandu, what allows **(O.LR)** to be violated.

[2] Clark (2004) suggests that the order in which quantifiers are chosen should not be completely arbitrary. Instead, he claims that probabilities should be attached to each alternative. I will not discuss this proposal in this paper.

[3] The current falsifier is called Falsifier, the current verifier Verifier. They correspond to what Hintikka calls Nature and Myself, respectively. Game rules are noted on the right, as they are applied.

[4] This is Example (5.1.1) of Hintikka and Sandu (1991).

4.1 Inverse Linking is not the Exception

There are, however, cases in which both the inverse linking reading and the surface reading are plausible (7). Two possible contexts that disambiguate between the readings (and show their availability) are given in (7a) (surface reading) and (7b) (inverse linking).

(7) Every parent of two children should attend the meeting.

a. There's a meeting tomorrow about helping children develop good relationships with their siblings. Parents with only one child do not have to come.
But *every parent of two* (or more) *children should attend the meeting.* ($\forall > 2$)

b. Our meeting tomorrow will talk about children who get in fights on the school yard. In our class, most children do not get into trouble, so their parents do not have to show up at the meeting.
But *every parent of* (these) *two children should attend the meeting.* ($2 > \forall$)

The rule **(O.quant)** allows exactly these two interpretations. Since any quantifier is available for play independently of its position in the string (contra **(O.LR)**), *two* can be played before *every*, and the inverse linking can be derived in this way.

The game which derives the inverse linking reading (7b) proceeds by playing on *two*, yielding:[5]

(8) Every parent of Mary should attend the meeting, and Mary is a child.

This is then resolved to:

(9) Bob should attend the meeting, if Bob is a parent of Mary; and Mary is a child.

Note that *of Mary* is a restrictive modification of the noun phrase, and thus part of the restriction of the quantifier *every*, just like the common noun *parent* and just like restrictive relative clauses specifically mentioned in every quantifier game rule.

The restrictive prepositional phrase must be treated in the same way as a restrictive relative clause. This also reflects the semantics: the *two children* are part of the restriction of the embedding quantifier, not part of its nuclear scope. (10) is an example very close to Hintikka and Sandu's original inverse linking sentence (6), and its most prominent reading is indeed the inverse linking one.

(10) Some parent of every student should attend the meeting.

The only way to play a game successfully on this sentence is when the restriction of *some* is treated as such. Otherwise, even in the inverse linking case one would derive the following ungrammaticality:

(11) Bob of Mary should attend the meeting, if Bob is a parent; and Mary is a student.

Using this account of restrictive prepositional phrases, the game deriving the surface scope reading (7a) proceeds as follows. First, **(G.every)** is played, yielding:

[5] I assume a rule for *two* similar to one given for *at least n* in Clark (2004). Such a rule would require the verifier to choose two individuals, of which the falsifier then picks one.

(12) Bob should attend the meeting, if Bob is a parent of two children.

Further play on *two* produces:

(13) Bob should attend the meeting, if Bob is a parent of Mary, and Mary is a child.

It should not worry us that the final sentences for the surface reading and the inverse reading are the same, since it is the entire game that matters. In the surface case, each parent of any two children is obliged to attend, since the parent 'Bob' is chosen without reference to particular children. In the inverse linking case, the parent is chosen *after* the children have already been picked, and is therefore informationally dependent on the children. This is the essence of the two distinct readings for the sentence.

One note about Hintikka and Sandu's original example (6) is in order: with the approach proposed here, nothing prevents us from deriving the surface scope reading for their sentence as well. The first step in the game, playing on *some sailor*, yields the following (grammatical) sentence:

(14) Jack Tar is drunk, and Jack Tar is a sailor on every ship in some harbour.

The question is to what extent this sentence makes sense. If it is actually the case that Jack Tar can be said to be a sailor on every ship in a harbour, then this reading is fine. More likely, however, there is no such sailor in the world, in which case this reading is not ungrammatical but simply false.

4.2 A More Complex Example

This section shows that the new quantifier ordering principle **(O.quant)** is still too permissive. For several quantifiers in a clause, the rule predicts all permutations of quantifiers to yield a possible reading. This makes the correct predictions for a suitable example, such as (15).

(15) a. (At least) two social workers gave a doll to each/every child.

b. $\forall > 2 > \exists$:
$\forall y[\mathrm{child}(y) \rightarrow \exists x[\mathrm{socialworkers}(x) \wedge |x| \geq 2 \wedge$
$\forall x'[x' \subset_i x \rightarrow \exists z[\mathrm{doll}(z) \wedge \mathrm{gives}(x', z, y)]]]]$ (Joshi et al., 2003, (5))

The sentence (15a), with the three quantifiers Q_1, Q_2 and Q_3, has the prominent reading spelled out in (15b). The same ordering of quantifiers, $Q_3 > Q_1 > Q_2$ is ruled out in examples with nested quantifiers, such as (16).

(16) Two politicians spy on someone from every city. (Larson, 1985, (12))

The sentence (16) also has three quantifiers, and it is therefore predicted, according to the game rule **(O.quant)**, to have $3! = 6$ different readings. One of the readings is generally excluded because it has very weak truth conditions, namely one in which $\exists > 2 > \forall$.[6]

[6] This is presented in Joshi et al. (2003). But note that the following sentence has exactly the same structure as the one above while the reading in question does not have equally weak truth conditions.

(i) Two politicians talked to every representative of some country.

The inverse linking reading $\forall > 2 > \exists$ (corresponding to the reading exhibited in (15b)) is also excluded, but this is not a logically impossible reading (Hobbs and Shieber, 1987). It turns out that in cases with nested quantifiers two nested quantifiers must be interpreted next to each other, not allowing other quantifiers to intervene. GTS does nothing to prevent this unavailable reading. The derivation in GTS, using the machinery introduced so far, would proceed as follows:

(17) Two politicians spy on someone from Philadelphia, if Philadelphia is a city. **(G.every)**

Bob spies on someone from Philadelphia, if Philadelphia is a city and Bob is a politician. **(G.two)**

Bob spies on Bill, if Philadelphia is a city and Bob is a politician, and Bill is from Philadelphia. **(G.some)**

$\Rightarrow * \forall > 2 > \exists$

The intuition is that quantifier phrases embedded into each other build a unit that nothing can intervene in. I formalise this by incorporating it into the ordering rule for quantifiers **(O.quant)**.

(O.quant)$'$: Within a clause, game rules for quantifiers can be applied earlier than **(O.LR)** would allow them. Once a quantifier has been chosen for play, rules for its embedded or embedding quantifiers take precedence over everything else.

Another possibility to account for the facts would be to keep the ordering rule **(O.quant)** from above, and state the additional constraint in the individual rules for quantifiers, for instance by making reference to possible embedded quantifiers in the structural description. I will not pursue this option further since it would mean a multiplication of the quantifier rules (at least one additional rule per quantifier).

Finally, the modified ordering principle **(O.quant)$'$** is still problematic: the sentence (18) shows why.

(18) Two sailors on some ship in every harbour are drunk.

This example sports three quantifier phrases that are nested into each other. The judgments are the same as above for (16). That is, all scope orderings are possible, as long as the outermost quantifier (*two sailors*) does not intervene between the other two. Given this data, a potential ordering principle for quantifiers becomes very complex:

(O.quant)$''$: Within a clause containing the nested quantifier phrase (QP_1 (QP_2 (QP_3))), where QP_1 and QP_3 may be arbitrary nestings of zero or more quantifiers, QP_2 may be played on at any time. Then, other ordering rules are suspended, until QP_3, followed by QP_1, has been processed.

This principle is interesting since it differs from **(O.LR)** and **(O.comm)** in being more detailed. The new rule also depends on more structure being explicitly marked. The simpler examples with non-embedded noun phrases (e.g., (3)) reduce to the trivial case in which QP_1 and QP_3 in **(O.quant)** are both zero. The other ordering principles then regulate the rest of the derivation.

5 The Nature of Ordering Principles

At first, ordering principles for quantifiers seem to differ from previous suggestions. In this section I will show that new principles are not so different from the others, and discuss an alternative way of deriving scope ambiguity in GTS.

5.1 Ordering Principles are Hierarchical

(O.quant)$''$ looks different from previous ordering principles because it makes an explicit reference to other ordering rules. It speaks *about* the principles that determine the application of game rules. However, this is not so new. **(O.any)** already implicitly overruled **(O.LR)**—this hierarchy of ordering principles has simply been made explicit. For **(O.LR)** and **(O.comm)**, their respective importance has been put into the phrasing of the principles, specifying that **(O.comm)** takes precedence, since **(O.LR)** applies "in one and the same clause". The hierarchy of the ordering principles discussed so far, made entirely explicit, should therefore be the following:

(19) *Hierarchy of Ordering Principles:*
(O.subgames) $\gg$ **(O.comm)** $\gg$ **(O.any)** $\gg$ **(O.quant)**$''$ $\gg$ **(O.LR)**.

The rule orderings, seen in this way, provide the *structure* (or *syntax*) on which the game itself derives the semantics. If this is the task of the ordering principles, they may become very complex. However, such machinery seems necessary if the semantics should be able to derive any and all of the readings for a given sentence. Each structure, or each sequence of game rules, if it is grammatical, yields a distinct reading of the sentence. For each such game, a separate winning strategy can exist.

5.2 Ordering Principles Should not be Violated

Hintikka and Sandu (1991, p. 76) discuss an inverse linking example which allows two distinct interpretations (their 5.1.10):

(20) Every book on some interesting topic by any author is interesting.

Hintikka & Sandu derive two games and two readings by allowing **(O.LR)** to be violated in some cases, and letting *any author* be played on before *some interesting topic*. This approach, if permitted in every sentence, could of course allow the derivation of two readings in the simple quantifier scope ambiguity cases such as (3).

In general, though, ordering principles should not be violable. For example, allowing the rule **(O.comm)** to be violated would lead to an abundance of impossible interpretations for common sentences. For example, what would then prevent us from taking (21) to mean that for every professor, there is a student that believes the professor own a Porsche? But the sentence does not have this reading.

(21) Some student believes that every professor owns a Porsche.

For this reason, I will take the ordering rules to be inviolable (unless explicitly overruled). The hierarchy makes explicit which ordering principles take precedence over others. This provides a better tool than what we have had in the literature for controlling the order in which games are played on natural-language sentences, and prevents the possibility of deriving unavailable readings for sentences such as (21).

6 COORDINATION

In this section, I will discuss the question of how scope and coordination interact. These interactions lead to the question of whether finding a single, hierarchical set of ordering principles is possible.

6.1 SIMPLE COORDINATION IN GTS

Logical conjunction only recognises coordination of entire terms. In natural language, coordination of other types of constituents is common. Examples of constituent coordination are:

(22) *John and Mary* went to the movies.

(23) *John and Mary* met.

(24) *Every man and every woman* met.

The basic idea for constituent coordination is that it is distributive, in other words whatever is said about the coordinated constituents (for example, whatever is predicated of two coordinated NPs) should hold for each of them separately. Thus, (22) means that John went to the movies and Mary went to the movies. I adapt the game rule **(G. $\wedge$)** from formal languages to English:

(G.and): If the game $\mathsf{G}(\mathsf{S};\mathsf{M})$ has reached an expression of the form:

$$\mathsf{Z}-\mathsf{X} \text{ and } \mathsf{Y}-\mathsf{W}$$

then the current falsifier chooses a $\mathbf{c} \in \{\mathsf{X},\mathsf{Y}\}$, and the game continues as $\mathsf{G}(\mathsf{Z}-\mathsf{c}-\mathsf{W};\mathsf{M})$.

This does not work for all coordinations. One problem are predicates that are not distributive but collective, like the ones in (23–24). Their interpretation will pattern with the interpretation of plurals (like *they met*), and I will not address this interesting issue here.

6.2 COREFERENCE AND CONJUNCT ORDER

Consider:

(25) *Some father and his son* laughed.

Example (25) is interesting because it uncovers a deep problem with a solution: if only one of the conjuncts is ever inspected (as is done with sentential coordination for formal languages), some coreference effects cannot be obtained. If the game is played on a model in which there is a father and his (only) son, the falsifier has a winning strategy by choosing the conjunct “his son laughed”. A common approach to pronouns allows the verifier to choose freely from a list (choice set) of potential referents, which is compiled from previously picked individuals (Janasik et al. 2002, Clark 2006). Now verifier cannot find a referent for the pronoun, and therefore loses.

To avoid this problem, the game rule must be changed so that both conjuncts will actually be checked, in a certain order. To do this, the game will be split into two subgames. Furthermore, the ordering principle **(O.subgames)** (see Hintikka and Sandu 1991) guarantees that the subgames will be played in the order they are listed.[7]

[7] [See Gabriel Sandu’s paper in this volume for related discussion.]

(G.and)′: If the game $\mathsf{G}(\mathsf{S};\mathsf{M})$ has reached an expression of the form:

$\mathsf{Z}-\mathsf{X}$ and $\mathsf{Y}-W$

then the game continues as $\mathsf{G}(\mathsf{Z}-\mathsf{X}-W, \mathsf{Z}-\mathsf{Y}-W;\mathsf{M})$.

Together with **(O.subgames)**, **(G.and)′** predicts that the following sentence with the alternative conjunct order runs into problems. This is, in fact, the case because the sentence (26) is ungrammatical.

(26) **His son and some father* laughed.

6.3 Scope Restrictions

Consider next:

(27) *Every man and every woman* solved a puzzle.

Sentence (27) introduces a scope ambiguity similar to ones I discussed in Section 3. The sentence has two readings, one in which *a puzzle* has wide scope and one in which *a puzzle* is outscoped by both universal quantifiers (presumably independently, as I will assume the distributive reading according to which every person worked on one puzzle on their own).

Nothing in the ordering principle **(O.quant)″** guarantees just these two readings. Instead, one obtains all four combinatorially possible scope orderings. A further fact about the data is that when *a puzzle* has wide scope, it must necessarily be the same puzzle for both men and women (that is, it must scope over *and* as well).

One can obtain the wide scope of *a puzzle* using the ordering rule for quantifiers from the previous section. If **(G.a)** is played first on the sentence, the coordination can be split up later and we get the desired result. The ordering of **(G.a)** at the beginning of the game is optional, though. The coordination has to be processed first in order to get the surface scope, $\forall > \exists$. But then, nothing prevents the players from applying **(G.a)** before **(G.every)** in one of the individual subgames. Because each subgame should be independent from the others, it is not guaranteed that their structure is parallel.

One approach is to require the conjunction *and* that coordinates NPs to be processed late. That is, the individual parts are played before the rule for *and* is applied for NP coordination. Then, both *every* quantifiers would have to be replaced by constants first, before **(G.and)** splits the game. The ordering principle **(O.and)** requires **(G.and)′** to apply last in the case of NP coordination.

(O.and): In an expression of the form:

$\mathsf{Z}-[_{\mathsf{NP}}\ \mathsf{X}]$ and $[_{\mathsf{NP}}\ \mathsf{Y}]-W$,

a game rule must not apply to *and* if it can apply to X or if it can apply to Y.

This ordering principle permits the following games on (27):[8]

[8] At the moment, it is not clear how some additional readings that result from interleaving the play on *a* somewhere in between the two *everys* might be excluded.

(27) *Every man and every woman solved a puzzle.*

a. Every man and every woman solved Rubik's Cube, and Rubik's Cube is a puzzle. **(G.a)**
John and every woman solved Rubik's Cube, Rubik's Cube is a puzzle, and John is a man. **(G.every)**
John and Mary solved Rubik's Cube, Rubik's Cube is a puzzle,... **(G.every)**
John solved RC, Mary solved RC, RC is a puzzle,... **(G.and)**′
$\Rightarrow \exists > \forall$

b. John and every woman solved a puzzle, and John is a man. **(G.every)**
John and Mary solved a puzzle, and John... **(G.every)**
John solved a puzzle and Mary solved a puzzle,... **(G.and)**′
John solved Rubik's Cube, Mary solved the Tower of Hanoi,... **(G.a)**
$\Rightarrow \forall > \exists$

These two games derive exactly the two possible readings for (27).

7 Conclusion

The ordering principles that govern the application of game rules during the course of a game are one part of GTS that has not yet been worked out in sufficient detail. It is easy to find counterexamples that abuse the game rules when they are simple string-matching rules. Therefore, some syntactic structure has to be assumed. This fact makes it not so much different from traditional semantic approaches, which require certain very particular syntactic trees to work on (e.g., LF). It is an open question to what extent the claim that GTS does not pose constraints on syntax and works independently from it can be maintained.

Quantifier scope ambiguities are a special problem for GTS because if there is no LF, and no movement, all quantifiers are interpreted in situ. That is, *temporal ordering* during the game determines the scope of quantifiers, because once the rule has been applied, everything left in the clause is in the nuclear scope of the quantifier.

So far, the only rule touching that issue was **(O.LR)**, which requires left-to-right processing. This is of course too strict, especially when taking into account that some sentences have more than one reading due to quantifier scope ambiguities.

Therefore, I proposed an additional ordering rule **(O.quant)**″ which allows quantifiers to be optionally interpreted 'earlier' than in their actual surface position. This worked well for the simple case, but immediately opened up new questions about scope islands and scope ordering in nested quantifiers.

Furthermore, NP coordinations raise problems of scope ambiguities and scope restrictions. A mechanism that can force quantifiers to be interpreted 'together' (or rather, 'next to each other'), as suggested above for nested quantifiers, may help in this case. Conjoined quantifiers also seem to take scope together, either both above or both below other scopal elements.

It was shown here that the variability of natural language coordination is a potential problem for GTS (just as for other semantic approaches). As with quantifier scope ambiguities, the data supporting ordering principles is, in part, contradictory. It seems that an absolute ordering of game rules may not be possible after all. A possible extension would be a statistical ranking of rules that is sensitive to the context. The path chosen here emulates syntactic structure through the ordering rules. This leads to considerable complication in the rules but has been shown to derive correct interpretations for a variety of sentences.

ACKNOWLEDGEMENTS

I would like to thank Ahti-Veikko Pietarinen, Uwe Scheffler and the anonymous reviewers for comments on drafts of this paper. Furthermore, I am grateful to Robin Clark for introducing me to GTS.

REFERENCES

Clark, R. (2004). Game rules for numeric and majority quantifiers. Manuscript, University of Pennsylvania.

Clark, R. (2006). Quantifier games and reference tracking. In: *Logic and Games: Foundational Perspectives* (O. Majer, A.-V. Pietarinen and T. Tulenheimo, eds.).

Hintikka, J. (1973). *Logic, Language Games and Information.* Oxford University Press, Oxford.

Hintikka, J. (1982). Game-theoretical semantics: insights and prospects. *Notre Dame Journal of Formal Logic*, **23**, 219-241.

Hintikka, J. (1996). *The Principles of Mathematics Revisited.* Cambridge University Press, Cambridge.

Hintikka, J. and J. Kulas (1985). *Anaphora and Definite Descriptions.* D. Reidel, Dordrecht.

Hintikka, J. and G. Sandu (1991). *On the Methodology of Linguistics.* Blackwell, Oxford.

Hobbs, J. R. and S. M. Shieber (1987). An algorithm for generating quantifier scopings. *Computational Linguistics*, **13**, 47-63.

Janasik, T., A.-V. Pietarinen and G. Sandu (2002). Anaphora and extensive games. In: *Chicago Linguistic Society* **38**: *The Main Session* (M. Andronis, E. Debenport, A. Pycha and K. Yoshimura, eds.), pp. 285-295. Chicago Linguistic Society, Chicago.

Joshi, A. K., L. Kallmeyer and M. Romero (2003). Flexible composition in LTAG, quantifier scope and inverse linking. In: *Proceedings of the 5th IWCS*, pp. 179-194. Tilburg.

Larson, R. (1985). Quantifying into NP. Manuscript, MIT.

Lorenzen, P. (1955). *Einführung in die operative Logik und Mathematik.* Springer, Berlin.

Pietarinen, A.-V. (2001). Most even budget yet: Some cases for game-theoretical semantics in natural language. *Theoretical Linguistics*, **27**, 20-54.

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 12

TWO NOTIONS OF SCOPE

Gabriel Sandu
University of Helsinki
CNRS Paris I

I offer a dynamic version of game-theoretical semantics (GTS) which accounts for the distinction between logical and binding scope. The present paper modifies an argument contained in Sandu & Janasik (2003).

1 Two Notions of Scope

Usually, a standard quantifier is associated with both a *logical* (priority) and a *binding* scope. The former indicates a relation of logical dependence and independence as in the formula

$$\forall x(\exists y A(x,y) \rightarrow B(y)), \tag{1}$$

where the universal quantifier is logically prior to the existential quantifier and implication, which, in turn, is logically prior to the existential quantifier. The binding scope, on the other hand, relates to the segment in which a free variable is said to be bound by the quantifier. In the traditional Frege-Russell logical languages, if a variable is in the binding scope of a quantifier, then it is also in its logical scope. Logicians and philosophers (Peter Geach, Jaakko Hintikka, David Kaplan and Hans Kamp, among others) noticed that this is no longer so for natural language as the following examples indicate:

A girl smiles. She is happy. (2)

If a girl smiles, she is happy. (3)

In (2), the indefinite *a girl* is a head of *she*. If we construe the head-anaphor relation on the analogy 'quantifier-bound variable', then (2) is an example of a variable being in the binding scope of a quantifier but not in its logical scope. In Hintikka & Sandu (1989), we followed Hintikka (1987) and introduced two kinds of brackets to distinguish between the two notions of scope. In the following logical representations of (2) and (3), the parentheses indicate the binding scopes and the square brackets the logical scopes:

$$\exists x([G(x) \wedge S(x)] \wedge H(x)) \tag{4}$$

$$\exists x([G(x) \land S(x)] \rightarrow H(x)). \tag{5}$$

The main question of the paper is: Can we have a semantical interpretation which distinguishes the two notions of scope and is such that the existential quantifier in (5) has the force of a universal quantifier?

2 Two Attempts in Game-Theoretical Semantics

2.1 The Subgame Interpretation

Hintikka and his associates (Carlson & ter Meulen 1979, Hintikka & Kulas 1983, 1985) attempted to give an answer to our question in game-theoretical semantics (GTS) using the notion of a subgame. The basic idea is to divide an overall semantic game into several subgames, each of which is played out completely before the players move on to consider the next one. The notion of a subgame was intended to capture 'the conditional character of conditionals', and to avoid the problems associated with the analysis of 'if X then Y' as the material conditional '¬X or Y'. It is clear that the correct analysis of the conditional must provide a warrant for the passage from the truth of the antecedent to the truth of the consequent.

Indeed, according to the original idea of GTS, given a game $G(X \rightarrow Y)$ on a conditional $X \rightarrow Y$, the game $G(Y)$ on the consequent should be played only if $G(X)$ has turned out to be true. Furthermore, $G(Y)$ should be played in a way that depends on the mechanisms which led to the verification of X. Exploiting the fact that in GTS the truth of a sentence A amounts to the existence of a winning strategy for the verifier in the game $G(A)$, the subgame interpretation takes the truth of the whole conditional to be defined in terms of a mechanism by which the strategy used in verifying the antecedent X is somehow 'remembered' in playing the game on the consequent Y. What this 'remembering' amounts to becomes particularly evident in the case of pronominal anaphora. Typically a game-theoretic strategy in the antecedent reduces to the choice of an individual which can serve as the head for a subsequent pronoun via a choice set which is the book-keeping device keeping track of the individuals introduced in the course of the game. More generally, the notion of a subgame allows a natural extension of GTS from the sentential level to the level of a discourse. A fragment of discourse can now be conceived of as a 'supergame' consisting of several subgames played on successive sub-sentences.

Despite its rather long history and diligent application, the notion of subgame has never been made very precise but instead has been used more or less heuristically. The main difficulty with the mechanism of subgames in the Hintikka-Carlson-Kulas framework is that it makes the relation of an anaphoric pronoun to its head dependent upon the notion of truth. In other words, if we follow this idea in the game associated with (2), the subgame associated with the consequent is played only if the antecedent was first shown to be true. If the antecedent is shown to be false, the second subgame is not played at all. In other words, in conditionals with false antecedents there is no need to establish any anaphoric link whatsoever.

2.2 IF Logic

The second attempt in the game-theoretical literature has been to take the scope distinction to be a phenomenon of *informational independence* in the sense of the theory of games. Hintikka & Sandu (1989) introduced IF ('independence-friendly') languages designed to represent arbitrary patterns of dependences and independences of quantifiers, connectives and other logical operators. In that programmatic paper, we considered two kinds of examples.

First, we wanted to represent quantifier patterns as in the paradigmatic formula

$$\forall x \forall y (\exists z/\{\forall y\})(\exists w/\{\forall x, \exists z\}) R(x, y, z, w), \tag{6}$$

where the idea is that

1. $\exists z$ is only in the logical scope of $\forall x$ (and not in the logical scope of $\forall y$), and
2. $\exists w$ is only in the logical scope of $\forall y$ (and not in the logical scopes of $\forall x$ and $\exists z$).

We have here a partial order of the four quantifiers, an idea which goes back to Henkin (1961) who used a different notation:

$$\begin{pmatrix} \forall x & \exists z \\ \forall y & \exists w \end{pmatrix} R(x, y, z, w). \tag{7}$$

Drawing on some earlier work of Hintikka, we were persuaded that natural-language provides plenty of examples involving partially ordered quantifiers as well as partially ordered modal (tense) operators standardly interpreted as quantifiers over possible worlds (temporal moments). Examples are:

> John believes that there are people who persecute him, but some of them are in reality merely trying to get his autograph. (8)

> Once I did not believe that I would now be living in Tallahassee. (9)

In (9), for instance, the second tense operator (*would*) is in the logical scope of the epistemic operator (*believe*) but not in the logical scope of the tense operator (*did*). The semantical role of *now* and *really* is to get one back to the actual world or the present moment of time. Inspired by Hintikka's and Saarinen's work in the seventies (see Saarinen 1979), I devised a game interpretation for such prefixes of modal operators which I labelled the *back-looking interpretation*. The thing to be emphasised from the perspective of the present paper is that, indeed, the back-looking interpretation is an example of the partial ordering of the logical scopes of modal operators but it does not, however, throw light on the two notions of scope.

The second group of examples we focussed on involved, in addition to partially-ordered quantifiers and modal operators, richer combinations of quantifiers, modal operators and connectives such as those in the following example inspired by David Kaplan:

$$\Diamond[\forall x(A(x)] \rightarrow B(x)). \tag{10}$$

Here the diamond is the possibility operator. The brackets indicate that the diamond is in the logical scope of the implication. Formula (10) is ill-formed on the standard logical syntax, for the binding scope of the universal quantifier extends over implication which logically dominates the diamond which in turn dominates the universal quantifier. The meaning we associated with this formula is: there is at least one alternative world such that, whatever is an A in this world is, as a matter of fact (= in the actual world), a B. Hintikka & Sandu (1989) did not contain a detailed analysis of (10), but we thought it provided an intuitive example of a class of logical constants on which the relation of 'being in the logical scope of' is not well-founded:

- '$\forall x$' is in the logical scope of the diamond.
- '$\Diamond$' is in the logical scope '$\rightarrow$'.
- '$\rightarrow$' is in the logical scope of '$\forall x$'.

Diagrammatically:

$$\Diamond \geq \forall x \geq \rightarrow \geq \Diamond. \tag{11}$$

Unlike in the previous examples, which are notational variants of Henkin's partially ordered quantifiers and for which a natural interpretation can be given in terms of games of imperfect information (Pietarinen 2001 gives a good survey of the literature), we have never been able to devise an interpretation for these kinds of circular prefixes. In fact, I now think it is really difficult to make sense of (10) if the logical scopes of the constants occurring in it have this kind of circular dependence. This is one of the reasons why it may be useful to think of examples like (10) in a different way.

We may think of (10) as presupposing already the two notions of scope, taking the logical scopes of the quantifier, implication and diamond to be totally ordered:

$$\rightarrow \geq \Diamond \geq \forall x. \tag{12}$$

That is, '$\forall x$' is in the logical scope of the diamond which is in the logical scope of the implication, but the binding scope of the quantifier extends over the implication in order to reach the variable x in $B(x)$. In this case, the ordering of the logical scopes of the connectives would not bring a solution to the problem of the two notions of scope, but the other way around: the two notions of scope are taken as primitive. This allows us to account for examples such as (10), provided, of course, that we find a natural interpretation for the two notions. As a matter of fact, this is how it should be: For how could a relation which affects only the ordering of the logical scopes of quantifiers and connectives also affect their binding scopes?

In what follows I am going to sketch a game interpretation which distinguishes the two kinds of scopes and which yields, for certain sentences, the same prediction as dynamic predicate logic (Groenendijk & Stokhof, 1991). In other words, I am going to sketch a version of dynamic GTS.

One remark is in order here. The fact that the two notions of scope do not coincide, and thereby that first-order logic cannot, given its standard interpretation, account for the intended reading of these sentences, can be taken in two ways. Either the analogy between the pairs ⟨head, anaphor⟩–⟨quantifiers, variables⟩ should be given up (Neale, 1990), or one should look for alternative logical representations. I will opt for the latter, pointing out, however, some of its limitations.

I will begin by a short sketch of two alternative approaches suggested to the problem at hand.

3 Discourse Representation Theory (DRT)

In the original version of DRT (Kamp 1981, Kamp & Reyle 1993), natural-language discourse is the input to a construction algorithm which converts its sentences into *Discourse Representation Structures* (DRS). DRSs are specified in a formal language L consisting of a set A of individual constants together with a set of n-place predicate constants to which we add a set U of discourse markers $x_1, x_2, \ldots$. The individual constants and the discourse markers form the set of terms of our language L extended with U.

A DRS consists of a set of discourse markers and a set of atomic or complex conditions. We will not be concerned here with all the intricacies of DRT but demonstrate instead its application to sentences (2) and (3).

In the first stage, the first subsentence of (2) (*A girl smiles*) is processed, the result being:

$$D_1 : (\{x\}, \{\mathit{girl}(x), \mathit{smiles}(x)\}) . \tag{13}$$

Next, the second sentence (*She is happy*) is processed:

$$D_2 : (\{y\}, \{\mathit{happy}(y), x = y\}) . \tag{14}$$

Finally, we have an operation ';' that *merges* the two DRSs into a single one:

$$D_1; D_2 = (\{x, y\}, \{\mathit{girl}(x), \mathit{smiles}(x), \mathit{happy}(y), x = y\}). \tag{15}$$

A DRS is true in a model M if and only if there is an assignment from a set of markers to the entities of the domain of M that satisfies all the conditions of the DRS. In the present case, this means that $D_1; D_2$ is true if and only if there is an assignment $g\colon \{x, y\} \to \mathrm{Dom}(M)$ such that the conditions $\mathit{man}(x)$, $\mathit{entered}(x)$, $\mathit{smiled}(y)$ and $x = y$ are satisfied. The reader may check that, in the model M, this interpretations makes (2) materially equivalent to the first-order sentence

$$\exists x(G(x) \land S(x) \land H(x)) . \tag{16}$$

The DRS corresponding to (2) is formed by an operation which combines D_1 and D_2 into a single DRS, $D_1 \Rightarrow D_2$, defined via the stipulation

$$D_1 \Rightarrow D_2 := \neg(D_1; \neg D_2) . \tag{17}$$

Again, it may be checked that this interpretation renders (3) materially equivalent to the sentence

$$\forall x(G(x) \land S(x) \to H(x)) . \tag{18}$$

4 Dynamic Predicate Logic (DPL)

Usually some version of dynamic logic is used in the formalisation of reasoning about programs which, for the present purpose, are taken to consist of sets of pairs of assignments in a model M, where an assignment is a function from the set of variables to the universe of the model. A formula of predicate logic is then interpreted as a set of pairs of assignments. Roughly, a pair (g, h) is in the interpretation of the formula φ if and only if, when φ is evaluated with respect to g, h is a possible outcome of the evaluation procedure. We shall be interested only in a few clauses relevant to the present discussion.

According to this interpretation, the formulas are divided into two classes:

$$||\psi \land \theta|| = \{(g, h) : \exists k((g, k) \in ||\psi|| \text{ and } (k, h) \in ||\theta||)\} \tag{19}$$

$$||\exists x\psi|| = \{(g, h) : \exists k(k[x]g \text{ and } (k, h) \in ||\psi||)\} \tag{20}$$

$$||\psi \to \theta|| = \{(g, h) : h = g \text{ and } \forall k((g, k) \in ||\psi|| \implies \exists j((k, j) \in ||\theta||))\} . \tag{21}$$

We may check that

$$||\exists x P(x) \wedge Q(x)|| = \{(g, h) : (h(x) \in P^M \text{ and } h(x) \in Q^M)\}. \tag{22}$$

(Here P^M is the interpretation of P in M.) Satisfaction in M with respect to the assignment s ($M \models_s \varphi$) is defined as

$$M \models_s \varphi \quad \text{if and only if} \quad \exists g((g, s) \in ||\psi||). \tag{23}$$

It can be checked that the satisfaction of $\exists x P(x) \wedge Q(x)$ in DPL is equivalent to the satisfaction in standard predicate logic of $\exists x(P(x) \wedge Q(x))$, and the same holds of the pair $\exists x P(x) \rightarrow Q(x)$, $\forall x(P(x) \rightarrow Q(x))$. Thus the interpretation of $\exists x P(x) \wedge Q(x)$ nicely distinguishes between the two scopes: the conjunction has logical priority over the existential quantifier whose binding scope extends to the right conjunct. Moreover, unlike in the DRT approach, the result is achieved with the syntax of first-order languages (cf. Groenendijk & Stokhof 1991).

5 Game-Theoretical Semantics for Formal Languages

5.1 Standard Interpretation

We fix a standard first-order language containing $\wedge$, $\vee$, $\neg$, $\forall x$ and $\exists x$ as its logical constants. The full game-theoretical interpretation for this language goes back to Hintikka & Kulas (1983). An alternative interpretation in terms of games in extensive form was first explicitly introduced in Sandu & Pietarinen (2001) in connection with IF-languages. The extensive-form of a game represents a semantical game $\mathcal{G}(\varphi, M, s)$ of perfect information (s is an assignment) as a set of histories build up according to the rules of the game. Here I prefer to represent these rules in the form of tableau rules for building up semantic trees for classical logic in the style of Evert Beth and Jaakko Hintikka. Recall that these semantic tableaux are formed by associating each connective and quantifier with its own rule. For example:

$$\begin{tikzcd} A \wedge B \arrow[d] \\ A \arrow[d] \\ B \end{tikzcd} \tag{24}$$

$$\begin{tikzcd} & A \vee B \arrow[dl] \arrow[dr] & \\ A & & B \end{tikzcd} \tag{25}$$

Let s be an assignment in the relevant model which, for the sake of simplicity, contains only two elements, a and b. The game rules for the first-order language are given below. The labellings '$\exists$' and '$\forall$' indicate the player who is making the move, and 'c' indicates an exchange of the roles of the players.

Conjunction:

$$\begin{tikzcd} & \mathrm{A} \land \mathrm{B}, \mathrm{s} \arrow[dl] \arrow[dr] \arrow[d, phantom, "\forall"] & \\ \mathrm{A}, \mathrm{s} & & \mathrm{B}, \mathrm{s} \end{tikzcd} \tag{26}$$

(Player ∀ chooses one of the conjuncts and the play goes on with the chosen conjunct and the initial assignment.)

Disjunction:

$$\begin{tikzcd} & \mathrm{A} \lor \mathrm{B}, \mathrm{s} \arrow[dl] \arrow[dr] \arrow[d, phantom, "\exists"] & \\ \mathrm{A}, \mathrm{s} & & \mathrm{B}, \mathrm{s} \end{tikzcd} \tag{27}$$

Negation:

$$\begin{tikzcd} \neg \mathrm{B}, \mathrm{s} \arrow[d, "\mathrm{c}"] \\ \mathrm{B}, \mathrm{s} \end{tikzcd} \tag{28}$$

(Players exchange roles for the remaining part of the game.)

Existential quantification:

$$\begin{tikzcd} & \exists x \mathrm{B}, \mathrm{s} \arrow[dl] \arrow[dr] \arrow[d, phantom, "\exists"] & \\ \mathrm{B}, \mathrm{s}(x/a) & & \mathrm{B}, \mathrm{s}(x/b) \end{tikzcd} \tag{29}$$

(Player ∃ chooses one of the elements of the universe of the model to be the interpretation of x.)

Universal quantification:

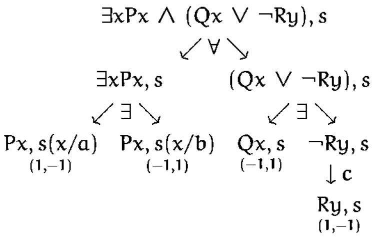
$$\begin{tikzcd} & \forall x \mathrm{B}, \mathrm{s} \arrow[dl] \arrow[dr] \arrow[d, phantom, "\forall"] & \\ \mathrm{B}, \mathrm{s}(x/a) & & \mathrm{B}, \mathrm{s}(x/b) \end{tikzcd} \tag{30}$$

If the universe of the model is infinite, then the last two rules would result in an infinite number of branches.

Each application of one of the above rules reduces the complexity of the formula to which the rule is applied. A maximal branch is a branch to which no rule can be applied any longer because it is labelled by (C, t), C an atomic formula. Each maximal branch represents a play of the game, and it results in a win for exactly one of the two players. It is a win for the player ∃ under exactly two conditions: (i) the assignment t satisfies the formula C and the play contains an even number of role exchanges; and (ii) t does not satisfy C and the play consists an odd number of role exchanges. Otherwise, the maximal branch is a win for the player ∀, which amounts to the dual cases of (i) and (ii).

The crucial notion is that of the truth in M (the existence of a winning strategy for ∃) and the falsity in M (the existence of a winning strategy for ∀). The former is defined as a method for player ∃ to win every play against any move of his opponent. And likewise for falsity. This notion can be made more precise, but we prefer to illustrate it with an example. Consider the game played with the formula $\exists x Px \wedge (Qx \vee Ry)$, the model M with the universe $\{a, b\}$ such that $P^M = \{a\}$, $Q^M = \{b\}$, and $R^M = \{b\}$, and the assignment s such that $s(x) = a$ and $s(y) = a$.

Here is the game tree:



(31)

The winning strategy for the player ∃ is: If ∀ chooses left, then ∃ chooses left; if ∀ chooses right, then ∃ chooses right.

The following propositions are easily proved for any φ, model M and an assignment g:

Proposition 1.

1. Player ∃ has a winning strategy in $\mathcal{G}(\varphi, M, g)$ iff player ∀ has a winning strategy in $\mathcal{G}(\neg\varphi, M, g)$.

2. Player ∃ has a winning strategy in $\mathcal{G}(\neg\varphi, M, g)$ iff player ∀ has a winning strategy in $\mathcal{G}(\varphi, M, g)$.

Proof. The two games are identical except that the moves done in one of the games by one player are done in the other by the opponent, and vice versa. Also, the rules of winning and losing are likewise reversed. □

We perform the following operations on φ:

- Replace every quantifier ∃ (resp. ∀) by ∀ (resp. ∃);
- Replace every connective ∨ (resp. ∧) by ∧ (resp. ∨);
- Prefix every atomic subformula in φ with a negation sign ¬;
- Erase the negation ¬ from every negated atomic subformula of φ.

Call the result φ^*. Now we have:

Proposition 2.

1. Player ∃ has a winning strategy in $\mathcal{G}(\varphi, M, g)$ iff player ∀ has a winning strategy in $\mathcal{G}(\varphi^*, M, g)$.

2. Player $\forall$ has a winning strategy in $\mathcal{G}(\varphi, M, g)$ iff player $\exists$ has a winning strategy in $\mathcal{G}(\varphi^*, M, g)$.

Proof. Analogous to that of Proposition 1. □

Proposition 3 (Negation Normal Form).

1. Player $\exists$ has a winning strategy in $\mathcal{G}(\neg\varphi, M, g)$ iff player $\exists$ has a winning strategy in $\mathcal{G}(\varphi^*, M, g)$.
2. Player $\forall$ has a winning strategy in $\mathcal{G}(\neg\varphi, M, g)$ iff player $\forall$ has a winning strategy in $\mathcal{G}(\varphi^*, M, g)$.

Proof. Suppose player $\exists$ has a winning strategy in $\mathcal{G}(\neg\varphi, M, g)$. Then by Proposition 1, player $\forall$ has a winning strategy in $\mathcal{G}(\varphi, M, g)$ and by Proposition 2, player $\exists$ has a winning strategy in $\mathcal{G}(\varphi^*, M, g)$. □

5.2 Dynamic Interpretation

We add to the syntax of first-order logic two dynamic connectives: ';' for dynamic conjunction and '$\Rightarrow$' for dynamic implication. The syntax is enriched with the clause:

- If A and B are formulas, so are $(A; B)$ and $(A \Rightarrow B)$.

We could actually operate directly on the syntax of first-order logic and give a dynamic interpretation of conjunction and implication, as is done in the DPL approach, but we prefer to have both a 'static' and 'dynamic' conjunction and implication.

The tableau rule for dynamic conjunction is:

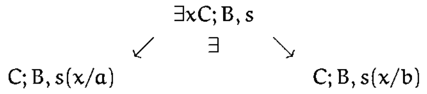
$$\begin{tikzcd} & A; B, s \arrow[dl] \arrow[dr] \arrow[d, phantom, "\forall"] & \\ A, s & & B, s \end{tikzcd} \tag{32}$$

A may be an atomic formula, the negation of an atomic formula, or a universally quantified formula. Notice that in all these cases the dynamic conjunction $A; B$ is treated as standard conjunction.

Here are the remaining clauses:

$$\begin{tikzcd} & (C \vee D); B, s \arrow[dl] \arrow[dr] \arrow[d, phantom, "\exists"] & \\ C; B, s & & D; B, s \end{tikzcd} \tag{33}$$

$$\begin{tikzcd} & (C \wedge D); B, s \arrow[dl] \arrow[dr] \arrow[d, phantom, "\forall"] & \\ C; B, s & & D; B, s \end{tikzcd} \tag{34}$$

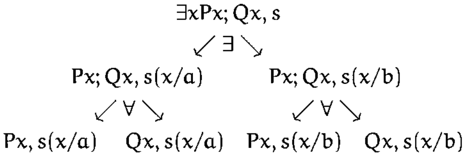


(35)

$$\begin{tikzcd}(C;D);B,s \arrow[d] \\ C;(D;B),s\end{tikzcd} \tag{36}$$

The clause $\exists x C; B, s$ deserves special attention. The conjunction does not have logical priority over the existential quantifier: if that were to be the case, it would be $\forall$ who would choose a conjunct. Before that happens, the assignment s is extended to bind the variable x which will eventually occur in B.

Again, an example will help to see the impact of the rule. As before, we take the universe of the model to consist only of two elements, $\{a, b\}$, but the resulting interpretation holds for the general case. The assignment s is arbitrary. Here is the game tree:



(37)

We may convince ourselves that in the general case, player $\exists$ has a winning strategy in the game associated with the model M, the formula $\exists x Px; Qx$, and the assignment s, if and only if there is an individual c such that both Px and Qx are satisfied when x takes the value c. In other words:

$$M, s \models_{\text{GTS}} \exists x Px; Qx \text{ if and only if } M, s \models_{\text{Tarski}} \exists x(Px \land Qx). \tag{38}$$

Finally, we have a rule for A having the form of $(C; D)$:

$$(C; D); B \quad \text{is played as} \quad C; (D; B).$$

This rule allows to decrease the complexity of the left-hand side formula concerning dynamic conjunction. To take an example, when the formula $(C; D); (B \land E)$ is reached, the game is played as $C; (D; (B \land E))$.

Dynamic implication '$\Rightarrow$' does not correspond to a game-rule but is taken to be a defined symbol:

$$A \Rightarrow B := \neg(A; \neg B). \tag{39}$$

The game tree of $G(\exists xPx \Rightarrow Qx, M, s)$, where M and s are as before is:

$$\begin{tikzcd}
 & & \neg(\exists xPx;\neg Qx), s \arrow[d, "c"] & & & \\
 & & (\exists xPx;\neg Qx), s \arrow[dl, "\forall"'] \arrow[drr] & & & \\
 & Px;\neg Qx, s(x/a) \arrow[dl, "\exists"'] \arrow[dr] & & & Px;\neg Qx, s(x/b) \arrow[dl, "\exists"'] \arrow[dr] & \\
Px, s(x/a) & & \neg Qx, s(x/a) & Px, s(x/b) & & \neg Qx, s(x/b)
\end{tikzcd} \tag{40}$$

Given the negation, one may verify that the existential player has a winning strategy in the game, iff she can win against any move by her opponent. Namely, no matter which element d her opponent produces, she is able to choose a formula which results in a win for her, that is, $\neg P(d) \vee Q(d)$. In other words:

$$M, s \models_{GTS} \exists xPx \rightarrow Qx \text{ if and only if } M, s \models_{\text{Tarski}} \forall x(Px \rightarrow Qx)\,.$$

References

Carlson, L. and A. ter Meulen (1979). Informational independence in intensional contexts. In: *Essays in Honor of Jaakko Hintikka* (E. Saarinen, ed.), pp. 61-72. D. Reidel, Dordrecht.

Groenendijk, J. and M. Stokhof (1991). Dynamic predicate logic. *Linguistics and Philosophy*, **14**, 39-100.

Henkin, L. (1961). Some remarks on infinitely long formulas. In: *Infinistic Methods. Proceedings of the Symposium on Foundations of Mathematics, Warsaw, Panstwowe.* pp. 167–183. Wydawnictwo, Naukowe, Pergamon Press, New York.

Hintikka, J. (1987). Is scope a viable concept in semantics? In: *Proceedings of the Third Eastern States Conference on Linguistics*, 259-270.

Hintikka, J. (1996). *The Principles of Mathematics Revisited.* Cambridge University Press, New York.

Hintikka, J. and J. Kulas (1983). *The Game of Language.* D. Reidel, Dordrecht.

Hintikka, J. and J. Kulas (1985). *Anaphora and Definite Descriptions: Two Applications of Game-theoretical Semantics.* D. Reidel, Dordrecht.

Hintikka, J. and G. Sandu (1989). Informational independence as a semantical phenomenon. In: *Logic, Methodology and Philosophy of Science* (J. Fenstad et al., eds.), pp. 571-589. Elsevier.

Hintikka, J. and G. Sandu (1997). Game-theoretical semantics. In: *Handbook of Logic and Language* (J. van Benthem and A. ter Meulen, eds.), pp. 361-410. Elsevier, Amsterdam.

Hogdes, W. (1997). Compostional semantics for a language with imperfect information. *Logic Journal of the IGPL*, **5**, 539-563.

Kamp, H. (1981). A theory of truth and semantic representation. In: *Formal Methods in the Study of Language* (J. Groenendjik, T. M. V. Janssen and M. Stokhof, eds.), pp. 277-322. Mathematisch Centrum Tracts, Amsterdam.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic*. Kluwer, Dordrecht.

Neale, S. (1990). *Descriptions*. MIT Press, Cambridge, Mass.

Pietarinen, A.-V. (2001). *Semantic Games in Logic and Language*. Doctoral Dissertation, University of Helsinki.

Saarinen, E. (ed.) (1979). *Game-Theoretical Semantics: Essays on Semantics by Hintikka, Carlson, Peacocke, Rantala, and Saarinen*. D. Reidel, Dordrecht.

Sandu, G. and T. Janasik (2003). Dynamic game semantics. In: *Meaning: The Dynamic Turn* (J. Peregrin, ed.), pp. 215-240. Elsevier, Amsterdam.

Sandu, G. and A.-V. Pietarinen (2001). Partiality and games: Propositional logic. *Logic Journal of the IGPL*, **9**, 107-127.

Sandu, G. and A.-V. Pietarinen (2003). Informationally independent connectives. In: *Games, Logic, and Constructive Sets* (G. Mints and R. Muskens, eds.), pp. 23-41. CSLI Publications, Stanford.

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 13

SEMANTIC GAMES AND GENERALISED QUANTIFIERS

Ahti-Veikko Pietarinen
University of Helsinki

This chapter proposes to marry generalised quantifiers with game-theoretic semantics (GTS). It is argued that generalised quantifiers are no impediment to a game-theoretic interpretation. To this effect, semantic game rules for various types of generalised quantifiers are defined. Moreover, game semantics is argued to surpass relational semantics in that it provides (i) a generic method of dealing with context-dependent quantifiers in terms of strategic content and (ii) a general semantics for branching generalised quantifiers.

1 INTRODUCTION

An appealing alternative to the standard relational semantics for generalised quantifiers is game-theoretic semantics (GTS).[1] Towards this end, in this chapter some semantic game rules for determiner phrases that generalise the existential and universal quantifiers are formulated. It is argued that, minimally, GTS is no worse off when paralleled with standard relational semantics. Furthermore, GTS provides a general, systematic and dynamic framework for dealing with complex quantifier phrases, especially with reference to context-dependent quantifiers such as the complement quantifiers *most... the rest* or *three... the others*, and with reference to quantified phrases in sentences or discourse involving branching generalised quantifiers with reciprocal phrases.

The game-theoretic interpretation has its roots in Charles S. Peirce's (1839–1914) logical studies (Hilpinen, 1982; Pietarinen, 2005). What has not been recorded before is that Peirce recognised the importance of not only first-order but also generalised quantification. These two approaches can now be merged. Therefore, in order to put the discussion into a historical perspective, the development of these logical ideas is briefly surveyed in the appendix.

[1] For original studies on the relational semantics for generalised quantifiers, see Mostowski (1957), Lindström (1966), Barwise & Cooper (1981) and Higginbotham & May (1981), and on GTS see Hintikka & Kulas (1983), Hintikka & Kulas (1985), Hintikka & Sandu (1991, 1997), and Saarinen (1979). For recent publications, see e.g. Sandu & Janasik (2003), Pietarinen (2001a, 2004b) and Sandu & Pietarinen (2001).

2 GTS for Natural Language

Any sentence of English defines a game between two players, the Verifier (V, H∃loïsé, Myself) and the Falsifier (F, ∀bélard, Nature). V strives to show that the given sentence is true in a given model, and F strives to show that the sentence is false in it.[2] The game rules for quantificational expressions such as *some*, *every*, *a(n)* and *any* prompt a player to choose an individual from the relevant domain (choice set) I, labelling the individual with a name if it does not have one already. The game continues with respect to an output sentence defined by the game rules.

Analogously to semantic games correlated with expressions of formal languages, a play of the game terminates when such components (corresponding to atomic formulas) are reached where further applications of game rules are no longer possible. Their truth in a given interpretation determines whether V (atomic truth) or F (atomic falsity) wins.

An example of the game rule for *some* is as follows.

(G.some) If the game has reached a sentence of the form

X – *some* Y who Z – W,

then the verifier V chooses an individual from I, say b. The game continues with respect to the sentence

X – b – W, b is a Y, and b Z.

Here "who Z" (or "where Z", "when Z" etc.) is the entire relative clause, and the main verb phrase W and the head noun in Y are in the singular. For simplicity, the relative clause markers are mostly omitted. The linguistic context X may be arbitrary.

The rule for negation is as follows:

(G.not) If the game has reached a sentence of the form neg(A), the players exchange roles (and also the winning conventions will change). The game continues with respect to A.

The operation neg(A) is a functor forming sentential negations of A.

A strategy is a complete rule telling players at any contingency what the actions of players are. If a strategy dominates, in other words if it leads a player to a winning position no matter how the opponent chooses to play, that strategy is a winning one. In logic, the existence of winning strategies for V in a given model is equated with truth, and the existence of winning strategies for F in a given model is equated with falsity.

Since the notion of scope does not surface in the syntactic structure of natural-language sentences, GTS has traditionally used ordering principles to provide the needed orders of application for the game rules. In the syntactic structure, a node n_1 is said to be in a higher clause than (or "c-commands") the node n_2 if the first branching node immediately dominating n_1 also dominates n_2, but not vice versa.

The following general ordering principles may thus be derived (see e.g. Hintikka & Kulas 1983; Scheffler this volume):

(O.left-to-right) For any two phrases in the same clause, a game rule must not be applied to the one on the right if a rule can be applied to the one on the left in the clause.

[2] See Pietarinen (2003b) for a cross-section of the kinds of games that have recently been operationalised in logic and in science.

(O.c-command) A game rule must not be applied to a phrase in a lower clause if a rule can be applied to a phrase in a higher clause.

Special ordering principles may override general ones. For instance, the following special ordering principles may be applied:

(O.any) (G.any) has priority over rules such as (G.not), (G.conditional), (G.or) and some modal rules such as (G.can), (G.may), (G.possible) and (G.likely).

(O.some) (G.some) has priority over (G.not).

The semantic-game derivation of the meaning of sentences may be represented in the form of a tree, beginning with the complete sentence or discourse at the root $r \in H$ and ends with the terminal histories in $Z \subseteq H$ associated with those expressions that correspond to atomic formulas.[3]

In game theory, these explicit representations of total derivational histories recording the past actions in the game are generally known as extensive forms. Extensive-form games provide a rich structure for the meaning of discourse, as they replace the notion of a choice set I with the notion of 'accessible actions' that the players may pick during the game from sets of elements associated with non-terminal histories $H \setminus Z$. Terminal histories are mapped by payoff functions $u_i \colon Z \to \{1, -1\}$ to wins (1) and losses (−1) for a player $i \in \{V, F\}$. Non-terminal histories may also be interspersed with deictic elements from the environment, such as those given by initial chance moves by a third player, Nature.[4]

3 Definite Descriptions

One specific issue to which GTS has been applied is the semantics of definite descriptions (Hintikka & Kulas, 1985). This analysis has some merit over the treatment of *the* in standard theories of generalised quantifiers.

The main game rules for definite descriptions to this effect are the following:

(G. Russellian *the*) When a game has reached a sentence of the form

$X - \textit{the}\ Y$ who $Z - W$,

an individual, say b, is chosen by Myself, whereupon a different individual, say d, is chosen by Nature. If these individuals do not already have names, the players give them names, which are assumed to be 'b' and 'd'. The game is then continued with respect to

$X - b - W$, b is a Y, b Z, but d is not a Y who Z. (Hintikka & Kulas, 1985, pp. 37–38)

(G. anaphoric *the*) Like (G. Russellian *the*), but the selections are relativised to a choice set I whose members have been introduced earlier in the game by either player. (Hintikka & Kulas, 1985, p. 48)

[3] See Janasik et al. (2003) and Pietarinen (2004b) for details.

[4] Nature does not mean here the Falsifier, who is often called that in the literature. See my "The semantics/pragmatics distinction from the game-theoretic point of view" in this volume on a trichotomy contexts that follows from the game-theoretic perspective on linguistic meaning.

Observe how the game-theoretic approach to definite descriptions differs from the usual determiner rules, such as those first proposed in Barwise & Cooper (1981):

$$M \models \mathit{The}_{E}^{\text{singular}}(A, B) \quad \text{if and only if} \quad \mathit{All}_{E}(A, B) \text{ and } |A| = 1$$
$$M \models \mathit{The}_{E}^{\text{plural}}(A, B) \quad \text{if and only if} \quad \mathit{All}_{E}(A, B) \text{ and } |A| > 1.$$

How are we to extend these rules to apply on the level of discourse? The answer is not forthcoming from standard relational semantics. In GTS, on the other hand, the answer lies in the strategic content and the accessibility of actions in earlier parts of the histories of the games.[5] Supplemented with strategic precepts that guide players actions, GTS spells out a dynamic interpretation of a variety of meanings for definite descriptions.

The sense here in which GTS may be seen to be richer than the standard theory of generalised quantifiers is that games allow for various pragmatic and discourse-oriented overtones in the incorporation of a wealth of collateral information, syntactic clues, and other contextual features into the content of players' strategies.

4 Relational Semantics

Kalish & Montague (1964) and Montague (1969) were among the first to propose to interpret natural-language NPs as generalised quantifiers. The early work of William Woods (1968) in the context of machine-assisted natural-language translation should also be mentioned, as well as the much earlier source of Stanislaw Leśniewski (Simons, 1994). According to the received approach first formulated in Mostowski (1957) and Lindström (1966), and systematised for linguistic applications in Barwise & Cooper (1981) and Higginbotham & May (1981), a quantifier is a relation on the power set of a domain E satisfying certain constraints, such as extensionality (EXT), conservativity (CONS), universality (UNIV) and isomorphism (ISOM).

This semantics is relational, since it interprets quantifiers in terms of relations between individuals or between relations of individuals. In other words, it interprets generalised quantifiers as higher-order relations and so takes NPs to be set-theoretic constructions.[6]

For example, according to a garden-variety of such relational semantics, the sentence *All men walk* is true in M if and only if the set of men A in the domain E of M is included in the set of walkers B in E. A binary relation is asserted to hold between A and B, notated by $D_E(A, B)$, in which D_E is a determiner that picks out the relation between subsets of the domain E, namely that $D_E \in \wp(\wp(E) \times \wp(E))$.

Those D_E that count as determiner denotations in natural language are typically taken to satisfy a few basic properties, such as EXT, CONS, UNIV and ISOM:

- EXT (Domain Independence): If $\{A_1 \ldots A_n\} \subseteq E \subseteq E'$ then $D_E(A_1, \ldots, A_n)$ if and only if $D_{E'}(A_1, \ldots, A_n)$. This property is intended to capture that the part of the universe E' lying outside the noun denotation is irrelevant to the meaning of the determiner.
- CONS (Domain Connectivity): $D_E(A_1, \ldots, A_n, B)$ if and only if $D_E(A_1, \ldots, A_n, (A_1 \cup$

[5] See Janasik et al. (2003) for further elaboration on how to encode the notion of accessible (or remembered) strategies into the description of extensive-form semantic games associated with anaphoric discourse. Sandu & Janasik (2003) suggest a dynamic treatment of singular anaphora in GTS. See also Sandu and Clark this volume.

[6] If the sets that function as values of quantificational notions are not logically individual objects, the generalised quantifiers are in fact indexical and their interpretation must, therefore, be substitutional (see Appendix).

$\cdots \cup A_n) \cap B)$. This property is intended to capture that only the part of the verb argument common to the noun argument matters to the meaning of the determiner.

- UNIV (Domain Restriction): CONS + EXT.

- ISOM (Topic Neutrality): If f is a bijection from E to E' then $D_E\ (C_1, \ldots, C_n)$ if and only if $D_{E'}\ (f(C_1), \ldots, f(C_n))$. In words, D's do not distinguish between different elements of the universe or universes.[7]

A number of determiners that violate one of these constraints has been proposed:

- *Many, few* and *not enough* depend on E and so do not satisfy EXT.

- *Only* (if a determiner at all) denotes a superset relation and so does not satisfy CONS; exceptives such as *all but* or *most... except* denote a relation between the verb phrase and the complement of the NP and so do not satisfy CONS.

- Possessives (e.g. *John's*$_E$ (A, B)) depend on the set P_{Subject}, where P_i is the set of things possessed by the subject i, and so do not satisfy ISOM. Likewise, the meaning of exceptives depends on the very elements that constitute such exceptions.

Among the facts that these constraints aim at capturing is that the subject of the sentence bears some special function in relation to its predicate.

As far as binary determiners are concerned, which were not considered in Mostowski (1957), the effect of $D_E\ (A, B)$ may be described by the difference $|A - B|$ and intersection $|A \cap B|$. For n-ary D_E's that satisfy CONS, EXT, and ISOM, difference and intersection are intended to separate determiners into classes of logical and non-logical ones.

My remark here concerns the overall methodological significance of these constraints. For example, Barwise & Cooper (1981) derive them as empirical generalisations from data. They were not motivated by the properties of the theory of generalised quantifier per se. In the light of GTS, however, conservativity stems from the idea that, once a suitable game rule has been applied to an expression, the game proceeds with respect to the sentence in which the denotation of the subject noun captures, or relativises, the respective domain $E^* \subseteq E$ from which the players are to seek and pick elements that will satisfy the predicate property. This may be done irrespectively of any need to look outside a given restricted portion of the domain in question. Hence, CONS receives a theoretic backing by GTS in terms of the concrete actions taken by the players. Similar motivations may be envisaged for other constraints, such as EXT and thus UNIV.

The interpretational rules for some commonplace determiners in relational semantics include the following:

[7] Sometimes also termed QUANT for monadic determiners.

$$\begin{array}{lll}
\mathrm{M} \models \mathit{All}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & \mathrm{A} \subseteq \mathrm{B} \\
\mathrm{M} \models \mathit{Some}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & \mathrm{A} \cap \mathrm{B} \neq \emptyset \\
\mathrm{M} \models \mathit{No}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & \mathrm{A} \cap \mathrm{B} = \emptyset \\
\mathrm{M} \models \mathit{Fewer\ than\ five}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & |\mathrm{A} \cap \mathrm{B}| < 5 \\
\mathrm{M} \models \mathit{All\ but\ two}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & |\mathrm{A} - \mathrm{B}| = 2 \\
\mathrm{M} \models \mathit{Neither}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & |\mathrm{A}| = 2 \text{ and } \mathrm{A} \cap \mathrm{B} = \emptyset \\
\mathrm{M} \models \mathit{Most}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & |\mathrm{A} \cap \mathrm{B}| > |\mathrm{A} - \mathrm{B}| \\
\mathrm{M} \models \mathit{More}_{\mathrm{E}_1}\ \mathit{than}_{\mathrm{E}_2}\,(\mathrm{B},\mathrm{C}) & \text{if and only if} & |\mathrm{A} \cap \mathrm{C}| > |\mathrm{B} \cap \mathrm{C}| \\
\mathrm{M} \models \mathit{John's}^{\text{singular}}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & \mathit{All}_{\mathrm{E}}\ \mathrm{P}_{\text{John}} \cap (\mathrm{A},\mathrm{B}) \text{ and } |\mathrm{P}_{\text{John}} \cap \mathrm{A}| = 1 \\
\mathrm{M} \models \mathit{John's}^{\text{plural}}_{\mathrm{E}}\,(\mathrm{A},\mathrm{B}) & \text{if and only if} & \mathit{All}_{\mathrm{E}}\ \mathrm{P}_{\text{John}} \cap (\mathrm{A},\mathrm{B}) \text{ and } |\mathrm{P}_{\text{John}} \cap \mathrm{A}| > 1.
\end{array}$$

Simple as these rules are, as such they are not intended to account for the inevitable contextual and pragmatic aspects arising in natural language. Several refinements to these rules have thus been proposed. The hallmark of the relational meaning of quantifiers nonetheless is that they apply to sets and subsets of the objects of the domain by stipulating some specific conditions and thus properties of those sets and relations between them as denotations according to some given sentence.

To recap, a generalised quantifier defines relations between subsets of the domain of the model. According to the standard notation, *type* $\langle n_1 \ldots n_k \rangle$ is a finite sequence of positive numbers indicating how many and which relations there are. For instance, a monadic predicate P^1 denoting a subset of a domain E gives rise to a generalised quantifier of type $\langle 1 \rangle$, which maps $\wp(E)$ into a set of truth-values $\{F, T\}$. A monadic determiner gives rise to a type $\langle 1, 1 \rangle$ quantifier, which maps $\wp(E)$ into $\langle 1 \rangle$. Pairs of properties map $\wp(E) \times \wp(E)$ into $\langle 1 \rangle$, giving rise to generalised quantifiers of type $\langle \langle 1, 1 \rangle, 1 \rangle$.

A further division is routinely made between *monadic* and *polyadic* quantification. As to the monadic quantifiers, the arguments are sets that are interpretations of nouns and intransitive verbs. When NPs are taken as objects of transitive and ditransitive verbs, polyadic generalised quantifiers may be obtained.

Landman (2000) has passed judgement on this division and argues that polyadic quantifiers cannot be based on similar logicality principles than monadic ones, and, in particular, cannot be composed of them in any trivial manner. Some replies to these complaints are outlined in the concluding section.

In the remaining sections it is argued that, at the very least, GTS equals the relational attempts, and that there are perspectives in which GTS may be seen to tackle questions that have not been—and in some cases to be considered in Section 6 are likely to be difficult to be—adequately addressed from the perspective of relational semantics.

5 GAME RULES FOR GENERALISED QUANTIFIERS

Let us begin with some examples of semantic rules for monadic quantifiers. For simplicity, it is assumed throughout that the universe is finite.[8]

[8] This is no GTS constraint, since in that semantics quantifiers are objectually interpreted and the correlated games may be played on countable as well as on uncountable domains.

5.1 Monadic Quantification

5.1.1 Type $\langle 1 \rangle$ *Quantification* $(\wp(\mathsf{E}) \to \{\mathsf{F}, \mathsf{T}\})$

These types contain expressions such as *no*, *neither... nor* and *all the*.

(G.no) If the game has reached a sentence of the form

$X - no\, Y\ Z - W$,

F chooses an individual, say d, and the game continues with respect to the sentence

X – d – neg(W), if (d is a Y and d Z).

neg(*W*) is a sentential negation operation on the main verb phrase *W*.

As an example of an application of this rule, the sentence 'No musician who plays violin likes to play accordion' is mapped to "John doesn't like to play accordion, if John is a musician, and John plays violin."

The expression *neither... nor* is also a type $\langle 1 \rangle$ quantifier. The game rule (G.neither... nor) is similar to (G.neither), with the sole exception that the quantifier in (G.neither... nor) can apply to more than two objects.

Other than these, type $\langle 1 \rangle$ has a very limited occurrence in natural language. This is shown by the impossibility of defining the quantifier *most* in terms of this type alone, at least without the notion of relativisation (Westerståhl, 1995, pp. 365–386).

5.1.2 Type $\langle 1, 1 \rangle$ *Quantification* $(\wp(\mathsf{E}) \to \langle 1 \rangle)$

These quantifiers comprise the best-studied class of generalised quantifiers. Let us consider only a few examples.

(G.many) If the game has reached a sentence of the form

$X - many\, Y\ Z - W$,

V chooses individuals, say $d_1 \ldots d_n$, and the game continues with respect to the sentence

$X - d_1 \ldots d_n - W$, $d_1 \ldots d_n$ are Y, and $d_1 \ldots d_n$ Z, where

- the number n of individuals is $n > n_0$ for some finite n_0, and
- the relative frequency of $d_1 \ldots d_n$ among the individuals satisfying 'x is a Y' and 'x Z' is larger than some constant c, $0 < c < 1$.

This is by no means the sole possibility of a semantic rule for *many*. One might wish to dispense with the normal frequency constant c and replace it with the proportion $n/|E|$, in which $|E|$ is the size of the domain, or define *many* to be greater than a number defined as a function on the domain $|E|$ (Westerståhl, 1985). Other possibilities also exist, some of them recorded in Lappin (2000). The crux is that (G.many) alone does not do the whole job: the strategies that guide players' actions, not the defining game rules, will determine what counts as *many*.

Correspondingly, the game rule (G.few) is a complement to (G.many), not to (G.most), which goes along the following lines (cf. Hintikka & Kulas 1983):[9]

[9] This is unlike in Mostowski's original study, in which the quantifiers *Most* and *Few* were paralleled.

(G.most) If the game has reached a sentence of the form

$\mathrm{X} - \textit{most}\ \mathrm{Y}\ \ \mathrm{Z} - W,$

V chooses individuals, say $b_1 \ldots b_n$, and the game continues with respect to the sentence

$\mathrm{X} - b_1 \ldots b_n - W$, $b_1 \ldots b_n$ are Y, and $b_1 \ldots b_n$ Z, where

- n is about at least n_0, where n_0 is an approximate finite number of individuals counting as *most*, and
- n_0 is at least half of the total number of individuals satisfying "x is a Y" and "x Z".

For instance, the sentence "Most Chinese ride a bike" translates into "$b_1 \ldots b_n$ ride a bike, and $b_1 \ldots b_n$ are Chinese." It may occasionally be reasonable to require only the latter condition, which is often the case in relational semantics.

Tense-wise, *most* corresponds to the temporal phrase *most of the time*, not to phrases like *almost always*. Various nuances in this basic treatment of *most* are commonplace.

Consider next the quantifier *neither*:

(G.neither) If the game has reached a sentence of the form

$\mathrm{X} - \textit{neither}\ \mathrm{Y}\ \ \mathrm{Z} - W,$

V chooses two individuals, say d and e, whereupon F chooses an individual, say f, and the game continues with respect to the sentence

> neg(X – d W), neg(X – e W), d is a Y, e is a Y, d Z, e Z, and if f is a Y, and X – f Z, then f = d or f = e.

For example, "Neither of the women who saw the exhibition liked the sculpture" translates into "Mary didn't like the sculpture, Joan didn't like the sculpture, Mary is a woman, Joan is a woman, Mary saw the exhibition, Joan saw the exhibition, and if Angela is a woman, Angela saw the exhibition and Angela liked the sculpture, then Angela is Mary or Angela is Joan."

The latter choice by F guarantees that there are no other individuals being Y who Z.

The rule for the hemilogical exceptive (G.every... but) might run as follows.

(G.every... but) If the game has reached a sentence of the form

$\mathrm{X} - \textit{every}\ \mathrm{Y}\ \textit{but}\ \mathrm{Y}'\ \ \mathrm{Z} - W,$

F chooses individuals, say d, $e_1 \ldots e_n$, and the game continues with respect to the sentence

> X – d – W, d is a Y, d Z, and X – neg(e_1 – W), e_1 is a Y′, e_1 Z,..., X – e_n – neg(W), e_n is a Y′, and e_n Z.

According to this rule, the sentence "Every student but the usual ones who signed up arrived on time" receives the interpretation "John arrived on time, John is a student, John signed up, and Mary didn't arrive on time, Mary is the usual one, Mary signed up, Kathy didn't arrive on time, Kathy is the usual one, and Kathy signed up." Mutatis mutandis, one gets a rule (G.no... but). Note that this rule does not satisfy CONS.

The next rule evaluates sentences with *enough*, as in "Enough important members attended the meeting." This is in many respects like *many*, as it does not satisfy EXT, either.

(G.enough) If the game has reached a sentence of the form

$X -$ *enough* $Y\ Z - W$,

V chooses individuals, say $b_1 \ldots b_n$, and the game continues with respect to the sentence

$X - b_1 \ldots b_n - W, b_1 \ldots b_n$ are Y, and $b_1 \ldots b_n$ Z, where

- n is about at least n_0, where n_0 is an approximate finite number of individuals counting as *enough*.

The rule is very similar for *not enough*. Because of their domain-dependence, **(G.enough)** and **(G.not enough)** are examples of rules running the risk of non-logicality in relational semantics.

5.1.3 Type $\langle\langle 1,1\rangle,1\rangle$ *Quantification* $(\wp(E) \times \wp(E) \to \langle 1\rangle)$

The bulk of the quantifiers of type $\langle\langle 1,1\rangle,1\rangle$ deal with comparative statements such as *fewer... than*, *three more... than* and *at least three times as many... as*, which have two NPs applied to one predicate. An example is for *fewer... than*:

(G.fewer... than) If the game has reached a sentence of the form

Fewer $X -$ *than* $Y - W$,

V chooses individuals, say $d_1 \ldots d_n, e_1 \ldots e_k$, whereupon F chooses an individual f, and the game continues with respect to the sentence

$d_1 \ldots d_n$ are X, $e_1 \ldots e_k$ are Y, and $d_1 \ldots d_n, e_1 \ldots e_k$ W, and if f is an X and f W, then $f = d_1$ or $\ldots$ or $f = d_n$, where $n < k$.

F's selection is a certifying move stating that $d_1 \ldots d_n$ are all X who W.

An application of this rule takes "Fewer adults than children enjoy cartoons" into "John and Jill are adults, Tom, Tim and Sandy are children, and John, Jill, Tom, Tim and Sandy enjoy cartoons, and if Bill is an adult and Bill enjoys cartoons, then Bill is John, Bill is Jill,... or Bill is Sandy."

Other rules of this type progress from straightforward modifications to this example.

5.1.4 Type $\langle 1,\langle 1,1\rangle\rangle$ *Quantification*

Let us consider a quantifier *more... than* that has one noun property and two predicate expressions. The subscript is meant to distinguish this expressions from lexically similar expressions of another type.

(G.more... than$_1$) If the game has reached a sentence of the form

More $X - Z$ *than* W,

V chooses individuals, say $d_1 \ldots d_n$, and the game continues with respect to the sentence

$d_1 \ldots d_n$ are X, $d_1 \ldots d_k$ Z, $d_{k+1} \ldots d_n$ W, where $k > n - k$.

(Here $k < n$.) An example is:

More students came early than left late. (1)

That this example illustrates no trivial or unproblematic behaviour of these quantifiers is shown by the sentence:

Students came early more often than left late. (2)

In this sentence, the quantifier *more... than* is applied to the event of coming rather than to individual students, delivering the meaning that perhaps on several occasions, students were early rather than late. In contrast, (1) speaks about one particular occasion with fixed starting and ending dates.

These sentences have posed some problems for the treatment of generalised quantifiers in terms of relational semantics. However, further extensions of GTS may be equipped with quantification over moments of time, intervals, events, states, and similar proper occasions.[10]

5.1.5 Type $\langle\langle 1,1\rangle,\langle 1,1\rangle\rangle$ Quantification

These types consist of two noun phrases applied to two predicates.

(G.more... than$_2$) If the game has reached a sentence of the form

More $X - Z$ *than* $Y - W$,

V chooses individuals, say $d_1 \ldots d_n, e_1 \ldots e_n$, whereupon F chooses an individual f, and the game continues with respect to the sentence

$d_1 \ldots d_k$ are X, $d_1 \ldots d_k$ Z, $e_1 \ldots e_n$ are Y, $e_1 \ldots e_n$ W, and if f is a Y and f W, then $f = e_1$ or ... or $f = e_n$, where $k > n$.

An example of a type $\langle\langle 1,1\rangle,\langle 1,1\rangle\rangle$ quantifier is "More students came early than teachers left late."

5.2 Polyadic Quantification

In standard theories of generalised quantifiers, polyadic quantifiers refer to constructions in which quantifiers may themselves be objects, as in sentences with transitive verbs: "Most critics reviewed just four films" or "At least three girls gave more roses than lilies to John." In this subsection only a small fraction of possible polyadic quantifier schemas are examined, and iterations of monadic quantifiers, for instance, are not be dwelled upon.

5.2.1 Type $\langle\langle 1,1\rangle,2\rangle$ Quantification

An example of type $\langle\langle 1,1\rangle,2\rangle$ quantification is "Different students answered different questions." A game rule runs as follows:

(G.different... different) If the game has reached a sentence of the form

Different $X - Z$ *different* Y,

[10] Pietarinen (2001a) suggests some implementations.

F chooses individuals, say d_1, d_2, e_1, e_2, and the game continues with respect to the sentence

d_1, d_2 are X, e_1, e_2 are Y, and if $d_1 \neq d_2$ then $d_1 - Z - e_1$ and $d_2 - Z - e_2$,
where $e_1 \neq e_2$.

"Different students answered different questions" is now mapped to "John and Mary are students, e_1 and e_2 are questions, and if John and Mary are different then John answered e_1 and Mary answered e_2, where e_1 and e_2 are different." The iteration and the semantic dependency of two applications of *different* aim at establishing a match between students and questions such that the logical force of the sentences equals one in which no two students answered exactly the same set of questions, and that there are at least two students and questions. Therefore, the sentence differs in meaning from, for instance, "More than one/at least two student(s) answered different questions."

(G.every...the same) If the game has reached a sentence of the form

Every X – Z *the same* Y,

F chooses individuals, say d_1, d_2, e_1, e_2, and the game continues with respect to the sentence

d_1, d_2 are X, e_1, e_2 are Y, and if $d_1 \neq d_2$ then $d_1 - Z - e_1$ and $d_2 - Z - e_2$,
where $e_1 = e_2$.

An example is "Every student answered the same questions", which maps to "John and Mary are students, e_1 and e_2 are questions, and if John and Mary are different then John answered e_1 and Mary answered e_2, where e_1 and e_2 are the same." In other words, the logical force here is that no two students answered two different questions.

5.2.2 Type ⟨2, 2⟩ Quantification

The type $\langle 2, 2\rangle$ quantification brings pairs into relation with transitive verbs.

(G.most...are) If the game has reached a sentence of the form

X – *most* Y Z – *are* W,

V chooses (unordered) pairs of individuals, say $(b_1, e_1) \ldots (b_n, e_n)$, and the game continues with respect to the sentence

$X - b_1 \ldots b_n, e_1 \ldots e_n$ are Y, $b_1 \ldots b_n, e_1 \ldots e_n$ Z, and (b_1, e_1) are W,
(b_2, e_2) are W, ..., and (b_n, e_n) are W, where

- n is about at least n_0, where n_0 is an approximate finite number of individuals counting as *most*, and
- n_0 is at least half of the total number of pairs of individuals satisfying "(x, y) are Y" and "(x, y) Z".

For example, the sentence "Most neighbours living in the countryside are friends" has *most* applied to pairs rather than to individuals, making it to have the logical force of, say, "John, Mary, Bill, Sue, Jack and Tim are neighbours living in the countryside, John and Mary are friends, and Bill and Sue are friends." This feature distinguishes *most... are* from its monadic counterpart.

Some other types of polyadic quantifiers are also considered in Section 6.3.

This ends our exposition of semantic game rules for a fragment of non-standard quantifiers in natural language. Let us turn next to a couple of wider implications and amplifications.

6 Context Dependence and Branching

6.1 Complement Quantification

It was assumed throughout that all choices are made relative to the set $I \subseteq E$. We may associate this notion with the record of complete histories $h \in H$ as provided by the theory of extensive-form games. The set I as well as H may also be adjoined by contextually and deictically determined elements from distinct sources. The need for such relativisation arises in a variety of circumstances. An example is "Most students came late. Two of them were kept in." In interpreting the latter sentence in this mini-discourse, the domain of quantification has to be restricted to those individuals delineated by *most*.

This poses some problems in standard theories of quantification. In the theory of semantic games, however, the players' actions may always be relativised to the choice set I or the set of histories H, thus specifying the legitimate subdomain derived from earlier parts of discourse (see also Clark this volume; Sandu 1991, 1993, this volume).

Related proposals that resort to context sets (Westerståhl, 1985) or discourse-representation structures (Kamp & Reyle, 1993) also have some drawbacks: these theories are not equipped to provide an explanation as to why the latter sentence in the following pair is ungrammatical:[11]

Three donkeys were beaten. Two bolt. (3)

*Three donkeys weren't beaten. Two bolt. (4)

GTS explains the unacceptability of (4) in terms of one player overturning the responsibility of finding out the meaning of this mini-discourse to the opposite player, who acts according to a rival strategic purpose. Hence, the choices by the opponent for the quantifier *three* in the antecedent sentence are not included into the choice set I; in other words they are inaccessible to the opponent from the history $h \in H$ that has been traversed in reaching the position in the game in which the values for the quantifier *two* in the consequent sentence are to be chosen.

Furthermore, we customarily interpret I on the left argument:

$$D_E^I(A, B) \text{ if and only if } D_E((I \cap A), B). \qquad (5)$$

This ensures that previous discourse is effective and is taken into consideration when interpreting the noun argument A.

Lacking a comparable notion of histories reached in any play of a semantic game, the relational theory of meaning falls short of dealing with context-dependent complement quantifiers

[11] Assuming that donkeys never bolt unless beaten.

such as *most... the rest* and *three... the others*. This may be illustrated by examples such as the following:

Most students left the building early. The rest stayed inside. (6)

Three participants took a taxi. The others walked to the hotel. (7)

Here the semantic game for the context-dependent complement quantifiers *Most... the rest* and *Three... the others* employs the extensive-game notion of histories that spells out the relevant subdomain for the subsequent expressions *the rest* and *the others* when co-occuring with the quantifiers *most* or *three* to which they are interpretationally linked.

The game rules are applied by facilitating them with complements of choice sets:

$$\mathsf{D}_{\mathsf{E}}^{\mathsf{I}}(\mathsf{A},\mathsf{B}) \text{ if and only if } \mathsf{D}_{\mathsf{E}}(((\mathsf{E}-\mathsf{I})\cap\mathsf{A}),\mathsf{B}). \tag{8}$$

Accordingly, context-dependent quantifiers are not independent determiners operating on their own, but are interpretable only if there is a record of past actions whose primary function is denoting the cases in which certain determiners do not obtain their corresponding values.[12]

Allied to the above determiners are complement quantifiers over locations such as *here... elsewhere* as well as those over times such as *tomorrow... some other time* and related adverbials.

Consider also the following discourse of three clauses:

Dozens of students rallied. Most of them were peaceful. A few smashed windows. (9)

Here the player selects individuals for the determiner *a few* not from the set of past actions as defined after the application of a game rule for *most* in the intermediate sentence, but from the record of actions as constrained by the game rule applied to the numeral phrase *dozens of* in the initial sentence.

One may think that context-dependency might simply be encoded into a model-theoretic forcing relation analogously with assignments of values to free variables. However, it is not sufficient to do just that: such a model-theoretic relation keeps no record of the variability of the relevant domains in discourse in the manner comparable with anything like the contextual notions of the choice set or the records of possible and actual derivational histories in an extensive-form semantic game.

6.2 STRATEGIES

What is particularly important in the game-theoretic interpretation is the notion of a strategy. As mentioned, it is a complete rule telling at every contingency (however probable or improbable) how the respective player should act. In general, strategies may be taken to be (set-valued) functions from non-terminal histories $\mathsf{h} \in \mathsf{Z}$ and sequences of non-terminal histories $\mathsf{h}_1 \ldots \mathsf{h}_k$ to individuals and sets of individuals.

[12] The sole exception to this rule seems to be the quantifier *No* that has no existential import. It prompts a choice by the falsifier, after which the game continues with respect to the negated verb phrase and other required conditions. In such a case, we can naturally have anaphoric pronouns which may be either plural or singular ("No participant took a taxi. They/He/She/Everybody preferred to walk.").

How are the contextual features actually reflected in such strategies? For example, how do we interpret the following sentence?

Most students received many good marks. (10)

The point is that *Many* should appear within the scope of *Most*. We may attempt to represent this in a fashion analogous to how polyadic quantification is expressed:

$$Most_{\mathsf{E}}\,(\mathsf{A}, Many_{\mathsf{E}}(\mathsf{B}, \mathsf{C})). \tag{11}$$

However, the verb phrase is not transitive or ditransitive. Thus (11) is synonymous in meaning with "Most students received marks and many marks are good." No amount of nesting makes the polyadic quantification agree with the intended reading.

The sentence (10) may be thus be symbolised by relativising the domains to the choice set I:

$$Most^{\mathsf{I}}_{\mathsf{E}}\,(\mathsf{A}, \mathsf{B})\; Many^{\mathsf{I}}_{\mathsf{E}}\,(\mathsf{B}, \mathsf{C}). \tag{12}$$

The formula (12) involves two generalised quantifiers *most* and *many*. It might be interpreted so that the denotations for *many* are amongst the individuals falling within the range of *most*. Otherwise, the sentence might be understood as being true even if there are, say, just three good marks among twenty-odd students in a thirty-person class.

What is still needed is a semantic method that accounts for functional dependence between generalised-quantifier phrases. Such method is in the offing in GTS, which uses strategy functions to facilitate the dependence.

Accordingly, in representing (12), two strategies are evoked:

$$\exists \mathsf{F} \exists \mathsf{G}\, Most^{\mathsf{I}}_{\mathsf{E}}\,(\mathsf{A}\,\mathsf{F}(\mathsf{A}))\, Many^{\mathsf{I}}_{\mathsf{E}}\,(\mathsf{F}(\mathsf{A})\;\mathsf{G}(\mathsf{F}(\mathsf{A}))). \tag{13}$$

Here $\mathsf{F}(\mathsf{A}) = \mathsf{B}$ and B is the set of many marks received for A, in which A is the set of most students. $\mathsf{G}(\mathsf{B}) = \mathsf{C}$ and C is a set of good marks of B in which B is the set of many marks received for A. The interpretation this sentence gets is "Most students received marks, and many marks that most students received were good."

By generalising this idea we get a generic definition of what a 'strategic' normal form for generalised quantifiers across different types is, namely one in which arrays of functionals represent players' winning strategies. These arrays may be thought of as solution concepts of the semantic games correlated with discourse.

6.3 Branching Generalised Quantification

Another form of polyadic quantifiers is branching, which is of the general type of $\langle\langle 1^{\mathsf{k}}\rangle, \mathsf{k}\rangle$. An example for first-order quantifiers first given in Hintikka (1973) is

Some relative of every villager and some friend of every townsman hate each other. (14)

The intended reading here ought to accord with the quantifier structure which expresses that (i) every relative of a villager and every friend of a townsman hate each other, and that (ii) the choice of a relative y depends only on the choice of a villager x and the choice of a friend y depends only on the choice of a townsman z:

$$\begin{pmatrix} \forall x & \exists y \\ \forall z & \exists u \end{pmatrix} ((V(x) \land T(z)) \to (R(x,y) \land F(z,u) \land H(y,u) \land H(u,y))). \tag{15}$$

The Skolem normal form of (15) is

$$\begin{aligned} \exists f \exists g \forall x \forall z\, ((V(x) \land T(z)) \to (R(x,f(x)) \land F(z,g(z)) \land \\ \land H(f(x),g(z)) \land H(g(z),f(x)))). \end{aligned} \tag{16}$$

Since we can envisage functional forms for generalised quantification, it is also possible to derive a general semantics for branching quantifiers. Such a definition has nonetheless posed difficulties in the literature. The received definitions rely strongly on the monotonicity properties of noun phrases.[13] Barwise (1979) proposed a definition for monotonically-increasing quantifiers, and Westerståhl (1987) generalised the definition to monotonically-decreasing and some non-monotonic quantifiers.[14]

For example, (17)–(19) are customarily taken to illustrate monotonically-increasing, monotonically-decreasing and non-monotonic branching, respectively:

Most women and most men have all dated each other. (17)

Few woman and few men have all dated each other. (18)

Exactly one woman and exactly one man have all dated each other. (19)

Barwise (1979) proposed the following interpretation for monotonically-increasing quantifiers:

$$\begin{pmatrix} Q_1 x \\ Q_2 y \end{pmatrix} \psi(xy) := \exists X \exists Y (Q_1 x Xx \land Q_2 y Yy \land \forall x \forall y (Xx \land Yy \to \psi(xy))). \tag{20}$$

Reversing the implication in (20) defines monotonically decreasing quantification. However, Barwise (1979) did not consider this to be a genuine example of branching.[15] Given that linear quantifiers may enjoy different monotonicity properties and yet receive uniform semantic treatment, Barwise's definition is quite limited.

[13] Generalised quantifier Q is monotonically increasing if and only if: If $Qx\varphi(x)$ and $\forall x(\varphi(x) \to \psi(x))$, then $Qx\psi(x)$. Likewise, Q is monotonically decreasing if and only if: If $Qx\psi(x)$ and $\forall x(\varphi(x) \to \psi(x))$, then $Qx\varphi(x)$. Q is non-monotonic if and only if it is neither monotonically increasing nor decreasing.

[14] According to Beghelli et al. (1997), the most successful linguistic applications of the properties of monotonicity are the licensing conditions for negative polarity items. Pietarinen (2001b) argues that the semantic behaviour of such items and their contexts of licensing may be explained in terms of the NPI-thesis, which does not turn on monotonicity properties at all. Some doubt is thus cast on the alleged usefulness of monotonicity properties in the semantics of natural-language quantification.

[15] Despite the fact that its meaning may be given by the following definition:

$$\begin{pmatrix} Q_1 x \\ Q_2 y \end{pmatrix} \psi(xy) := \exists X \exists Y (Q_1 x Xx \land Q_2 y Yy \land \forall x \forall y (\psi(xy) \to Xx \land Yy)). \tag{21}$$

In order to obtain a more general definition, Westerståhl (1987) suggested a decomposition of determiners into positive D^+ and negative D^- parts. However, his proposal assumes EXT and CONS, and so we cannot interpret the sentences such as the following:

Many townsmen and many villagers hate each other. (22)

All except youngsters and few relatives like each other. (23)

It may be contested whether coherent interpretations exist for these sentences, but those cannot even be attempted using Westerståhl's definition. At all events, (22) appears to read legibly, though it also has non-branching distributed readings ("Many townsmen hate each other and many villagers hate each other").

Sher (1990, 1997) proposed maximality on the pairs related by D. The idea is that a sentence must come out as true also in the arbitrary enlargements of a model. This allows mixed monotonicity properties within an arbitrary branching sentence. Spaan (1995) has refined the proposal and defines maximality on the cardinality of the sets related by D. Spaan's strategy avoids the parallel of non-monotonic sentences that prima facie may appear contradictory, such as

Exactly one man was seen by exactly one women. (24)

Exactly two men were seen by exactly two women. (25)

In the works of Sher and Spaan we thus appear to have a general definition of branching generalised quantifiers.[16]

Unfortunately, the definition only works for monadic quantifiers. In contrast to the aforementioned proposals, in GTS branching is interpreted as informational loss: strategies do not get perfect input from earlier histories of the extensive game. Such information loss induces equivalence relations on the histories of the game on which strategies are defined. This idea can now be applied for generalised quantifiers to yield a general semantics for branching: if two determiner phrases are independent (e.g., cumulative or branching), no information concerning the application of a rule for one determiner may propagate to the application of a rule for the other, independent determiner. This is accomplished in GTS by taking games to be those of imperfect information.

The properties of functionals are affected accordingly. What thus ensues is a general definition of branching as soon as the strategies that use only partial information of previous discourse

[16] The generalised definition of generalised branching quantifiers Sher proposes is:

$$\begin{pmatrix} Q_1 x \\ Q_2 y \end{pmatrix} \psi(x,y) := \tag{26}$$

$$\exists X \exists Y (Q_1 x Xx \land Q_2 y Yy \land \forall x \forall y (Xx \land Yy \to \psi(x,y)) \land \forall Z \forall W (\forall x \forall y ((Xx \land Yy \to Zx \land Wy) \land (Zx \land Wy \to \psi(x,y))) \to \forall x \forall y (Zx \land Wy \to Xx \land Yy))).$$

Here the conjunct

$$\forall Z \forall W (\forall x \forall y ((Xx \land Yy \to Zx \land Wy) \land (Zx \land Wy \to \psi(x,y))) \to \forall x \forall y (Zx \land Wy \to Xx \land Yy)) \tag{27}$$

of (26) expresses that $\langle X, Y \rangle$ is a maximal pair in a model.

are defined with reference to information sets instead of single actions. The reason why branching has posed difficulties in attempts based on relational theories of generalised quantifiers is that due to substitutional interpretation, quantifiers do not display functional dependencies and each determiner block has to be evaluated in isolation.[17]

The definition of branching in Sher (1997) is also based on reductions in information, albeit quite differently from games of imperfect information. Sher's definition traces the essential dependencies between quantifiers and relates them via sets in different rows of the quantifier structure, which gives rise to maximal pairs with respect to the satisfaction of the predicate part of the formula. The notion of information is not similarly functional as in semantic games, however. Information flows relationally, among quantified variables occurring in the same rows, and it is such relations that delineate independent quantifiers from one another.

7 Conclusions

A couple of concluding remarks and wider perspectives are in order.

(a) Landman (2000) has argued that motivations for monadic quantifiers do not carry over to polyadic ones, as the problem is that we do not know which operations in natural language give rise to polyadic quantifiers. This is, according to Landman, because the existence of such operations need to be settled grammatically, even though polyadic quantifiers are not lexical items. According to Landman, the proper way to derive polyadic generalised quantification is through operations on monadic quantifiers.

In contrast, GTS does not presuppose any separate syntactic operation to form polyadic quantifiers, because they are interpreted directly on linguistic input. Inherent in the theory is the assumption that no hard-and-fast set of natural-language quantifiers exists to be constrained by some operations on lexical items. The distinction between monadic and polyadic is simply to facilitate a comparison between GTS and the standard account. Hence, new operations need not be defined whenever new determiner types are discovered. From the game-theoretic point of view, there is little difference between monadic and polyadic quantifiers. In other words, asking which polyadic quantifiers are expressible in natural language is something that cannot be fully settled by empirical generalisations from data, for instance by closure on suitable operations on lexical quantifiers.

(b) Another consequence is that, since in GTS players choose sequences of individuals instead of sets, the so-called proportion problem is avoided. According to this problem (see e.g. Partee 1984), in sentences such as

Most farmers who own a donkey beat it, (28)

one has to quantify over pairs of a farmer and a donkey. But if so, there will be too many donkeys. However, in playing the semantic game on such sentences, no quantification over pairs, but rather a selection of suitable individuals, is taking place.

A similar phenomenon is illustrated by the following pair of examples at the level of simple discourse:

Most marbles are in the basket. *They are under the sofa. (29)

[17] See Hintikka & Sandu (1994) for a related criticism. A caveat is that generalised quantifiers of higher types are able to capture some sense of 'scope dependence' and therefore the kind of branching that escapes the standard relational semantics. But that would not be a general solution and it still relies on the substitutional interpretation.

Most marbles are not in the basket. They are under the sofa. (30)

In (30), the sequence of elements corresponding to the quantifier *Most* is added to the choice set. The elements of that set are thus available to be the value of the pronoun *They* in the second sentence. Also, the negation in the first sentence does not affect the availability, since it does not contribute to the exchange of players' roles concerning the choice for the quantifier *Most.* Why is the first discourse (29) not similarly interpretable? None of the marbles selected to be the values of the quantifier *Most* are acceptable to satisfy the plural pronoun, and there is moreover no mismatch between the plural and the singular, not even if we assume that those marbles that do not satisfy the noun phrase were a set of more than one element.

What is going on? The answer lies in the incompatibility of the two verb phrases W ("are in the basket") and W' ("are under the sofa"). The elements satisfying the former cannot satisfy the latter and vice versa in (29). Assuming the basket is not placed under the sofa, only a complementary quantifier such as *The remaining ones* in place of the illicit pronoun *They* would render the latter sentence in (29) acceptable and the whole discourse interpretable. This is because the intersection of the set of elements satisfying W and the set of elements satisfying W' would in that case be empty.

(c) The examples get even trickier when we intersperse complement quantifiers with bridging. Suppose we hear the following discourse:

They got married. The other is happy. (31)

Notwithstanding the conversational implicatures which are abundant here, the complement quantifier *The other* does not have any lexical head at all whose value could be included into the choice set. Nor does its complement, whoever the unhappy spouse may be. The legitimate values for the proper interpretation of discourses like this must be provided by contextual and collateral considerations bound to the strategic but rule-governed system of the game.

Examples such as these are good test benches for the range of applicability of our overall theories of quantification. To conlude, then, meanings of a variety of quantifier phrases are characterisable in a unifying game-theoretic fashion. Some such quantifiers belong to linguistic categories other than those classified according to standard accounts of generalised quantification based on relational semantics. Conversely, many quantificational expressions exist that have little to do with generalised quantifier theory per se, but which nonetheless fall within the purview of GTS.[18] So GTS is as well off as relational semantics.[19] What is more, GTS provides a general and dynamic framework for dealing with complex and context-dependent quantifier phrases. With reference to branching it also proves its worth.

[18] See Pietarinen (2001a) for some of the examples and their treatment in GTS.

[19] We have focussed on natural language and hence a mathematical treatment of the relationship needs to be studies separately. Taking a binary form of the weak logic $\mathsf{L}(Q)$ of Keisler (1970), where a first-order model M expands to a model (M, r) for some binary relation r on subsets of the domain of M, a match between GTS and a weak logic would amount to the characterisation that, assuming the axiom of choice, for any $\mathsf{L}(Q)$-sentence φ and $\mathsf{L}(Q)$-model (M, r), $(\mathsf{M}, \mathsf{r}) \models_{\text{weak logic}} \varphi$ iff $(\mathsf{M}, \mathsf{r}) \models_{\text{GTS}} \varphi$. A weakened relationship obtains as soon as r varies when passing from weak logic to GTS according to some relation R: assuming the axiom of choice, for any $\mathsf{L}(Q)$-sentence φ and $\mathsf{L}(Q)$-model (M, r) such that R, $(\mathsf{M}, \mathsf{r}) \models_{\text{weak logic}} \varphi$ iff $(\mathsf{M}, \mathsf{r}') \models_{\text{GTS}} \varphi$. The more general the restrictions on R, the more precise will the match-up be.

ACKNOWLEDGMENTS

Supported by the Osk. Huttunen Foundation, the Jenny and Antti Wihuri Foundation, the Osk. Öflund Foundation, and the Academy of Finland (1103130, *Logic and Game Theory*). A preliminary version was presented at the *Sinn und Bedeutung IV* conference in Amsterdam in December, 1999.

REFERENCES

Barwise, J. (1979). On branching quantifiers in English. *Journal of Philosophical Logic*, **8**, 47-80.

Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, **4**, 159-219.

Beghelli, F., D. Ben-Shalom and A. Szabolcsi (1997). Variation, distributivity, and the illusion of branching. In: *Ways of Scope Taking* (A. Szabolcsi, ed.), pp. 29-69. Kluwer, Dordrecht.

van Eijck, J. (1996). Quantifiers and partiality. In: *Quantifiers, Logic and Language* (J. van der Does and J. van Eijck, eds.), pp. 105-144. CSLI Publications, Stanford.

Higginbotham, J. and R. May (1981). Questions, quantifiers and crossing. *The Linguistic Review*, **1**, 41-80.

Hilpinen, R. (1982). On C. S. Peirce's theory of the proposition: Peirce as a precursor of game-theoretical semantics. *The Monist*, **62**, 182-189.

Hintikka, J. (1973). Quantifiers vs. quantification theory. *Dialectica*, **27**, 329-358. (Reprinted in *Linguistic Inquiry*, **5**, 1974, 153-177.)

Hintikka, J. and J. Kulas (1983). *The Game of Language: Studies in Game-Theoretical Semantics and its Applications*. D. Reidel, Dordrecht.

Hintikka, J. and J. Kulas (1985). *Anaphora and Definite Descriptions*. D. Reidel, Dordrecht.

Hintikka, J. and G. Sandu (1991). *On the Methodology of Linguistics*. Blackwell, Oxford.

Hintikka, J. and G. Sandu (1994). What is a quantifier? *Synthese*, **98**, 113-129.

Hintikka, J. and G. Sandu (1997). Game-theoretical semantics. In: *Handbook of Logic and Language* (J. van Benthem and A. ter Meulen, eds.), pp. 361-410. Elsevier, Amsterdam.

Janasik, T., A.-V. Pietarinen and G. Sandu (2003). Anaphora and extensive games. In: *Chicago Linguistic Society* **38**: *The Main Session* (M. Andronis et al., eds.), pp. 285-295. Chicago Linguistic Society, Chicago.

Kalish, D. and R. Montague (1964). *Logic: Techniques of Formal Reasoning*. Harcourt, New York.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic*. Kluwer, Dordrecht.

Keisler, H. J. (1970). Logic with the quantifier 'there exists uncountably many'. *Annals of Mathematical Logic*, **1**, 1-93.

Landman, F. (2000). *Events and Plurality*. Kluwer, Dordrecht.

Lappin, S. (2000). An intensional parametric semantics for vague quantifiers. *Linguistics and Philosophy*, **23**, 599-620.

Lindström, P. (1966). First order predicate logic with generalized quantifiers. *Theoria*, **32**, 186-195.

Montague, R. (1969). English as a formal language. In: *Formal Philosophy* (R. Thomason, ed.), pp. 188-221. Yale University Press, New Haven.

Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae*, **44**, 12-36.

Partee, B. H. (1984). Nominal and temporal anaphora. *Linguistics and Philosophy*, **7**, 243-286.

Peirce, C. S. (1931-8). *Collected Papers of Charles S. Peirce* (C. Hartshorne and P. Weiss, eds., Vols. 1-6). Harvard University Press, Cambridge, Mass.

Peirce, C. S. (1976). *The New Elements of Mathematics* (C. Eisele, ed., Vol. 4). Mouton, The Hague.

Peirce, C. S. ed. (1983/1883). *Studies in Logic, By Members of the Johns Hopkins University (1883)*. John Benjamins, Amsterdam.

Peirce, C. S. (1998). *The Essential Peirce* (Peirce Edition Project, Vol. 2). Indiana University Press, Bloomington.

Pietarinen, A.-V. (2001a). Most even budged yet: Some cases for game-theoretic semantics in natural language. *Theoretical Linguistics*, **27**, 20-54.

Pietarinen, A.-V. (2001b). What is a negative polarity item? *Linguistic Analysis*, **31**, 165-200.

Pietarinen, A.-V. (2003a). Peirce's game-theoretic ideas in logic. *Semiotica*, **144**, 33-47.

Pietarinen, A.-V. (2003b). Games as formal tools versus games as explanations in logic and science. *Foundations of Science*, **8**, 317-364.

Pietarinen, A.-V. (2004a). Logic, language games and ludics. *Acta Analytica*, **18**, 89-123.

Pietarinen, A.-V. (2004b). Semantic games in logic and epistemology. In: *Logic, Epistemology and the Unity of Science* (S. Rahman, J. Symons, D. Gabbay and J. P. Van Bendegem, eds.), pp. 57-103. Springer, Dordrecht.

Pietarinen, A.-V. (2005). *Signs of Logic: Peircean Themes on the Philosophy of Language, Games, and Communication*. Springer, Dordrecht.

Pietarinen, A.-V. and L. Snellman (2006). On Peirce's late proof of pragmaticism. In: *Truth and Games* (T. Aho and A.-V. Pietarinen, eds.), pp. 275-283. Acta Philosophica Fennica **78**, Societas Philosophica Fennica, Helsinki.

Saarinen, E. (ed.) (1979). *Game-Theoretical Semantics*. D. Reidel, Dordrecht.

Sandu, G. (1991). Choice set quantifiers. In: *Acta Philosophica Fennica*, **49** (L. Haaparanta et al., eds.), pp. 252-264. Societas Philosophica Fennica, Helsinki.

Sandu, G. (1993). On the logic of informational independence and its applications. *Journal of Philosophical Logic*, **22**, 29-60.

Sandu, G. (2000). Partially interpreted generalized quantifiers. In: *Philosophy and Logic: In Search of the Polish Tradition* (J. Hintikka, T. Czarnecki, K. Kijana-Placek, T. Placek and A. Rojszczak, eds.), pp. 93-108. Kluwer, Dordrecht.

Sandu, G. and A. Pietarinen (2001). Partiality and games: Propositional logic. *Logic Journal of the IGPL*, **9**, 107-127.

Sandu, G. and T. Janasik (2003). Dynamic game semantics. In: *Meaning: The Dynamic Turn* (J. Peregrin, ed.), pp. 215-240. Elsevier, Oxford.

Sher, G. (1990). Ways of branching quantifiers. *Linguistics and Philosophy*, **13**, 393-422.

Sher, G. (1997). Partially-ordered (branching) generalized quantifiers: A general definition. *Journal of Philosophical Logic*, **26**, 1-43.

Simons, P. (1994). Leśniewski and generalized quantifiers. *European Journal of Philosophy*, **2**, 65-84.

Spaan, M. (1995). Parallel quantification. In: *Quantifiers, Logic and Language* (J. van der Does and J. van Eijck, eds.), pp. 281-309. CSLI Publications, Stanford.

Westerståhl, D. (1985). Logical constants in quantifier languages. *Linguistics and Philosophy*, **8**, 387-413.

Westerståhl, D. (1987). Branching generalized quantifiers and natural language. In: *Generalized Quantifiers. Linguistic and Logical Approaches* (P. Gärdenfors, ed.), pp. 269-298. D. Reidel, Dordrecht.

Westerståhl, D. (1995). Quantifiers in natural language. In: *Quantifiers: Logics, Models and Computation* (M. Krynicki et al., eds.), pp. 359-408. Kluwer, Dordrecht.

Woods, W. (1968). Procedural semantics for a question-answering machine. *AFIPS Conference Proceedings*, **3**, 457-471.

Zeman, J. J. (1986). Peirce's philosophy of logic. *Transactions of the Charles S. Peirce Society*, **22**, 1-22.

APPENDIX: PEIRCE ON GENERALISED QUANTIFICATION

The prehistory of generalised quantifiers has not been much studied (but see Simons 1994). It is well-known that Gottlob Frege (1848–1925) considered quantifiers as variable-binding operators denoting second-order relations. But Charles Peirce (1839–1914) noticed the need for having generalised notions of quantifiers alongside the quantifiers that bind individual objects.

Together with the Johns Hopkins mathematician Oscar H. Mitchell (1851–1889), Peirce invented quantifiers of first-order logic in the early 1880s (Peirce, 1983/1883). In the Peirce-Mitchell first-order logic, the sign Σ, denoting the relative sum $\Sigma_i P_i$ of terms in the algebraic sense, corresponds to the substitutionally interpreted existential quantifier, and the sign Π, denoting the relative product $\Pi_i P_i$ of terms, corresponds to the substitutionally interpreted universal quantifier. One of the first instances of the term 'Quantifier' occurs in his 1885 paper "On the Algebra of Logic: A Contribution to the Philosophy of Notation" (CP 3.396; W5: 162–190).[20]

As to the various ways of generalising the idea, Peirce wrote in 1893:

> Two varieties of [selective pronouns] are particularly important in logic, the *universal selectives,*... such as *any, every, all, no, none, whatever, whoever, everybody, anybody, nobody.* These mean that the hearer is at liberty to select any instance he likes within limits expressed or understood, and the assertion is intended to apply to that one. The other logically important variety consists of the *particular selectives,... some, something, somebody, a, a certain, some or other, a suitable, one.*
>
> Allied to the above pronouns are such expressions as *all but one, one or two, a few, nearly all, every other one,* etc. Along with pronouns are to be classed adverbs of place and time, etc.
>
> Not very unlike these are, *the first, the last, the seventh, two-thirds of, thousands of,* etc. (CP 2.289, 1893, *Speculative Grammar: The Icon, Index, and Symbol*)

A couple of years later, he stated further:

> A subject should be so described as to be neither Universal nor Particular; as in *exceptives (Summulae)* as "Every man but one is a sinner." The same may be said of all kinds of numerical propositions, as "Any insect has an even number of legs." But these may be regarded as Particular Collective Subjects. An example of a Universal Collective subject would be "Any two persons shut up together will quarrel." A collection is logically an individual. (CP 2.324, c.1902–03, *Speculative Grammar: Propositions*)

Peirce did not interpret these quantificational expressions in the relational way as quantifying over sets and then expressing relations that would hold between objects and predicates. Instead of sets (the use of which he was inclined to repudiate in logic altogether), he referred to collections the theory of which he struggled to develop over an extended period of time. By what he termed the method of "hypostatic abstraction", collections are logically turned into individual objects, since their existence was taken to depend upon the existence of certain concrete individuals.[21] According to Peirce, collections are the precise counterpart to the notion of a set which is vague and imprecise.

[20] Originally appeared in *American Journal of Mathematics* 7, 1885, 180–202. The reference CP is to Peirce (1931–8) by volume and paragraph number, and W is to *Writings of Charles S. Peirce: A Chronological Edition*, The Peirce Edition Project, Bloomington: Indiana University Press, by volume and page numbers.

[21] See MS 690, 1901. Let me quote some textual sources to this effect: "The Object of every sign is an Individual, usually an Individual Collection of Individuals" (CP 8.181); "Collections are not *grades* of any kind, but are single things" (CP 4.663), cf. CP 4.179, 4.345, 4.370, 4.532, 4.649, 4.655; MS 690. See Zeman (1986) for a study of Peirce's notion of hypostatic abstraction.

The reason why Peirce did not approve sets as values for generalised quantifiers was that by mid-1890s, he realised the limitations of substitutional interpretation of quantification in which quantifiers were interpreted as indices and which he had earlier assumed for his algebra of logic. Since our universes of discourse may be uncountable, the substitutional interpretation is bound to fail.

Instead of the substitutional interpretation, Peirce envisioned a game-theoretic interpretation, in which quantifiers are symbols which are interpreted objectually through habits.[22] Given such an interpretation, the Utterer and the Interpreter of the given assertion pick logically individual collections as intended by the unsaturated predicate terms (rhemas) of a given formula or an assertion. Peirce believed that through such habitual and strategic actions, the ability to grasp plural expressions in natural language follows.

According to such a game-theoretic interpretation, "*selective* pronouns [quantifiers] ...inform the hearer how he is to pick out one of the objects intended" (CP 2.289). This interpretation may be effectuated by resorting to the terminology of game theory, something which in the literal sense was not yet available during Peirce's lifetime. The choices of the players are made such that, "In the sentence 'Every man dies,' 'Every man' implies that the interpreter is at liberty to pick out a man and consider the proposition as applying to him" (CP 5.542, c.1902, *Reason's Rules*). Likewise, in the sentence containing particular selectives, the Utterer will act. The role of the winning strategy that is central in GTS was played in Peirce's semantic theory by the notion of a habit that has "definite general tendencies of a tolerably stable nature".[23]

Among Peirce's observations was also that, "When there are several quantified subjects, and when quantifications are different, the order in which they are chosen is material" (CP 2.338, c.1895). Here Peirce was speaking of the order and the priority of the symbols Σ and Π in a first-order formula. He thus recognised the importance of the dependencies that these quantifiers give rise to when arranged in a linear order. But the recognition of the importance of the quantifier order is an altogether general observation and vindicates the finding that he did not take quantifiers to function substitutionally, because in that case the behaviour of quantifier strings would be reduced to algebraic substitutional equations, which means that such equations are already given in a certain distinct form, that is, in a form that reflects various quantifier dependencies. This is yet another blow to the functionality of substitutional interpretation. Since such dependencies are essential in the theory of generalised quantifiers as well as in first-order logic, the substitutional interpretation is bound to fail in both.

Further evidence for the kinship of Peirce and the game-theoretic interpretation for natural language is found in the parallel of the following two formulations:

> "Any man will die," allows the interpreter... to take any individual of that universe as the Object of the proposition, giving, in the above example, the equivalent "If you take any individual you please of the universe of existent things, and if that individual is a man, it will die." (Peirce, 1998, p. 408)

This is very similar, both in spirit and letter, to the interpretation GTS assigns to sentences containing the universal *any* (the specific terminology is explained in Section 2):

(G.any) If the game has reached the sentence

$X - \textit{any}\ Y$ who $Z - W$,

then Nature may choose an individual and give it a proper name (if it did not have one already), say 'b'. The game is continued with respect to

$X - \mathrm{b} - W$, b is a(n) Y, and (if) b Z. (Hintikka in Saarinen 1979)

[22] See Pietarinen (2005) as well as chapter "The semantics/pragmatics distinction from the game-theoretic point of view" in this volume.

[23] MS 280: 30, c.1905, *The Basis of Pragmatism*; see Pietarinen (2005) and Pietarinen & Snellman (2006) for further discussion on strategic and methodological outlook that Peirce had in his logical and philosophical investigations.

Here Nature (i.e., the Falsifier) corresponds to Peirce's Interpreter and Myself (i.e., the Verifier) to the Utterer.

On many occasions Peirce refers to "hemilogical quantifiers" (Peirce, 1983/1883, p. 203) in addition to universal and existential ones. They were taken to mean phrases such as *all but one, all but two* and so on. For example, the algebraic quantifiers $\Pi', \Pi'' \ldots$ were taken to mean products of all individuals except one, except two, and so on. Peirce even attempted to characterise sentences containing generalised quantifier phrases such as "there are at least three things in the universe that are lovers of themselves" using such hemilogical quantifiers (Peirce, 1983/1883, p. 203).

Moreover, that some quantifiers are vague and some are general is no impediment to them being logical:

> Logicians confine themselves, apart monstrative indices themselves, to 'Anything' and 'Something' two descriptions of what monstrative index may replace the subject, the one description *vague* the other *general.* No others are required since such subjects, "All but one", "All but two", "Almost all", "Two thirds of the occasions that present themselves in experience", and the like are capable of logical analysis. (MS 288, 1905, *Materials for Monist Article: The Consequences of Pragmaticism*)

Worth noting is also that Peirce recognised not only the important linguistic fact that some languages have double concord while others do not but also the properties of downwards and upwards entailment that accompanies certain quantifiers: "There are but few languages in which two negatives make an affirmative. If *not* means "less than one" or "fewer than one" fewer than fewer than one is simply fewer than one. The new signs I propose make *some some,* all" (MS L 237, 12 November 1900, *Peirce to Christine Ladd-Franklin*).

Although in the version of GTS given in the body of this chapter the players choose sequences of individuals, these may as well be interpreted as collections, whereof by abstraction they become logically tantamount to an individual. The sundry addition in the game rules transforms the main verb into the singular. This step, sometimes termed 'collectivisation' in the literature (a passage from a set of individuals to an individual set, or from generalised quantifiers to plurals), will be significant as soon as the players' strategic decisions are at issue, because they are partial functions defined on (some of) the previous choices in a game with a unique output. The kind of collectivisation that takes place between game rules and strategies thus exhibits a version of hypostatic abstraction. It is an embodiment of Peirce's remark that certain "abstractions are individual collections" (CP 2.357, 1901, *Subject*).

The cumulative weight of these remarks is unmistakable. Had Peirce continued his development of generalised quantifiers, we would have witnessed a development of a game-theoretic interpretation for generalised quantifiers on a par with the game-theoretic interpretation for the existential and universal ones long before such generalisations were actually discovered and their theory systematised. Consequently, his suggestions are not only anticipations left for historians of logic to explore. There is ample room for research on the largely unexplored terrain of the directions into which the study of the logic of collective subjects might be advanced.

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 14

GAMES, QUANTIFIERS AND PRONOUNS

Robin Clark
University of Pennsylvania

In this paper, I will outline a game-based approach to reference tracking. Reference tracking is the ability to successfully assign referents to discourse anaphors. My central claim is that reference tracking is an example of how linguistic agents can strategically manage a resource; as such, it is amenable to a game-theoretic analysis. The technique I will develop relies on the management of a data structure, which I will call a *game board*; since all participants of the discourse are aware of how the game board is managed, speakers can strategically use this resource during the course of a conversation.

We turn, in Section 1, to a brief presentation of some data about how quantifiers introduce discourse entities and how these entities can be accessed by discourse anaphors. I do not intend to cover all of the possibilities here, but to treat a few interesting basic cases. In Section 2, we will turn to a brief discussion of Game-Theoretic Semantics (GTS) and a few rules for interpreting a small selection of quantifiers. The quantifier rules of GTS allow for a straightforward introduction of discourse entities. I will not specifically address the problem of scope ambiguities here, fixing my attention instead on the elementary case of how a single quantified expression establishes a discourse entity. I will briefly address the problem of scope in the conclusion of the paper.

In Section 3 we turn to a game-based discussion of discourse anaphora. The basic idea is that once discourse entities have been introduced, they can be treated as a resource available as public knowledge to the participants of the discourse. The participants can then treat the problem of associating referents with discourse anaphors as a game that can be solved rationally. I will argue that the referents for discourse anaphors can be found by solving for the Pareto-Nash Equilibrium of the game. The idea is that both the speaker and the hearer are involved in a strategic interaction and that the basic structure of the problem is a matter of public knowledge. Because of this mutual knowledge, the participants in the conversation are able to formulate coherent strategies dealing with reference tracking, the ability to correctly assign discourse referents to pronouns. In short, in Section 2 we approach the problem of establishing discourse entities using quantifiers and in Section 3 we solve the problem of choosing ways to refer to them.

1 Overview

My interest here will be twofold. There have been extensive discussions in the literature about how names, singular indefinites and some definite noun phrases introduce new discourse entities. Concrete proposals have been made about how these discourse entities are managing over the course of a conversation, particularly in the literature on Dynamic Semantics, Discourse Representation Theory (DRT) and Centering Theory.[1] Furthermore, while Dynamic Semantics and DRT have had a great deal to say about how some noun phrases introduce discourse entities (and others do not), they have had less to say about how these resources are managed. Centering Theory has had a great deal to say about how resources are managed, particularly with respect to topic-hood, but it has not been particularly concerned with how quantified noun phrases introduce these resources. I would like to consider here, first, how a broader range of expressions introduce discourse entities and, second, how these entities are then managed in the course of a conversation.

I will consider relatively simple texts like those exemplified in (1):

(1) a. No dean reads Proust. They prefer Stephen King.
b. At least 5 deans dropped acid. One jumped out the window.
c. At most 5 faculty members considered resorting to cannibalism. They changed their minds when they realized how much work it would be to hunt undergraduates.
d. Most deans are druids. They march about waving mistletoe.
e. More deans than faculty eat three square meals a day. They need to keep up their blood sugar.
(*They* being the deans)
f. More deans than faculty eat three squares a day. They want to keep their weight down.
(*They* being the faculty)

In each of the above cases, a quantifier introduces a discourse entity—for the moment, we will make no commitments as to the character of this entity—which is then the target of a pronoun in the next sentence. We should compare the small texts in (1) which involve inter-sentential anaphora with the example in (2) which involves anaphora within a single sentence:

(2) The doctor told John his pants were on fire.

Assuming that the doctor in (2) is male, then the sentence is perfectly ambiguous given no further information about the context; the pronoun *his* can refer either to John or the doctor. The pronouns in (1) behave quite differently from the one in (2). All else being equal, the pronouns in the second sentences of the texts in (1) are unambiguously dependent on an element in the preceding sentence.

Consider, first, the small text in (1)c. The pronoun *they* in the second sentence must refer to those faculty members, five or fewer in number, who considered resorting cannibalism. A similar judgment holds for (1)d; *they* must refer to those deans who are druids. The pronoun *they* in (1)a must refer to deans—note, though, that this pronoun has no plural antecedent in the preceding sentence. The indefinite pronoun *one* in (1)b must refer to an individual selected

[1] For example, see Beaver (2001), Groenendijk & Stokhof (1991), Musken, van Benthem & Visser (1997) or van den Berg (1996) on Dynamic Semantics. On DRT, see Kamp & Reyle (1993) or van Eijk & Kamp (1997). On Centering Theory, see Grosz, Joshi & Weinstein (1983, 1986), Joshi & Weinstein (1981) or Walker & Prince (1996), among others.

from among the deans who dropped acid. Finally, the examples in (1)e and (1)f involve multiply headed quantifiers.[2] My aim, in this paper, is to give an account of facts like those in (1). How do quantifiers create discourse entities and how do pronouns refer back to them in the course of a text?

In Section 2, I will introduce some elementary rules for the interpretation for a few quantifiers using GTS.[3] These rules take the form of simple games between a verifier and falsifier, and sometimes involve a pair of moves by each. These rules are not intended to provide an exhaustive theory of natural-language quantification in the GTS framework, but, rather, to provide a basis for studying how quantifiers establish entities which are, then, available to the participants of the discourse as a managed resource. One principle in stating these rules is that they must correctly account both for the truth-conditional contribution of the quantifier and for its discourse contribution. As a result, I cannot yet give an algebraic treatment of the rules in the manner of Keenan & Stavi (1986) since I have not to date discovered an algebraic method for algebraically constructing the proper discourse entity from the truth-conditional component of the game rule. The discussion here is mainly intended to provide a basis for later discussion.

In Section 3, we turn to the management problem of the discourse entities introduced by names and quantifiers. The analytical framework I adopt will be rather different than the GTS framework in Section 2, although I will stay within a game setting. In particular, I will not conceive of the game as being a zero-sum contest between a verifier and a falsifier but a cooperative game between a speaker and an audience. This approach is much closer in spirit to the framework of Parikh (2001, 2006), which is grounded in classical game theory; we differ largely in my use of zero-sum games in the form of GTS-style rules as opposed to Parikh's use of cooperative games.[4] We suppose that the speaker and the audience are at odds insofar as the speaker prefers to use the shortest expression possible to refer to an entity while the audience prefers utter clarity and, all else being equal, wants the speaker to be unambiguous. Both the speaker and the audience prefer for the intended message to be transmitted. As a result of this shared preference, the speaker and audience must cooperate, even though their preferences might diverge on some points.

Speakers and hearers have a mutual interest in fixing the reference of phrases. Assuming that it is in the speaker's interest to avoid prolixity in favor of conciseness, she will tend to use pronouns. Hearers would prefer for the speaker to be as precise as possible so that the reference can be fixed with a minimum of effort and ambiguity. Hence, hearers would prefer that the speaker avoid pronouns unless their reference can be easily fixed. Thus, it is in the interests of the speaker and the hearers that there be a set of publicly known strategies regarding the use of pronouns.

In Section 4, I turn to some of the consequences of the above approach. In particular, I will argue that much of the truth-conditional semantics of natural language can be modeled by zero-sum games between a verifier and a falsifier. Pragmatics, on the other hand, involves games of cooperation between a speaker and an audience. The underlying formalism for both truth conditional semantics and pragmatics is otherwise identical, once the difference between zero-sum semantical games and cooperative pragmatic games is recognized.

[2] My judgment and the judgment of those I have asked is that (1)e is more natural than (1)f. We return to this below.

[3] See Hintikka & Kulas (1985), Hintikka (1996) and Hintikka & Sandu (1997). On quantifiers in GTS more specifically, see Clark (2004) and Pietarinen (2006).

[4] Myerson (1991) gives a clear exposition of classical game theory and some of its extensions.

2 Basic Quantifiers

In this section, I will give some game rules for a few quantifiers. Before doing so, however, we should consider, in broad outline, the structure of GTS. It seeks to give the truth conditions of a sentence, S, by means of a zero-sum game played between two players, a verifier and a falsifier, with respect to a model, $\mathcal{M}$. This game will be designated by:

$$\mathsf{G}(\mathsf{S}, \mathcal{M}).$$

The verifier bets that the sentence S hold in $\mathcal{M}$ while the falsifier bets the opposite.

The game proceeds roughly as follows. A logically active element—that is, an element associated with a game rule—is selected. The players play according to the rule and the original sentence is replaced by one or more sentences that do not contain the logically active element. Eventually, there are only simple sentences left containing only names and simple predicates that can be shown to be supported (or not) by the model. At this point, a winner is determined. If a simple sentence is supported by the model, then the verifier wins; otherwise, if the sentence is falsified by the model, then the falsifier wins. Notice that it is possible that the outcome is a draw, in which case the sentence has a third, indeterminate, truth value.

We should note that, as the game proceeds, the verifier and the falsifier may change roles. This can happen under negation, for example; the verifier will win on a negated sentence if and only if the falsifier wins on the unnegated form. The following rule will suffice:

> **(R.negation)**
> If the game $\mathsf{G}(\mathsf{S}; \mathcal{M})$ has reached an expression $\mathsf{P}^{(-)}$ which is the negation in English of P, then the players exchange roles, i.e. the verifier will become the falsifier and vice versa. The game goes on as $\mathsf{G}(\mathsf{P}; \mathcal{M})$.

As we will see below, it will be important to keep track of who the initial verifier and the initial falsifier are. I will refer to the initial verifier as "Eloïse" and the initial falsifier as "Abelard." It is important to note that there will be times when Eloïse is playing the role of falsifier and Abelard that of verifier.

We can explicate *truth* in terms of *winning strategy*. That is, the truth of sentence S in $\mathcal{M}$ can be defined as follows:

$\mathcal{M} \models_{\text{GTS}} \mathsf{S}^{+}$ if and only if there exists a winning strategy for the verifier in $\mathsf{G}(\mathsf{S}; \mathcal{M})$.

Falsity is defined as the dual of truth:

$\mathcal{M} \models_{\text{GTS}} \mathsf{S}^{-}$ if and only if there exists a winning strategy for the falsifier in $\mathsf{G}(\mathsf{S}; \mathcal{M})$.

Supposing that the verifier and the falsifier are playing a game of perfect information—that is, that they can see each other's moves, as in chess—the above rules are equivalent to the standard Tarskian treatment (see Hintikka & Sandu 1997, and the references cited there):

Theorem 1. Assuming the axiom of choice, for any first order sentence S and model $\mathcal{M}$, Tarski-style truth and GTS truth coincide; that is:

$$\mathcal{M} \models_{\text{Tarski}} \mathsf{S} \text{ iff } \mathcal{M} \models_{\text{GTS}} \mathsf{S}^{+}.$$

Before we turn to some game rules for English, let us consider some properties of scope in natural languages. Unlike artificial languages like first-order logic, the priority of logical operators in natural languages is not overtly marked by parentheses or other syntactic devices. Instead, ordering principles must be used to determine the order of play in a natural language game. In addition, the scope of quantifiers in natural languages can be broken down into two parts which normally coincide in artificial languages. One part consists of the way in which the quantifier interacts with other logically active elements in the sentence. These are normally thought of as scopal ambiguities, but, since they involve the order of play between the verifier and the falsifier, we will refer to this kind of scope as *priority scope.* In GTS, priority scope is modeled by the order in which logically active elements are selected. I will have little to say about such scope in this paper.[5]

The other type of scope involves the interaction of the quantifier with anaphors, specifically bound pronouns and reflexive items. As has been frequently noted in the literature,[6] these two scopes need not coincide. This is amply illustrated by examples like:

(3) A man entered the bar. He had a penguin on his head.

In (3), *a man* and *he* may refer to the same element. In this case, the binding scope of *a man* exceeds its priority scope. Not all quantifiers are capable of this, as illustrated by the relative peculiarity of the discourse in (4):

(4) Every poet has low self-esteem. She thinks it makes her interesting.

In (4), the quantifier *every poet* cannot bind the pronoun *she.* Thus, the binding scope of "every + noun" is narrower than the binding scope of "a + noun." To account for these properties, we will use a special store, the choice set I_S. The players will add and withdraw elements to and from this set as the game proceeds; that is, the contents of the game set can change dynamically, allowing for a variety of scoping properties; I will develop principles governing the management of the choice set below. Sentences in a discourse are sub-games in a larger "super-game;" as hinted above, while sentences are zero-sum competitive games, the super game is a cooperative game. I will develop this point in Section 3.

In order to establish the properties of the choice set, consider the simple cases in examples (5), (6) and (7):

(5) a. Every student passed the exam. She studied very hard.
b. Every student passed the exam. They studied very hard.
c. Every student thinks he's treated badly.

(6) a. No dean drank the Pernod. He prefers Cointreau.
b. No dean drank the Pernod. They prefer Cointreau.
c. No dean admits he drinks Pernod.

(7) a. A trustee danced nude on the table. He had been snorting cocaine.
b. A trustee danced nude on the table. They had been playing "Truth or Dare."
c. A trustee always claims that he has important business to do.

[5] See the papers by Gabriel Sandu and Tatjana Scheffler in this volume for discussion on these issues.

[6] See, for example, Hintikka & Kulas (1985), Kamp & Reyle (1993), and Groenendijk & Stokhof (1991), among a host of others.

Consider, first, the simple texts in (5) and (6). The (a)-examples show that neither *every student* nor *no dean* introduces a discourse entity that survives in the super game, although both can bind a pronoun inside the same sentence, as witnessed by the (c)-examples. The (b)-examples show that sets evoked by *every student* and *no dean* are available in the super game. Thus, the pronoun *they* in (5)b is understood as the set of students. Likewise, the pronoun (6)b is understood as the set of deans. Notice that in both cases the plural pronoun is dependent on a morphologically singular noun phrase.

The texts in (7) illustrate what happens when a quantifier introduces a discourse entity that survives in the super-game. In this case, a singular pronoun in the next sentence can denote that entity, witness the example (a); a plural pronoun in the next sentence can denote a set evoked by the entity as shown in the example (b) where *they* denotes the set of trustees. Finally, the example (c) show that the quantifier can bind a quantifier within the sentence, just as was the case in (5) and (6).

How are we to account for the examples in (5), (6) and (7)? At the very least, these examples show, first, that not all quantifiers have the same status when it comes to introducing full-fledged discourse entities and, second, that while all entities introduced by quantifiers are available in the game in which they are introduced, not all survive to the super game; some entities disappear once their game is over, although sets that these entities evoked may remain available.

I will suppose that the choice set is divided into two partitions: one part, call it $I_{current}$, contains entities evoked in the current game; the other part, call it $I_{discourse}$, contains all the sets and entities that have been played in all the sub-games that make up the discourse up to the current sub-game. At the end of a sub-game, the contents of $I_{current}$ are placed in $I_{discourse}$ with the following proviso:

(8) *Choice Preservation*
An entity is passed from $I_{current}$ to $I_{discourse}$ just in case it was selected by Eloïse. Otherwise, the set X from which the entity was selected is placed in $I_{discourse}$.

There is crucial use of Eloïse in the formulation of (8). It is only Eloïse, the initial verifier, who has the privilege of passing individual choice from $I_{current}$ to $I_{discourse}$.

Let us now turn to some quantifier rules and illustrate the effects of our data structures and Choice Preservation. Consider first a rule for *some* (as well as *a(n)*):

(R.some)
If the game $G(S;\mathcal{M})$ has reached an expression of the form:

$$Z - \text{some } X \text{ who } Y - W$$

Then the verifier may choose an individual from the appropriate domain, say b. The game is then continued as $G(Z - b - W, b \text{ is an } X \text{ and } bY; M)$. The individual b is added to the choice set I_S.

Suppose, to take a concrete example, that the game is being played on the sentence:

(9) A trustee drank.

As always, Eloïse bets the sentence is true and Abelard bets that it is false. Suppose that $\mathcal{M}$ contains five trustees, one of who is Oscar. Furthermore, suppose that the set of drinkers includes Oscar. Then Eloïse has a winning strategy. Given:

$$G(\text{a trustee drank}, \mathcal{M})$$

Eloïse chooses Oscar and the game continues as:

$$G(\text{Oscar is a trustee and Oscar drank}, \mathcal{M}).$$

In addition, Oscar is placed in the choice set, in particular Oscar is an element of $I_{current}$. Since Oscar is both a trustee and a drinker, Eloïse will win. By Choice Preservation, Oscar is transferred to $I_{discourse}$ at the end of the sub-game, since Eloïse was playing as the verifier when she chose Oscar. Suppose that the next sentence is:

(10) He loves Old Crow.

We will give a proper version of the rule for interpreting pronouns in Section 3; for the moment, I will adapt the rule from Hintikka & Kulas (1985):

> **(R.he)**
> When a semantical game has reached a sentence of the form:
>
> $$X - \text{he} - Y$$
>
> an individual of the appropriate kind (a person or an animal), say b, is selected by the verifier from $I_{discourse}$, whereupon the falsifier chooses another individual, say d, from $I_{discourse}$. The game is continued with respect to:
>
> $$X - b - Y,\ b \text{ is a male, but d is not male.}$$

The idea behind **(R.he)** is that both the verifier and the falsifier select entities from the discourse model. If the falsifier can choose an entity that is male and distinct from the one chosen by the verifier, then the falsifier wins immediately. Notice that this rule predicts, incorrectly, that the following short text is incoherent, since the discourse model will contain two male individuals, either of which is an appropriate target for *he*; the falsifier should, therefore, always win:

(11) A man saw a boy steal an apple. He chased him down the block.

We will fix this problem in the next section.

Returning to example (10), we will now play a game on:

(12) a. $G(\text{He loves Old Crow}, \mathcal{M})$
b. $I_{discourse} = \{\text{Oscar}\}$

In the above context, Eloïse can pick Oscar as the referent for *he* while Abelard has no choice and must also pick Oscar. The game continues as a sequence of three games:

(13) Oscar loves Old Crow, Oscar is male, but Oscar is not male.

Assuming that Oscar is male and Oscar loves Old Crow, then Eloïse wins all three games in (13); note that she bets that *Oscar is not male* is false and, hence, takes the role of the falsifier in playing on this part of the game.

Compare the above with the rule for *every*:

(R.every)
If the game $G(S; \mathcal{M})$ has reached an expression of the form:

$$Z - \text{every } X \text{ who } Y - W$$

Then the falsifier may choose an individual from the appropriate domain, say b. The game is then continued as $G(Z - b - W, b \text{ is an } X \text{ and } bY; M)$. The individual b is added to the choice set I_S.

In this case, the falsifier is permitted to choose an entity. The verifier wins just in case it is falsifier who cannot find an entity d that will serve as a counterexample while falsifier wins if a counterexample is available.

Notice that *every* has a very different behavior with respect to anaphora than *some/a(n)*:

(14) a. Every student thinks he's treated unfairly.
b. Every student passed the exam. She studied very hard.
c. Every student passed the exam. They studied very hard.
d. Every student wrote an essay. One spelled most of the words correctly. He must have had a dictionary.

The familiar example (14)a shows that the quantifier *every student* can bind a pronoun within a sentence. In the course of a play, falsifier would pick a counterexample, say Julius, and drop him into the choice set. According to **(R.he)**, the verifier can pick an element of I_S to substitute for the pronoun. If she picks Julius, she has a potential winning strategy.

Examples (14)b and (14)c show that things are not so simple. The oddness of (14)b shows that falsifier's particular choice is no longer available after the game has been played. This is different from the behavior of *some*; a text made of (9) followed by (10) shows that when **(R.some)** is played, Eloïse's choice remains available on the next sub-game.

We can account for the difference between the *some* and *every* if we suppose that the set of students is added to the choice set at the end of the game in which the rule is played. If this is correct, then the use of the pronoun in (14)b will violate the definiteness required by **(R.he)**, giving the falsifier a strategy that wins no matter what. That is, (14)b is necessarily false, and therefore useless for communication; no one would use it outside of a logic class or a linguistics paper.

The difference between *some* and *every* is a function of Choice Preservation. This constraint says that Eloïse's choices, when she makes them as the Eloïse, survive the game in which they are made; otherwise, the set from which a choice is made survives, but not the individual choice. To motivate this, consider the following:

(15) Mary didn't see a student.

The example in (15) has two readings. One in which negation has scope over *a student* and one in which the relative scopes are interchanged. Scope is a matter of the order in which logically active elements are chosen, so there are two possibilities. The players could first play **(R.some)** and then play the negation. Suppose they do so. Eloïse would then select a witness for *a student*, substituting the witness' name for *a student*, say *Irving*:

(16) Mary didn't see Irving.

Next, the rule for negation must be played. Eloïse and Abelard exchange roles and Eloïse now bets that the sentence:

(17) Mary saw Irving.

is false. On this reading, Eloïse's choice of Irving is preserved in $\mathrm{I}_{\mathrm{discourse}}$. If the text continues with:

(18) He was hiding in the sewers.

then Eloïse can select Irving for *he* and all is well.

Suppose, however, that negation is played first. In that case, Eloïse and Abelard exchange roles and continue the play on:

(19) Mary saw a student.

Abelard is now betting that the above is true. Now, **(R.some)** must be played. Abelard, in his new role as the verifier, must select a student to witness the sentence. Suppose he picks Julius. The game continues on:

(20) Mary saw Julius.

If the above is false, then Eloïse wins. But Julius does not survive in $\mathrm{I}_{\mathrm{discourse}}$, which, instead, contains the set of students as required by *Choice Preservation*. Now if the discourse continues with:

(21) He was away.

the result is necessarily false and, therefore, incoherent. On the other hand, a continuation like:

(22) They were in Cancun for Spring Break.

The result is fine with *they* interpreted as the students.

Let us now return to example (14)b, repeated here as (23):

(23) Every student passed the exam. She studied very hard.

Application of **(R.every)** to the noun phrase *every student* in the first sentence allows Abelard to select a counterexample, say Julie, who is also placed in $\mathrm{I}_{\mathrm{current}}$. Play continues on:

(24) Julie passed the exam.

Putting aside the treatment of the definite description, *the exam*, suppose the sentence is confirmed by the model and that Abelard has no winning strategy on it. At the end of the game, Julie is not passed into $\mathrm{I}_{\mathrm{discourse}}$; rather, the set of students is placed there. When the continuation:

(25) She studied very hard.

is encountered, Abelard can always win, no matter what student Eloïse chooses, rendering the text trivial and incoherent.[7] Notice that the continuation in (14)c, repeated here:

(26) They studied very hard.

is coherent. The set of students is placed in $\mathrm{I}_{\mathrm{discourse}}$ and is, therefore, available as a target for a discourse anaphor.

Let us turn finally to (14)d, repeated here:

[7] Note the crucial assumption here that there are more than two female students in the model. If there were only one, the continuation would presumably be coherent. We will repair this in Section 3.

(27) Every student wrote an essay. One spelled most of the words correctly. He must have had a dictionary.

In order to account for the small text in (27) we need another game rule:

> **(R.one)**
> When a semantical game has reached a sentence of the form:
>
> $$\mathrm{X} - \text{one} - \mathrm{Y}$$
>
> an individual, say b, is selected by the verifier from a set contained in $\mathrm{I}_{\text{discourse}}$. The game is continued with respect to:
>
> $$\mathrm{X} - \mathrm{b} - \mathrm{Y}$$
>
> The entity b is then added to $\mathrm{I}_{\text{current}}$.

Playing **(R.every)** on *every student* in the first sentence of (27) implies that Abelard can select a counterexample. Suppose Abelard selects a student, Oscar, and the play continues on:

(28) Oscar wrote an essay.

Once the game on the first sentence is terminated, Oscar is deleted, but the set of students is added to $\mathrm{I}_{\text{discourse}}$. The play commences on the next sentence of (27):

(29) One spelled most of the words correctly.

Playing **(R.one)** allows the verifier, still Eloïse in her original role, to select an element from the set of students. Suppose she picks Lester. The play continues on:

(30) Lester spelled most of the words correctly.

Once the play terminates on this sentence, Lester is transferred to $\mathrm{I}_{\text{discourse}}$ as stipulated in *Choice Preservation*. Suppose the play continues on the third sentence of (27):

(31) He must have had a dictionary.

Playing **(R.he)** allows Eloïse to select Lester as the example of a student who must have had a dictionary. Notice that this treatment at least potentially captures the inference that the student who spelled most of the words correctly is the same as the student who must have had a dictionary. In Section 3 we will develop a system that guarantees the inference.

Now that we have given some examples of how game rules for quantifiers establish discourse entities, I will give a few examples of some more complex game rules that we will use in our discussion in Section 3. These rules cover cases like elementary cardinal quantifiers (Example (32)a), some bounding quantifiers (Example (32)b) and an example of majority quantifiers (Example (32)c). More complex examples can be found in Clark (2004, in press) and Pietarinen (2006).

(32) a. At least four deans smoked crack.
b. At most five faculty members hunted pigeons.
c. Most grad students must eat grubs in the winter.

Let us turn, first, to elementary cardinals like those in (32)a. Further examples, with their relationship to discourse anaphors are given in (33):

(33) a. At least 5 deans smoked crack. They passed out.
b. At least 5 deans drank Mad Dog. He passed out.
c. At least 5 deans dropped acid. One jumped out the window.

We give the game rule below. The basic idea is that the verifier must choose a witness set from the model. The falsifier must select an entity from this witness set. The witness set, itself, is placed in the choice set. If the sentence containing the elementary cardinal is true, then clearly it should be the case that the current verifier can select some number of witnesses for the sentence equal to the cardinality specified by the determiner. The falsifier will be unable to select a counterexample from this set. Thus, the verifier will have a winning strategy if the sentence is true. If the verifier is unable to select such a set, then she will be forced to select a counterexample. The falsifier then has a winning strategy: Pick the counterexample from the verifier's witness set. Hence, the falsifier has a winning strategy when the sentence is false.

(R.at least n)
If the game $G(S; \mathcal{M})$ has reached an expression of the form:

$$Z - \text{at least } n \text{ } X \text{ who } Y - W$$

then the verifier may choose a set of entities from the domain $\mathcal{M}$, call it ver($\mathcal{M}$), such that $|\text{ver}(\mathcal{M})| \geq n$. The falsifier then selects an entity $d \in \text{ver}(\mathcal{M})$. Play continues on $Z - d - W$, d is an X and $d - Y$. The contents of ver($\mathcal{M}$) are placed in I_{current}.

The above rule can correctly account for the judgments in (33). It will establish a set of discourse entities—the witnesses selected by the verifier—that can act as an antecedent for a plural pronoun like *they* in (33)a. The falsifier's choice will not be preserved, accounting for the impossibility of taking *he* to be one of the deans in (33)b. However, the set of witnesses selected by the verifier can provide the basis for *one* anaphora, as shown in (33)c.

The rule **(R.at least n)** differs from the usual game rules in GTS in that one of the players is permitted to choose an entire set. Normally, sets are constructed indirectly via the players' choices of individuals. Since we are not concerned with the foundations of mathematics, I will freely include sets in the ontology for natural language.

We should also note the lack of availability of the falsifier's choice within sentences and the relationship between singleton sets and entities. Consider the following short texts:

(34) a. At least one student thinks he's smart. He brags about it all the time.
b. At least two students claim he's wealthy. {They/He} will not stop talking about it.

According to our rule, the verifier chooses a singleton set to witness the properties in (34)a and the falsifier must, trivially, choose the sole element of the singleton. This element is available for intra- and inter-sentence anaphora. Compare this with the situation illustrated in (34)b. In this case, the verifier must choose a set of cardinality of at least two to witness the property. The falsifier can choose one of the elements of this set. His choice appears never to be available for within sentence anaphora: *he* in the first sentence in (34)b can never refer to one of the students. But why should this be? In other cases, the falsifier's choice is available for intra-sentential anaphora, as in the following example:

(35) Every student thinks he's smart.

Compare example (35) with the example in (36):

(36) All students think he's smart.

In (36), *he* cannot be dependent on *all students*. The problem appears to be morphosyntactic. The relevant pronouns in (34)b do not agree with their putative within sentence antecedents with respect to number. This appears to be sufficient to block anaphora. Compare (36) with (37):

(37) All students think they're smart.

In my dialect, (37) is at least two ways ambiguous. It can mean that each student believes of himself that he is smart or that all students believe that the members of the set of students are smart. This indicates that morphological agreement with the antecedent is crucial for intrasentential anaphora. This is in contrast with intersentential anaphor where the antecedent and the pronoun need not agree in number:

(38) Every dean wore a puce body-stocking. They thought it was becoming.

In the above, *they* can refer to the set of deans even though the noun phrase that evokes this set, *every dean* is singular, a fact we accounted for above.

Next, consider a bounding quantifier as in (32)b, repeated here embedded in a small text:

(39) At most five faculty members hunted pigeons. {They/One/#He} couldn't catch a single bird.

As the texts show, a plural pronoun or indefinite can have a bounding quantifier as discourse antecedent. The plural pronoun, *they*, is interpreted as the five or fewer faculty members who hunted pigeons; that is, *they* picks out the witnesses of the properties used in the first sentence of the text. A singular definite pronoun cannot pick out a member of this set.

One might suppose that we could construct the game rule for *at most* n by taking the **(R.at least** $n + 1$**)** and having the verifier and the falsifier exchange roles. That is, we could simply use the boolean structure of the set of determiner denotations to construct a game rule (see, for example, Keenan & Stavi 1986). The falsifier, playing as verifier, would select a set of $n + 1$ witnesses and the verifier, playing as falsifier, would try to select a counterexample from that set. If she has a winning strategy then the model must contain at most n witnesses as required. Notice that this approach would construct an inappropriate discourse antecedent since the set ver($\mathcal{M}$) would contain a non-witness, which is unacceptable.

I would argue that the correct rule would take the discourse effects of the quantifier into account. The following game rule is correct semantically and has the desired discourse effects:

(R.at most n)
If the game $G(S; \mathcal{M})$ has reached an expression of the form:

$$Z - \text{at most } n \text{ X who } Y - W$$

The verifier chooses a set of entities from the domain $\mathcal{M}$, call it ver($\mathcal{M}$), such that the cardinality of ver($\mathcal{M}$) is less than or equal to n. The falsifier chooses a disjoint set of entities from $\mathcal{M}$, call it fal($\mathcal{M}$), such that $|\text{ver}(\mathcal{M}) \cup \text{fal}(\mathcal{M})| > n$. The game then continues on:

> Z—every ver($\mathcal{M}$) —W, Z—no fal($\mathcal{M}$)—W, every ver($\mathcal{M}$) is an X who Y, every fal($\mathcal{M}$) is an X who Y.
>
> The set ver(M) is placed in $I_{current}$.

The game works by allowing both the verifier and the falsifier to select sets of entities; the union of these sets must exceed n. If the falsifier has a winning strategy, then he can drive the cardinality of the set of witnesses to a number that is greater than n. Otherwise, the set of witnesses must be less than n and the verifier has a winning strategy.

Finally, let us consider a game rule for a more complex quantifier like *most*. In the general case, *most* is not compact and, therefore, cannot be expressed as a first-order function.[8] It is unclear how to capture the meaning of *most* via a finite game. In particular, the falsifier should have a winning strategy on:

(40) Most integers are not divisible by five.

although it is difficult to see how to do this except through an infinite game or by the mechanism of allowing one of the players to deliver a proof. Below, I give a rule which works in finite models:

> **(R.most)**
> If the game $G(S; \mathcal{M})$ has reached an expression of the form:
>
> $$Z - \text{most CN who } P_1 - W$$
>
> where CN is a common noun and P_1 is a predicate, then the verifier picks a set of objects, call it ver($\mathcal{M}$), of cardinality:
>
> $$\frac{|CN|}{2} + 1 .$$
>
> The falsifier may choose an individual $d \in$ ver($\mathcal{M}$) and the game continues as:
>
> $$G(Z - d - W, d \text{ is a CN and } d\ P_1; \mathcal{M}) .$$
>
> The set ver($\mathcal{M}$) is then added to the choice set I_S.

The game rule **(R.most)** requires that the verifier select a set whose cardinality is greater than half that of the set denoted by CN. The falsifier may then select an element of that set to test the sentence on. If the falsifier cannot select a counterexample from the set, then it must be that a majority of the elements denoted by CN have the requisite property and the verifier wins. Notice that the difference between **(R.most)** and the game rules for the cardinal determiners resides in the requirement that ver(M) be of a particular size.

Finally, the rule requires that the set ver(M) be placed in the choice set. The discourse effect of **(R.most)** should be similar to those of the cardinal determiners. That is, singular pronouns will not match but plurals and indefinites will:

[8] See, for example, the proof of this in Landman (1991) among many other sources.

(41) a. Most deans practice fortune-telling. He is a reader of tarot cards.
b. Most deans are druids. They march about waving mistletoe.
c. Most deans hunt small game. One caught a pigeon.

The reader can verify that the rules have the correct results in these sentences. A more complete set of rules for quantifiers is given in Clark (2004) and Pietarinen (2006). Now that we have a set of interesting quantifiers to work with, let us now turn to the problem of managing their discourse consequences.

3 Game Theory and Discourse Anaphora

We have so far considered zero-sum games between two idealized agents. These games simulate the truth conditions of a sentence and, indeed, provide an interesting definition of *truth* as the presence of a winning strategy for Eloïse (or, alternatively, as a verification procedure) relative to a model. It should be clear that these zero-sum games do not succeed in fully capturing communication. It is nonsensical to suppose that one participant in a conversation plays the verifier and the other the falsifier, for example. Interlocutors have a shared interest in successful communication. Speakers encode meanings strategically based on shared knowledge. If, for example, the speaker has reason to suppose that the hearer can successfully assign a referent to a pronoun, she will use a pronoun. If the speaker does not have ground for this supposition, then she will use some other means—a definite description, for example. The form chosen for the definite description will, in turn, depend on the speaker's assessment of the hearer's knowledge of the intended referent. All of this suggests that the interlocutors are playing a cooperative game. We will follow this line of reasoning, while maintaining the sensible notion that the games involve finding and selecting entities and sets of entities from the model and the choice set.

In this section, I will discuss an analysis, developed further in Clark & Parikh (2006), of some straightforward examples of discourse anaphora in terms of cooperative games. Here, I will develop an analysis of games involving first one and then two discourse anaphors in a clause. Clark & Parikh (2006) extends the analysis to include a variety of factors that can influence judgments of coreference, including contrastive stress, lexical semantics and real world decision problems.

Let us begin with the analysis of some very elementary cases of discourse anaphora. The analysis is focused on languages like English which lack null pronouns. It is straightforward to extend the game analysis in these directions.

I will assume that discourse entities are introduced by quantifiers, as discussed above. For example, a sentence like:

(42) A cop saw a hoodlum.

introduces two entities into the discourse model, one for the cop and one for the hoodlum. Given the sentence:

(43) He yawned.

the hearer is faced with a decision problem: Which element of the discourse model should she take as the referent of *he*? If we extend these considerations slightly, we should observe that the speaker is also faced with a decision; namely, given a context and a discourse model, $\mathcal{D}$, should

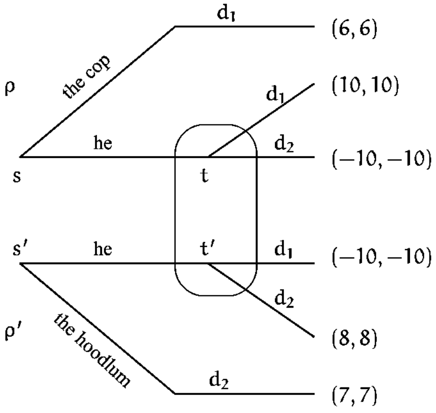


Figure 1: A game tree for simple discourse anaphora

he use a pronoun to refer to a particular element of $\mathcal{D}$ or should he use a definite description? Clearly there are costs for both choices. If the speaker chooses a pronoun, then he risks that the hearer will select an incorrect element of $\mathcal{D}$. If the speaker chooses a definite description, he reduces the risk of misunderstanding, but increases the amount of work that must be expended in producing and processing the utterance; definite descriptions are longer and syntactically more complex than pronouns.

The problem can best be solved as a game of partial information. The speaker and hearer share some common knowledge and have some interests in common. We can represent their common knowledge, their choices and their interests as a set of game trees. By finding the *Pareto-dominant Nash equilibrium* of the game, the players can most efficiently solve their problem and communicate.[9] Suppose, then, that the speaker has uttered the sentence in (42), introducing the following discourse entities:

(44) $d_1 =$ the cop
$d_2 =$ the hoodlum

Now suppose that the speaker wishes to encode the meaning that the cop yawned. Both the speaker and the hearer know that the speaker could refer either to the cop or the hoodlum using either a pronoun or a definite description. Having encoded the intended meaning, the hearer must decide whether to associate the expression with d_1, the cop, or d_2, the hoodlum.

[9] A Nash equilibrium is a strategy that offers each participant the best payoff given the strategies of the other players. That is, in a Nash equilibrium a player has no reason to change his or her strategy since any other move results in a lower payoff. A game may have several Nash equilibria. A Pareto-dominant Nash equilibrium is a Nash equilibrium whose payoffs are at least as high as the payoffs in any other Nash equilibrium; no other Nash equilibrium offers a better payoff.

The game tree in Figure 1 shows the various moves available to the speaker and hearer as well as the payoffs that they can expect once the choices have been made. Figure 1 shows two trees, one rooted in information state s and the other rooted in information state s'. Consider, first, the tree that is rooted at information state s. Information state s is associated with probability ρ and shows the case where the speaker intends to refer to d_1, the cop in the discourse model $\mathcal{D}$, while the tree rooted at information state s' (associated with probability ρ') shows the case where the speaker intends to refer to d_2, the hoodlum. The branches from the root show the possible moves that can be made by the speaker, while the branches emanating from these show the hearer's moves. The leaves show a set of ordered pairs of payoffs, where the first element is the payoff to the speaker while the second is the payoff to the hearer. Finally, the circled nodes are states which the hearer cannot distinguish.

Thus, if the speaker intends to refer to d_1 she can either use a definite description, *the cop*, or a pronoun, *he*. If she uses the definite description, she succeeds in referring to d_1 unambiguously but at a cost of some work for both her and the hearer; she must go to the effort of actually producing the definite description—which is work—and the hearer must go to the effort of processing it—which is more work. Furthermore, we will suppose that referring to a prominent element (the subject of the preceding sentence) with a full description rather than a pronoun entails some cost. Thus, I have shown a payoff of $(6, 6)$; the speaker and the hearer have communicated successfully, but at a cost. Suppose she uses a pronoun. Now, the hearer can either pick d_1, the intended referent, or d_2 the boy. In the former case, communication has succeeded at the cost of very little work. Both the speaker and the hearer are happy and get a payoff of $(10, 10)$. If, however, the hearer selects d_2, then communication has failed, an eventuality that both the speaker and hearer find unpleasant and wish to avoid. We therefore assign this outcome a payoff of $(-10, -10)$. The tree rooted at s' is nearly symmetrical, with d_1 and d_2 substituted for each other throughout the discussion. The one difference in payoffs—choosing d_2 for *he* has a payoff of $(8, 8)$ and not $(10, 10)$—reflects our assumption that it is slightly less efficient to pronominalize a less prominent element (in this case, an object).

Summarizing, the method of apportioning payoffs is based on the interaction of two principles:

(45) a. It is more costly to use longer expressions; pronouns are less expensive than names which are, in turn, less expensive than descriptions.

b. It is cheaper to refer to a more prominent element with a pronoun; it is correspondingly marked to refer to a more prominent element with a description or name when a pronoun could be used. Prominence is, here, calculated on the basis of the grammatical function the element plays in the preceding sentence.

The two principles in (45) rely on linguistic structure to establish the basic game trees. I should emphasize that contextual information can condition the probabilities associated with information states, with the result that the preferred strategy could change; see Clark & Parikh (2006) for more discussion.

Now, since there is no further information, let us suppose that $\rho = \rho'$. This means that information state s is as likely as information state s'. The Pareto-dominant Nash equilibrium is the strategy:

$$\{(s,\ \text{he}), (s',\ \text{the hoodlum}), (\{t, t'\}, d_1)\}.$$

That is, if the speaker is in information state s where he wishes to refer to d_1, she will use *he*. The hearer, who is now indeterminate between information state t and information state t', will

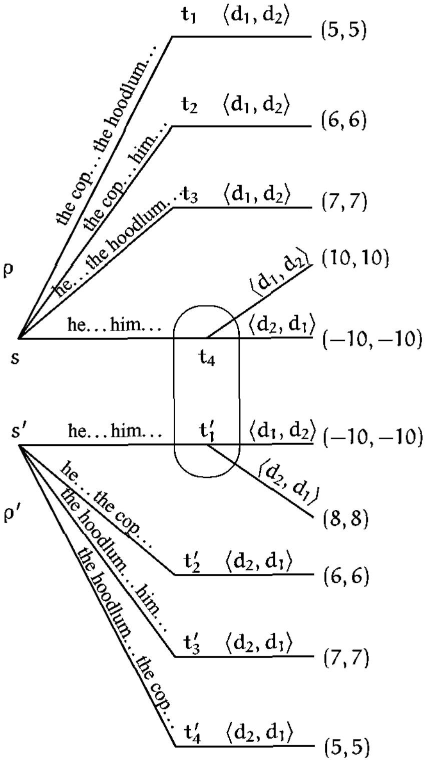


Figure 2: A game tree for two discourse anaphors

select d_1. If, on the other hand, the speaker is in state s' (where she wishes to refer to d_2, she will use *the hoodlum*); since the hearer's choice is determined unambiguously in the example, we have not included it in the strategy profile.

Now let us turn to the slightly more complex case where two pronouns are used:

(46) A cop saw a hoodlum. He chased him.

Again assuming that *a cop* invokes a discourse entity, d_1, and *a hoodlum* invokes entity d_2, there is a strong preference to take *he* as referring to d_1 and *him* as referring to d_2. The situation can again be represented as a game tree as shown in Figure 2.

The game involves two moves, as above. In this tree, I have shown only the sequence of the two possible referring expressions and the choice of two discourse entities. Thus, the speaker must decide whether to use two definite descriptions, a pronoun and a definite description or two pronouns.[10] Of course, if the speaker uses two pronouns, then the hearer cannot know with certainty which game tree he is in, a fact represented by circling the ambiguous information state nodes in the diagram. Both the speaker and the hearer can exploit properties of the grammar to narrow down the choices. For example, since the grammar rules out the case in which the two pronouns refer to the same entity, we need not include branches where the hearer chooses the same discourse entity twice.

The payoffs associated with each sequence of choices reflect the work of production and perception as well as success of communication. For example, the choice *the cop. . .the hoodlum. . .* guarantees successful communication but incurs work for both the speaker (in terms of production) and the hearer (in terms of perception). The choice *the cop. . .him. . .* is ranked slightly higher for both the speaker and the hearer since it involves successful communication with less work due to the replacement of one definite description by a pronoun. Notice that the best option for both the speaker and the hearer in terms of effort is *he. . .him. . .* where the hearer correctly chooses the discourse entities. The problem, of course, is that this encoding also runs the risk of miscommunication. A final factor I will take into account in determining payoffs is the relative prominence of an element; specifically, the subject of the preceding sentence is more likely to be the target of a discourse anaphor in the next sentence.

How should the speaker and the hearer play the game? If $\rho = \rho'$, as above, then the Pareto-Nash equilibrium is the strategy:

$$\{(s, \text{ he. . .him. . .}), (s', \text{ the hoodlum. . .him. . .}), (\{t_4, t_1'\}, \langle d_1, d_2\rangle)\}.$$

That is, if the speaker is in information state s, where he wishes to refer to the sequence of discourse entities $\langle d_1, d_2\rangle$, then he should use the pronouns *he* followed by *him*. The hearer should respond by picking the pair $\langle d_1, d_2\rangle$; that is, the choice where *he* $= d_1$ and *him* $= d_2$. If the speaker is in state s' then he should use *the hoodlum* and *him*, with hearer's choice being determined as shown in Figure 2.

Consider the following two short texts:

(47) a. A man saw a boy. He kicked the man in the shins.
b. A man saw a boy. The boy kicked him in the shins.

Although (47)a is interpretable, it is decidedly odd. The text in (47)b is entirely acceptable. The analysis of the game in Figure 2 correctly distinguishes between (47)a and (47)b on the assumption that the calculation of payoffs takes grammatical prominence into account.

Clark & Parikh (2006) assume that the game trees are associated with probability mass functions, ρ and ρ'. In fact, these probabilities are crucial in working out the Pareto-Nash equilibria of the games. In general, when there are n discourse entities to choose from, we will have n game trees whose roots are associated with probability mass functions $\rho_1, \ldots, \rho_n$, each ρ_i corresponding to the case in which discourse entity i is taken as the most prominent discourse entity. The game tree associated with each ρ_i would have 2^k branches, where k is the number of dis-

[10] One could, of course, represent each choice as a separate edge in the game tree. Thus, the speaker would first choose how to encode the subject of the sentence, the hearer would pick an entity from the choice set; then, the speaker would choose how to encode the second entity. The resulting game tree is more complex than the one in Figure 2 and harder to read. Since it does not really add information that is not already contained in the simpler figure, we have not shown it.

course anaphors in the expression since for each discourse anaphor, the speaker can select either a pronoun or a definite description.

The payoffs are structured to reflect the hierarchy in (48), where the grammatical functions in the ranking refer to the grammatical function played by the phrase that refers to the discourse entity in the sentence preceding the current sentence. The probability mass function ρ_1 is associated with an information state in which the speaker intends that the subject of the preceding sentence is selected as the target of a discourse anaphor or definite description, ρ_2 is associated with an information state where the indirect object is most prominent, and so forth. Consider the ranking in (48):[11]

(48) Subject > Indirect Object > Direct Object > Others.

Notice, though, that the relative prominence of elements is reflected in the payoffs of the game. The idea is that violating the ranking in (48) would carry a cost that is directly reflected in the payoffs given to the players. We will maintain this approach, although the ranking in (48) may be subject to linguistic variation (see Prasad 2003, and the references cited there). This approach suggests that cases in which the strategy profile provided by the Pareto-Nash equilibrium has apparently been violated are due to the fact that the probabilities associated with the information states have changed due to conditioning from other information sources, for example contrastive stress or lexical semantics.

I have left aside any discussion of apparent counterexamples to the strategies discussed above. As I have alluded to, the strategies can change depending on a variety of factors, for example contrastive stress (as in (49)a) or lexical semantics (as in (49)b):

(49) a. John called Bill a republican. Then hé insulted hím.
b. Mary insulted Sue. So she slapped her.

In (49)a, the usually judgment is that *he* can refer to Bill while *him* can refer to John. Equally, in (49)b *she* can refer to Sue and *her* to Mary. These judgments are the opposite of what we would predict on the basis of our simple analysis in this section. In both cases, however, the interlocutors can use contrastive stress or lexical semantics to condition their assessment of the subjective probabilities associated with the information states. This can result in a change in the Pareto-Nash equilibrium. These issues are discussed in more detail in Clark & Parikh (2006).

4 Conclusion

In this paper, I have developed a set of game rules for treating a broad class of quantifiers in English. A natural consequence of these rules is that they introduce new entities into the discourse model, treated here as the choice set, $I_{discourse}$. The game, here, is zero-sum and played on a model with the verifier and falsifier in direct competition.

The treatment of discourse anaphora, however, is rather different. In this case, there is little to be gained from conceiving of a verifier and falsifier in competition. Instead, as we have seen, it is more straightforward to suppose that the speaker and hearer have a mutual interest in establishing the content of discourse anaphors relative to $I_{discourse}$. Thus, the most natural analysis of these

[11] The ranking in (48) is the same as is assumed in much of Centering Theory. It is also in accord with our assumption that grammatical function correlates with ease of pronominalization. Centering theory is discussed in Joshi & Weinstein (1981,1983,1986) and Walker & Prince (1996) among many other sources.

elements is in terms of cooperative games. As we have seen, given plausible assumptions about the payoffs, the best interpretive strategy is given by calculating the Pareto-Nash equilibrium of the game. Since these games are public information, known to both the speaker and the hearer, the task of finding the referents for discourse anaphors is easily accomplished, with little risk of misunderstanding.

We can speculate that matters traditionally considered to be the province of semantics can be modeled by zero-sum games played on a model. The winning strategies for a sentence can therefore be used to characterize the class of structures that satisfy the sentence. Pragmatics, however, involves a different kind of information. Finding the intended referent for a pronoun, for example, is not so much a problem of truth conditions but, rather, a precondition for computing them. Pragmatic problems, such as presupposition and conversational implicature, are best treated as cooperative games between a speaker and a hearer rather than zero-sum competitions between virtual information agents.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the support of the NIH, grant NS44266. Thanks to Prashant Parikh for much useful discussion and to my students and colleagues at Penn for the many helpful comments.

REFERENCES

Beaver, D. I. (2001). *Presupposition and Assertion in Dynamic Semantics.* CSLI Publications, Stanford.

Clark, R. (2004). Game rules for quantifiers and discourse anaphora. Manuscript, University of Pennsylvania.

Clark, R. (2007). Games, quantification and discourse structure. In: *Logic and Games: Foundational Perspectives* (O. Majer, A.-V. Pietarinen and T. Tulenheimo, eds.).

Clark, R. and P. Parikh (2006). Game theory and discourse anaphora. Manuscript, University of Pennsylvania.

Groenendijk, J. and M. Stokhof (1991). Dynamic predicate logic. *Linguistics and Philosophy*, **14**, 39-100.

Grosz, B., Joshi, A. and S. Weinstein (1983). Providing a unified account of definite noun phrases in discourse. In: *Proceedings of the 21st Annual Meeting of the ACL*, pp. 44-50.

Grosz, B., Joshi, A. and S. Weinstein (1986). Towards a computational theory of discourse interpretation. Unpublished manuscript.

Hintikka, J. (1996). *The Principles of Mathematics Revisited.* Cambridge University Press, Cambridge.

Hintikka, J. and J. Kulas (1985). *Anaphora and Definite Descriptions: Two Applications of Game-Theoretical Semantics.* D. Reidel, Dordrecht.

Hintikka, J. and G. Sandu (1997). Game-theoretical semantics. In: *Handbook of Logic and Language* (J. van Benthem and A. ter Meulen, eds.), pp. 361-410. MIT Press, Cambridge, Mass.

Joshi, A. and S. Weinstein (1981). Control of inference: Role of some aspects of discourse structure-centering. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 385-387.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic.* Kluwer, Dordrecht.

Keenan, E. L. and J. Stavi (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, **9**, 253-326.

Landman, F. (1991). *Structures for Semantics.* Kluwer, Dordrecht.

Muskens, R., J. van Benthem and A. Visser (1997). Dynamics. In: *Handbook of Logic and Language* (J. van Benthem and A. ter Meulen, eds.), pp. 587-648. MIT Press, Cambridge, Mass.

Myerson, R. B. (1991). *Game Theory: Analysis of Conflict.* Harvard University Press, Cambridge, Mass.

Parikh, P. (2001). *The Use of Language.* CSLI Publications, Stanford.

Parikh, P. (2006). Radical semantics: A new theory of meaning. *Journal of Philosophical Logic.*

Pietarinen, A.-V. (2006). Semantic games and generalised quantifiers. This volume.

Prasad, R. (2003). *Constraints on the generation of referring expressions, with special reference to hindi*, Dissertation, University of Pennsylvania.

van den Berg, M. (1996). Dynamic generalized quantifiers. In: *Quantifiers, Logic, and Language* (J. van der Does and J. van Eijck, eds.), pp. 63-94. CSLI Publications, Stanford.

van Eijck, J. and H. Kamp (1997). Representing discourse in context. In: *Handbook of Logic and Language* (J. van Benthem and A. ter Meulen, eds.), pp. 179-237. MIT Press, Cambridge, Mass.

Walker, M. and E. Prince (1996). A bilateral approach to givenness: a hearer-status algorithm and a centering algorithm. In: *Reference and Referent Accessibility* (T. Fretheim and J. Gundel, eds.), pp. 291-306. John Benjamins, Amsterdam & Philadelphia.

This page intentionally left blank

Game Theory and Linguistic Meaning
Ahti-Veikko Pietarinen, editor
Current Research in the Semantics/Pragmatics Interface, Vol. 18

Chapter 15

THE SEMANTICS/PRAGMATICS DISTINCTION FROM THE GAME-THEORETIC POINT OF VIEW

Ahti-Veikko Pietarinen
University of Helsinki

This study examines the conceptual interplay between semantic and pragmatic aspects of linguistic meaning from the game-theoretic standpoint, and finds a negative result: that which is semantic and that which is pragmatic in language cannot be distinguished by means of the rule-governed and structural features of game theory. From that perspective, the sole difference is whether players entertain epistemic relationships with respect to the solution concepts and strategy profiles in the game-theoretic analysis of linguistic meaning. This means that, theoretically, the distinction is illusory.

1 ASSUMPTIONS

Let me start by outlining a few underlying assumptions that need to be acknowledged at the outset. First, I take meaning to loom in the relational action structure or the form that is essential in depicting games in their extensive forms. An extensive form of a game is a tree structure that lays bare the individual actions of the players as well as their responses to the actions of their adversaries. These games may be correlated with various things, such as formulas of logic, propositions, declarative and non-declarative assertions in natural language, or even some iconic and visual representations of our cognitive apparatus. One might be well advised to use the term 'signs', though this requires a separate argument which is beyond the scope of this paper.

In any event, that the structure is relational means that it is built from recurring interactions between those who utter and those who interpret the assertions. That the structure is extensive means, in the usual game-theoretic nomenclature, that it concerns not only the actual, but also the possible and counterfactual actions—the relational alternatives or referential multiplicities—of any particular or actual play of the game. Nevertheless, it is not the actions as such that correspond to the meaning, but the strategies, the exercise of which gives rise to actions.

That meaning is preserved in interactive structures has been prevalent throughout human inquiry. During the last fifteen years or so, interest in interaction has greatly expanded, bringing together masses of theoreticians and practitioners to bear on the topic. Computer scientists have begun looking closely into the idea to develop a general theory of semantics for programming languages (Abramsky, 2006). Linguists have incorporated interaction into their evolutionary and diachronic arguments for semantic and pragmatic change, though less often into game-theoretic outfits. For philosophers, the idea represents a time-honoured view of human discourse that has appeared in various metaphysical and logical guises ever since Plato's dialogues (Pietarinen 2003b, 2007a).

The individual disciplinary boundaries are not of too great a concern here; in each case the underlying terminology and the mathematical formalism is liable to be quite different, and geared to specialised theories. Yet the goal of the interdisciplinary enterprise is common: to get at the heart of meaning by methods that share general features, such as those analogous with how humans seem to accomplish this, through those concrete communicational and interactive practices and processes that take place between multiple agents with the application of multiple cycles of encounters throughout historical and evolutionary time.

The second assumption is that we can engage in semantics and pragmatics by applying the unifying conceptual framework, tools, and methods provided by game theory. We can engage in semantics, and indeed there is a time-honoured theory for doing so, by what is known in the trade as game-theoretic semantics (GTS; see Hintikka 1973, Hintikka & Kulas 1983, and papers by Clark, Pietarinen, Sandu and Scheffler in this volume, among others). Its motivations date back to certain venerable ideas in the history of philosophy, including Wittgenstein's language games (Wittgenstein 2000–, see the papers in this volume by Di Chio & Di Chio, Pietarinen and Sowa), Peirce's model-theoretic approach to logic (Peirce 1967, Pietarinen 2005b), and Kant's transcendental argument (Hintikka, 1973).

It is nearly as evident that we can study pragmatics by game-theoretic means. Such a methodology is cogently suggested by formal developments upon Grice's programme (Hintikka 1986, Parikh 2006). What is more, game-theoretic analyses of communication prompted, for the most part, by Lewis (1969), have burgeoned of late (see e.g. Allott 2006, Benz et al. 2006, Pietarinen 2006c, and papers in this volume by Guldborg Hansen, Alonso-Cortés and Miyoshi). This assumption is more contentious, however, since while linguistic interactions resonate closely with those of strategic interactions in anticipating the actions of others in order to increase, say, your communicative fitness, few agree on what the admissible, preferred ways of implementing this resemblance are or what they should be.

Pertinent questions include the following: What is the linguistic content of what actions represent? Are payoffs something that ought to be assigned to sets of such actions or do they go best with entire strategy profiles? Are there notions, such as Gricean intention or speaker's meaning, that do not naturally arise in, or are not well amenable to, game-theoretic analysis after all?

One of the consequences of this second assumption is that we can study semantics as well as pragmatics not only in terms of some well-chosen tools and methods of game theory but in terms of the logical and linguistic theory of GTS. And so it could as well be called game-theoretic pragmatics (see one of the early studies in this regard by Almog 1982, cf. Pietarinen 2001a).

Let me briefly justify. First, arguments for the usefulness of game theory in linguistic studies range over an area traditionally conceived as pertaining partly to the semantic and partly to the pragmatic study of meaning. GTS draws no a priori distinction between the two areas, however. The sundry postulation in the tradition of GTS has been that the classes of games that it studies and applies to various linguistic phenomena are strictly competitive rather than cooperative, and that the payoff structure is, for this reason, much simpler.

Second, although certainly a simplification of the theoretically and practically multifaceted notion of a game, GTS readily possesses genuine game-theoretic content. This is seen in the structure and formalism of those games that are capable of accommodating some basic notions such as actions, payoffs, strategies and different facets of information and its transmission. Hence GTS provides a platform for comparing semantics and pragmatics can from the game-theoretic point of view.

Since the intent of this paper is mostly conceptual and philosophical, it will focus not on a technical presentation of GTS but refer to the literature on the topic partly covered in the bibliography. My programmatic remark is that studies on the relationships between semantics and pragmatics take note of these profoundly philosophical and foundational questions and do not assume that understanding of the interplay will be considerably furthered simply by technical or empirical studies alone.

2 Analysis

Let me make five points that concern the role of GTS in the study of the semantics/pragmatics distinction as well as some of the relationships and the mutual points of contact between the two.

2.1 Truth, Meaning and Action

I have thus far spoken about meaning. However, GTS was originally devised to be a theory of material truth in the sense of merging truth-conditional semantics with a version of the verificationistic account of truth (Hintikka, 1973). Later, an array of studies appeared that aimed at extending the theory to cover not only expressions of logic, but also natural-language assertions (Saarinen 1979, Hintikka & Kulas 1983, 1985). In essence, this game-theoretic approach parallels truth with the existence of winning strategies for the utterer, who is the defender of the assertion (a.k.a. the verifier or Myself). Likewise, falsity is correlated with the existence of winning strategies for the interpreter, who is the opponent of the assertion (a.k.a. the falsifier or Nature). A similar thought emerged in Peirce's writings on logic (Hilpinen 1982, Pietarinen 2005b, Pietarinen & Snellman 2006). The notions of being true and being false are in this manner tied in with the existence of certain humanly attainable or humanly playable, rule-governed practices, activities and customs through which we come to observe and to realise the distributions of truth values that are linked with our linguistic assertions, assertoric practices and utterances.

One may see links with the philosophy of later Wittgenstein here, too, a point forcefully propounded by Hintikka & Hintikka (1986). Accordingly, both a version of verificationism and of truth-conditional semantics are attempted to be subsumed under a general theory for meaning, including aspects of how language is actually used.

But is meaning not something else or something more than just what correlates with material truth and verificationism? If the notion of truth agrees with the existence of some suitable strategies that show what the correct or optimal courses of actions are through the multiplicity of possible plays towards terminal positions, then the meaning is what gives rise to these actions together with all the alternative actions that might have come up in the course of playing the game. In other words, meaning is not the actions themselves nor is it to be found in observing, by sense observation or otherwise, the identities of any collection of available actions alone. Meaning is found in the more general mechanism or in the form that produces these actions. An entrepreneuring historian of ideas might attempt to relate such forms to the Aristotelian forms instantiated in the soul or in the mind. To describe the meaning is to refer to those actions that have been chosen or could have been chosen in the game associated with the assertions in question, but these actions themselves are not the meaning. To sketch a definition:

Definition. Meaning is that form of interactive processes that gives rise to the sum total of all actions, possible or actual, that arise, or may, will or would arise, as a consequence of playing the game across different contexts and in varying environments.

Two points must be highlighted. First, the sum total of all possible and actual actions referred to in this definition is what is exhibited by the extensive form of a game on an assertion. Hence the meaning involves considerations in the form of subjunctive conditionals: *If certain alternative actions were to be performed, then they would have certain consequences*. That some actions are merely possible has significant repercussions as to how we conceive the meanings of assertions to take shape from the vantage point of some particular play of the game that was in fact actualised. If possible actions were ineffectual to the development of the general mechanisms in which plays take their shape in the course of the game, our theory of meaning would be committed to relativism and, in the end, solipsism. For, if only action performed in this one, actual world of ours constitutes meaning, then all action constitutes meaning, language cannot be misused and there will be no false utterances. Moreover, no communication would be possible. Since communication clearly exists, what is meaningful to our general ways of acting upon our beliefs is constituted not by the actual actions and experiments upon utterances and assertions alone but by the application of actions and experiments and their modifications under any scenario and in any situation that may arise in the course of playing the game, as well as in the course of being prepared for repeating the games in whatever new situations or circumstances may come to pass.

Second, subjunctive considerations are empirically meaningful because they have practical consequences to our actions in the actual world. Since such actions can be correlated with the actual play of the game given by our preferred assemblies of strategy profiles, the alternatives to that play are the nitty-gritty of assessing the weight to be assigned to any particular choice illustrative of such profiles.

In summary, then, that which is to be taken into account in the definition outlined above includes those actions that lie on the off-equilibrium path, including the zero-probable actions. As any game theorist will be quick to confirm, a strategy of any practical use has to be prepared for unlikely actions as much as for those with higher probabilities.

2.2 Trichotomy of Contexts

Games are entities that by their very nature must be played in different circumstances, situations, locations and times. Hence the linguistically central notion of context has crucial undertones from the vantage point of game theory. There are several distinct but related ways in which it may enter the game-theoretic constitution of meaning. I will delineate three classes of contexts that game theory enables us to discern.

First, contexts may be encoded into 'chance moves'. They are actions performed by some third, fictitious player, call it Nature, who 'shuffles the deck' and by that manner functions as the 'probability generator' that designates individual contexts for each play. The idea is prevalent in game theory, but has not been applied to theories of semantics and pragmatics of language in full generality. Let us call linguistic contexts under this understanding A-contexts. For example, A-contexts relativise the winning strategy profiles not only to a possible-worlds type of semantic analysis, but also to 'contexts of play': namely to those circumstances or conditions under which each individual play takes place within the general framework of the whole game.

Chance moves that determine some key parameters of the game—for instance those concerning players' types—are often considered common knowledge. If not, the games are known as ones of incomplete information (see Harsanyi 1967 for the original account). Incomplete information is a prevalent phenomenon in game theory, economics and communication. It represents a 'veil of ignorance': players act while uninformed of the preferences and aims of the fellow participants. Formally, this is modelled by making the type-selecting chance moves members of the information sets in the extensive-form framework.

Such incompleteness is also commonplace in semantic and pragmatic theories of language. We are often not fully aware of the aims and purposes of our discourse participants, quite independently of whether we subscribe to cooperative communication. But such ignorance does not undermine the fact that we can be aware of several common characteristics of the game, without which the game would not be well defined. Section 2.5 on epistemology offers additional insight on the notion of common knowledge involved.

Second, notions of linguistic contexts may also be found in what is given by the earlier actions of the players along backtracked histories. We call these B-contexts; they arise in coreference, among others, and perhaps most conspicuously, in pronominal anaphora (see Clark's chapter in this volume as well as Janasik et al. 2003). B-contexts are also prevalent in interpreting sentences with multiple quantifiers and determiners (see chapters by Clark and Pietarinen in this volume), and are actively built in the course of the game, thus providing a dynamic and readily changeable notion of context. B-contexts imply that linguistic contexts, especially in communicative situations, are mutable and constantly accumulating, yet defeasible resources, and so rely on concepts with a game-theoretic and strategic character. B-contexts are prevalent in many pragmatic theories of linguistic meaning.

It is worth noting that B-contexts are created relative to actual plays of the game, since by traversing backwards we trace some particular histories that have already been realised. What this means in the treatment of anaphora is that the value to be found through such a process has already been selected earlier in the game and added into the set of such originally selected choices.

This is not an absolute requirement, however, since such values, and therefore the links for coreference may in certain cases be found from a supply of values totally different from those of players' previous choices. If this occurs, coreference is established through other types of context or a combination of them. A case in point is bridging: *Unfortunately, there is no live music in the club. Tonight, they are having a night off.* Such occurrences of coreference are nevertheless less frequent than those that B-contexts enable us to establish.

A third type of context can also be game-theoretically delineated. Contexts may be exogenous to the game, in whole or in part. In this regard, games are like open systems. They receive feedback and input from the environment within which they are situated. Let us call game-external contexts C-contexts. They are paramount to the Wittgensteinian notion of language games (see e.g. Sowa's chapter in this volume and Pietarinen 2003a), but highlight a wider phenomenon than what mere social factors can explain in linguistic comprehension. Focus, clefts, non-declarative moods, attitude descriptions and a multitude of any other type of similar modifiers are cases in point. They may well be partly grammaticalised, but the application and motivation for their use typically derive from the utterance's external surroundings. For example, C-contexts are commonplace in the interpretation of hedges (Almog, 1982). C-contexts also guide the selection of the values of coreferential expressions not found simply by looking at what has occurred with respect to some earlier parts of discourse, which is the case with the aforementioned bridging. This by no means prevents such choices—normally based on collateral observation and information—from being genuine parts of the game in the sense that they would not be congenial parts of what constitutes the strategy profiles of the game. Hence, they are in that very sense part and parcel of what constitutes the preferred solution concepts of the game. In that sense, it would be incorrect to state that C-contexts are altogether and absolutely exogenous to the theory of games.

Any three types of context may be combined, which is also likely to occur with natural language. As noted, anaphora may be attempted to be resolved by resorting to B-contexts. Often, however, deictic information, syntactic clues and other devices for extracting the required information are also indispensable in decoding the meaning of the utterance, and these typically appeal to C-contexts. Moreover, since A-contexts deal with what constitutes the common ground of language users, such as the common properties of the genus *Homo*, its knowledge and competence concerning the language in question, and its behaviour in communicative situations, are evidently also relevant to the meaning of anaphora.

One may think of A- and B-contexts as representing the 'narrow' understanding of context—B being the narrowest, referential and indexical one, and A being broader, but less expansive than C, which in turn may be regarded as the 'wide' context. The match is by no means perfect. For instance, the variability of 'context sets' is prevalent in type A as much as in type B, since chance moves need not be restricted to initial moves of a game, and since the variance is affected by the accessibility of information acquired from collateral cues and empirical observation as much as from what is asserted as the discourse unfolds. There is no hard and fast division between factual and conceptual information in the constitution of such contexts.

Moreover, C-contexts depend on the universes of discourse which are mutually understood to be the case, but which the players explore 'as they speak'. Such universes may be restricted in various ways, and the players need not be totally acquainted with them at the outset.

It is a remarkable feat of game theory to subsume such a variety of contexts within a single theoretical framework. It is equally notable how closely the three types of contexts pertinent in linguistic meaning fit the formal apparatus of game theory.

2.3 Who Plays the Games?

Strategic interactions move on two conceptually distinct levels. The first level is constituted by actual communicative actions and practices taking place between utterers and interpreters. The study of such actions pertains mainly to discourse analysis and the study of interpersonal communication to which methods of conversational games may be applied (see Miyoshi, this volume, among others).

Of greater concern, however, are the theoretical underpinnings of communicative interactions and the mediation of meaning in them. These structures are studied in GTS by applying game rules to the input data, which amounts to exchanges between two theoretical agents (the 'verifier' and the 'falsifier'). Agents are introduced to make the underlying conceptual mechanisms of interaction, not necessarily actual communicative interaction, better understood. Peirce once described it as a "sop to Cerberus" in order to emphasise the significance of logic in the study of meaning by making a resounding allegory with some common familiar phenomena. At the same time, he avoided falling back on a full-blown psychology or appeal to any singular human behaviour.[1] In brief, the sop expresses a refusal to identify agency with psychology.

In its somewhat limited sense and within the boundaries of linguistic theorising, it may be apposite to explain the sop as that modicum of rationality which needs to be injected into semantic and pragmatic theories of language in order to make them conform to one another. The sop will have to be thrown in order to make these theories mutually respectful towards certain principles and maxims of communication that became famous in many more or less like-minded philosophical theories of language, including those of Donald Davidson and H. Paul Grice (Pietarinen, 2004c).

In normative approaches to game theory we encounter a similar sop. The purpose of games is not to be motivated with experimental findings on how humans actually reason in making their strategic decisions in interactive settings, but with how they would rationally act (linguistically or otherwise), given the background that games are adapted to describe. There is little room for psychological processes in ordinary game theory, just as there is little room for psychological processes of reasoning in ordinary theories of logic. What matters is whether the choices of rational decision makers are good or bad, just as what matters in logic and semantics is whether an account of reasoning is good or bad, or whether our assertions can divide circumstances into those in which the assertions turn out to be true and those in which they turn out to be false. The parallel between the normativity of actions and the normativity of logic was foreshadowed in Peirce's assertion that, just as with ethics, logic ought to be regarded as a normative science.

One of the remaining topics to be brought to the fore concern the epistemology of such games. Epistemic issues concerning game theory and linguistic meaning have, if truth be told, been underrepresented in the current literature, despite the major repercussions they bring to the question of what is semantic and what is pragmatic in language.

2.4 Factual Versus Conceptual Truths

It is well established that GTS provides both the truth-conditions for logical expressions and the standard of meaning for a plethora of natural-language phenomena. But what is the overall methodological reach of GTS in that regard, theoretically speaking? One question that arises is whether GTS can—and if so, how well does it—cope with the interpretation of non-logical concepts. Prima facie, non-logical constants, including proper names, require theoretic methods that fix their intended reference. When first encountered, proper names behave more like variables than static, immutable and directly-referring singular terms. In this sense they may well have a scope just as that of logical constants.

This prima facie possibility is indeed realisable and, as such, simultaneously both extends the scope of GTS and reveals what semantics for atomic formulas and singular terms might look like. What is crucial here is not so much the actual set of game rules that could be evoked to implement the idea than the concrete implications of such an extension (Pietarinen, 2006d). As it happens, if some actual, humanly playable and rule-governed practices similar to those associated with complex formulas and utterances are involved in fixing the meanings of singular terms and proper names, then what we are accustomed to think of as 'analytic' truths are no

[1] To quote in full, Peirce writes in a letter to Welby, "I define a Sign as anything which is so determined by something else, called its Object, and so determines an effect upon a person, which effect I call its Interpretant, that the latter is thereby mediately determined by the former. My insertion of 'upon a person' is a *sop to Cerberus*, because I despair of making my own broader conception understood" (Peirce 1977, pp. 80–81, *Letter to Lady Welby*, 1908).

longer strictly separate from 'synthetic' truths that have to do with those boundary conditions by which we go about interpreting our non-logical vocabulary.

Support for the reality of this entanglement comes from the interpretation of predicate symbols of our non-logical vocabulary. For example, the significance of such terms in contributing to the meaning of assertions lies in the fact that they define the points at which individual plays of a semantic game terminate. Since such points of termination are co-located with terminal histories, in which the application of strategies is no longer possible and in which the attainment of the purpose of the players is mutually assessed, they in that very concrete sense are part and parcel of the game-theoretic construction of meaning.

What this also means is that, just as with logical constants, why non-logical vocabulary contributes to the normativity of GTS is due to the fact that the meanings of the constituents of non-logical vocabularies are grounded on the common understanding of the criteria needed for the attainment of the satisfaction of non-logical constants.

Moreover, the game-theoretically definable contexts, especially the A- and C-types, are closely related to collateral information obtained from sources subject to collateral observation. But that information may be both factual and conceptual, pertaining as it does to the environmental situation as well as to information about other players' types and their goals, including the payoff structure of the game, which for many purposes is constituted by taking into account common knowledge among the players.

What these points entail for theories of linguistic meaning is that factual and conceptual truths both contribute to the meaning of linguistic phenomena, and thus cannot serve as an implement of demarcation between what is semantic and what is pragmatic in such phenomena.

2.5 On the Epistemology of Solution Concepts

The foregoing remarks point to an issue in the need for conceptual clarification. It concerns the overall significance we ought to lay on various epistemic notions that permeate the game-theoretic analysis of linguistic meaning.

What are these notions? Typical epistemic characterisation results for solution concepts, such as Nash Equilibria, state that, given certain assumptions concerning the players' knowledge or belief about the game, the given equilibrium is a solution concept. To characterise Nash Equilibrium, for example, it is assumed that rationality, the payoff structure of the game, and the available actions are all common knowledge among the players. Without this common knowledge, and especially with respect to coordination problems, there are no solution concepts and equilibrium will be unattainable.

From the vantage point of linguistic meaning, what is notable is that these assumptions are similar to those that constitute reasonable presuppositions for successful communication. Three facets to such presuppositions can be discerned. First, the common ground, which, as noted above, may include both factual and conceptual truths about the situation or about the types and characters of one's adversaries, is established via the constitution of common knowledge about such truths, given reasonable postulates of rationality. Second, common knowledge about payoffs holds if no situational uncertainty or incomplete information prescribed by the games in question exists. Third, that available actions are common knowledge means that the utterer and the interpreter are sufficiently familiar with the universe of discourse in question, and that the other party is likewise sufficiently familiar with it. As noted in relation to C-contexts, the players need not be fully acquainted with the universe of discourse. But altogether lacking such acquaintance or familiarity or the aforementioned assumptions concerning the establishment of the common ground would jeopardise the possibility of the emergence of any sensible system of communication.

It is to be noted, however, that such epistemic characterisation results say little or nothing at all about the players' epistemic attitude about strategies. Yet that relationship is essential as far as pragmatic phenomena are concerned. To successfully use language is not only to master the game in order to be able to understand assertions or to be capable of computing or decoding what they convey or are intended to convey, but also to master their meaning in the crucial sense in which that meaning is given as a consequence of those actions, which in turn make the assertions understandable and comprehensible. Meaning, conceived under this qualification, is in strategic considerations governing individual actions.

In other words, assertions carry a force that is not brought out merely in actions. Such forces have to do with general phenomena, and function by way of appealing to collateral observation, information and other contextual features. Such forces have had different facets in the literature: intentions, conventional and conversational implicatures, generalised conversational implicatures and presumptions are among the familiar ones. True, the game-theoretic action structure is described by the individual actions, but assertoric meaning refers to general notions abstract from descriptions of individual actions or sequences of actions, and in that sense pertain to strategies that govern these actions.

To put the point in alternative terms, utterances do not constitute game-theoretic structures. To be able to utter and interpret one's utterances readily presupposes that language works as it does, and that the assumptions regarding the mutual knowledge of the key parameters of such structures are fulfilled. It is this descriptive and semantic function that is analysed by games, not the possible intentions and purposes that the agents might entertain in conveying, say, non-declarative moods and attitudes.

For example, in reliably asserting or claiming something to have a certain quality, one must already be acquainted with a range of human practices and customs connected with expressions that we customarily or habitually relate with various things and entities possessing qualities of a similar kind, or with anything customarily or habitually connected with the given quality. Moreover, such acquaintance must be mutual, which is to say that any utterer or interpreter is also aware of the fact that others are similarly familiar with and aware of the application of such practices and customs, and so on ad infinitum.[2]

3 Conclusions

What the similarities and dissimilarities in studies of semantics and pragmatics look like are very much brought to the fore as we move on to identify and assess some of the repercussions of the foregoing discussion. From the game-theoretic vantage point, there is no fundamental difference in characterising the meaning of some linguistic phenomena as pertaining to semantics or as pertaining to pragmatics, for the structures and fundamental resources of the underlying games are identical in both cases. The sole difference that justifies characterisation appears in the epistemic attitudes that a player has towards solution concepts, including considerations of the knowledge of players' strategies applied in the generation of that structure. If no epistemic rela-

[2] Let us re-emphasise that in GTS, it is not the utterer and the interpreter of actual linguistic assertions that are taken to bear an epistemic relation to strategies, only the theoretical players of the semantic game.

tionship exists between the players and their own or their opponents' strategies, the phenomenon may well be thought to pertain to semantics proper. If players know, however incompletely or with substantial uncertainty, what the strategies they are following consist of or what their essential content is, the phenomenon in question may be said to involve features that can, in normal situations mutually recognised as such, be characterised as pragmatic.[3]

Over and above such general characterisation of this distinction, what is semantic and what is pragmatic cannot, I submit, be distinguished independently of these epistemic considerations. In this forceful sense, the semantics/pragmatics distinction[4] is but moonshine: there are no a priori grounds for demarcating between the two realms. In other words, there is little prospect for stepping from one realm into another without changing the fundamental ways in which we make references to the players' epistemology in the description of the solution concepts of the underlying game. The difference between the two emerges through collateral observation and experience by which the players come to form their beliefs and predictions, and in that manner become aware of and acquainted with the strategies and their content employed in the course of the game. In this overall sense, semantics and pragmatics indeed form a unity.

One of the main complaints that might be voiced against the use of game-theoretic principles in studying the varieties of linguistic meaning is that games do not really seem to incorporate the idea of intended meanings, intentions or implicatures into their framework (for related criticism, see e.g. Allott 2006 and Sally 2003). They appear only to model actual interaction with explicit, manifest and identifiable actions. In other words, it may be asked what the game really 'means' in addition to such explicit, mutually testable actions that are produced in some formal framework such as GTS.

By way of recapitulating the point already argued for, it turns out that such an attempted counterargument for applying game theory to the constitution of meaning is misplaced. Semantics and pragmatics both operate on similar criteria, and thus cannot be separated from one another based on such criteria. Contexts are not to be thought of as an adequate candidate for such criteria. As noted, they enter game-theoretic meaning in various forms all of which can be given detailed, rule-governed descriptions, irrespective of whether the purpose is to articulate truth, the use of expression, or assertoric force.

As a result, the main reason for the inseparability of these two components is twofold. First, semantic relations are given by strategies that emerge and are contested by the populations of language users, namely strategies that are winning or have some related quality of approaching equilibrium for one of the subpopulations. Second, pragmatic relations between the interlocutors are detected and maintained by what the content of the strategies actually is, which presupposes epistemic access to them. This, in turn, can be captured in various ways depending on the representational systems in which players' epistemic and closely-related propositional attitudes are modelled.

An essentially similar argument counteracts the alleged difference between diachronic (historical) semantics and diachronic pragmatics (Pietarinen, 2006a). On the one hand, semantic change, including studies in historical semantics, is accounted for by requiring that the strategies in question have an attribute of stability of some appropriate kind over the recurring encounters and in varying situations and environments (Pietarinen, 2006b). On the other hand, pragmatic change, including studies in historical pragmatics (see e.g. Jucker 1994, Pietarinen 2007b), is

[3] And in that case, players may be held liable for their assertions. In the spirit of Wittgenstein's *On Certainty*, they must be able to demonstrate that they are in the position to have that knowledge.

[4] See Bianchi (2004), Szabó (2005) and Turner (1999), among others.

accounted for precisely by the same means as semantic change, namely by requiring that these very same strategies be stable over possibly indefinitely repeating plays.

The sole difference between semantic and pragmatic change is that the latter change is linked with the players' knowledge of the strategies in use throughout multiple runs of the game. Pragmatic change, just as with pragmatics in general, concerns the epistemic attitudes the players entertain towards strategies in view of constructing the preferred solution concepts in an evolutionary game. One such example would be evolutionarily stable strategies in evolutionary game theory. Intentions and the recognition and interpretation of adversaries' intentions (such as presumptive meanings or conversational implicatures) instantiate such attitude relations. The sole prerequisite for having an epistemic relation to strategies is to possess a sufficiently broad, collaterally acquired common ground concerning certain key features of the game, including payoffs and types of players and their presupposed rationality.

What comes to be added in such evolutionary games is the question of what the necessary and sufficient amount and type of information, including information about strategies, should be that is transmitted from any one instance or circumstance of playing the game, or a sequence of such instances or circumstances, to another in order for that information to trigger changes in meaning. For example, given the extensive-form framework for evolutionary games (Cressman, 2003), we have a pertinent theory of interactive, strategic situations at hand in which game-theoretically grounded evolutionary accounts of diachronic meaning could be studied. At all events, this approach holds a good deal of promise in exhibiting the kind of structure of relational multiplicity that codifies both the actual and the possible actions made during the recurring encounters and repeated plays.

What, then, is the wider, unifying phenomenon upon which both semantics and pragmatics may be said to represent our present-day reflections on linguistic, and more generally sign-theoretic, meaning? This is a question which cannot be answered here in full. Allow me merely to allude to the scholastic *speculative rhetoric* as an example of a study concerned with both semantic and pragmatic meaning. Peirce's term of art was *methodeutic* (Pietarinen, 2005b). It was mistakenly taken for pragmatics by Charles Morris and Rudolf Carnap, and the later tradition following these two propagators was similarly misguided (Pietarinen, 2007c). Consequently, pragmatics acquired psychological and sociological undertones, thus confining it to the realm of human language users, so much so that it has mistakenly been considered a part of such disciplines. Morris's behaviouristic interpretation of Peirce's semeiotic necessitated this turn, as it glossed over one of the main points of Peirce's theory, namely the strategic core of the Maxim of Pragmaticism (Pietarinen, 2005b). In fact, Peirce himself had considered and rejected the use of possible general psychological notions as proper explanations in the general theory of meaning, including conceptions, beliefs (hopes, fears, etc.), desires and expectations, and was left with habits as explanations of human interpretive activities, the 'logical interpretants' of our linguistic signs (Pietarinen & Snellman, 2006).

Other general psychological notions commonly believed to underlie actions that follow from an agent behaving in a certain way are intentions and reasons constitutive of an intentional agency. Such notions are, however, as dispensable as beliefs, desires or expectations are as adequate explanations in pragmatic theories of meaning.

Given his individualistic behaviourism, Morris never had any real use for Peirce's non-psychological concept of a habit. Likewise, in stark contradistinction to Morris, the scholastic speculative rhetoric pertained to the study of scientific methods and to the theory of scientific inquiry that would play a part in a general theory of interpretation. Speculative rhetoric was not intended to be a study of the relationships between signs (linguistic or otherwise) and any of their singular and actual interpreters, but the study of the relationships between signs and interpretants invariant over contexts, environments, and periods of time. This makes any 'actual use' part of the wider enterprise of sign meaning and interpretation. Peirce argued that the engine of this broader branch of science was the Maxim of Pragmaticism. Just as with the game-theoretic notion of a strategy, the Maxim is guided not by empirical criteria, but by counterfactual considerations. Just as strategies contribute to a unified account of the purposes and goals of actions, the pragmatic and habitual resolutions and plans of acting in a certain way in certain kinds of situations contribute to a unified theory of the meaning of those assertions. And it does so without positing any unnecessary psychological apparatus.[5]

In recent studies revolving around semantics, pragmatics, logic and communication, game theory has proved its strength and richness. Perhaps this is a sign of a significant theoretical unification towards a more general theory of meaning observed long ago, but which still awaits us in the future.

REFERENCES

Abramsky, S. (2006). Socially responsive, environmentally friendly logic. In: *Truth and Games* (T. Aho and A.-V. Pietarinen, eds.), pp. 17-45. Acta Philosophica Fennica **78**, Societas Philosophica Fennica, Helsinki.

Allott, N. (2006). Game theory and communication. In: *Game Theory and Pragmatics* (A. Benz, G. Jäger and R. van Rooij, eds.), pp. 123-151. Palgrave Macmillan, Basingstoke.

Almog, J. (1982). Game-theoretical pragmatics for ambiguity out of pragmatic wastebasket. *Theoretical Linguistics*, **7**, 241-262.

Benz, A., G. Jäger and R. van Rooij (eds.) (2006). *Game Theory and Pragmatics*. Palgrave Macmillan, Basingstoke.

Bianchi, C. (ed.) (2004). *The Semantics/Pragmatics Distinction*. CSLI Publications, Stanford.

Cressman, R. (2003). *Evolutionary Dynamics and Extensive Form Games*. MIT Press, Cambridge, Mass.

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

Grice, H. P. (2001). *Aspects of Reason* (R. Warner, ed.). Oxford: Clarendon Press.

Harsanyi, J. C. (1967). Games with incomplete information played by 'Bayesian' players. Part I: The basic model. *Management Science*, **14**, 159-182.

Hilpinen, R. (1982). On C. S. Peirce's theory of the proposition: Peirce as a precursor of game-theoretical semantics. *The Monist*, **65**, 182-188.

Hintikka, J. (1973). *Logic, Language-Games and Information*. Oxford University Press, Oxford.

[5] Pietarinen (2005) argues that Grice was a moderate anti-psychologist with respect to meaning.

Hintikka, J. (1986). Logic of conversation as a logic of dialogue. In: *Philosophical Grounds of Rationality* (R. E. Grandy and R. Warner, eds.), pp. 259-276. Clarendon Press, Oxford.

Hintikka, J. and M. B. Hintikka (1986). *Investigating Wittgenstein.* Blackwell, Oxford.

Hintikka, J. and J. Kulas (1983). *The Game of Language: Studies in Game-Theoretical Semantics and Its Applications.* D. Reidel, Dordrecht.

Jucker, A. H. (1994). The feasibility of historical pragmatics. *Journal of Pragmatics*, **22**, 529-547.

Lewis, D. (1969). *Convention: A Philosophical Study.* Harvard University Press, Cambridge, Mass.

Parikh, P. (2006). Radical semantics: A new theory of meaning. *Journal of Philosophical Logic*, **35**, 349-391.

Peirce, C. S. (1967). Manuscripts, The Houghton Library of Harvard University, identified by Richard Robin 1967. *Annotated Catalogue of the Papers of Charles S. Peirce.* University of Massachusetts Press, Amherst.

Peirce, C. S. (1977). *Semiotics and Significs. The Correspondence Between Charles S. Peirce and Victoria Lady Welby* (C. Hardwick, ed.). Indiana University Press, Bloomington.

Pietarinen, A.-V. (2001a). Most even budged yet: some cases for game-theoretic semantics in natural language. *Theoretical Linguistics*, **27**, 20-54.

Pietarinen, A.-V. (2001b). What is a negative polarity item? *Linguistic Analysis*, **31**, 165-200.

Pietarinen, A.-V. (2003a). Logic, language games and ludics. *Acta Analytica*, **18**, 89-123.

Pietarinen, A.-V. (2003b). Games as formal tools versus games as explanations in logic and science. *Foundations of Science*, **8**, 317-364.

Pietarinen, A.-V. (2004a). Semantic games in logic and epistemology. In: *Logic, Epistemology and the Unity of Science* (S. Rahman, D. M. Gabbay, J. P. Van Bendegem and J. Symons, eds.), pp. 57-103. Kluwer, Dordrecht.

Pietarinen, A.-V. (2004b). Multi-agent systems and game theory—a Peircean manifesto. *International Journal of General Systems*, **33**, 294-314.

Pietarinen, A.-V. (2004c). Grice in the wake of Peirce. *Pragmatics & Cognition*, **12**, 295-315.

Pietarinen, A.-V. (2005a). Relevance theory through pragmatic theories of meaning. In: *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society*, pp. 1767-1772. Lawrence Erlbaum, Alpha.

Pietarinen, A.-V. (2005b). *Signs of Logic: Peircean Themes on the Philosophy of Language, Games, and Communication* (Synthese Library **329**). Springer, Dordrecht.

Pietarinen, A.-V. (2006a). Evolutionary game-theoretic semantics and its foundational status. In: *Evolutionary Epistemology, Language and Culture: A Nonadaptationist Systems-theoretical Approach* (N. Gontier, J. P. Van Bendegem and D. Aerts, eds.), pp. 429-452. Springer, Dordrecht.

Pietarinen, A.-V. (2006b). The evolution of semantics and language-games for meaning. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, **7**, 79-104.

Pietarinen, A.-V. (2006c). Peirce, Habermas and strategic dialogues: From pragmatism to the pragmatics of communication. *LODZ Papers in Pragmatics*, **1**, 197-222.

Pietarinen, A.-V. (2006d). IF logic and incomplete information. In: *The Age of Alternative Logic: Assessing Philosophy of Logic and Mathematics Today* (J. van Benthem et al., eds.), pp. 243-259. Springer, Dordrecht.

Pietarinen, A.-V. (2007a). Towards the intellectual history of logic and games. In: *Logic and Games: Foundational Perspectives* (O. Majer, A.-V. Pietarinen and T. Tulenheimo, eds.).

Pietarinen, A.-V. (2007b). On historical pragmatics and Peirce's pragmatism. *Linguistics and the Human Sciences.*

Pietarinen, A.-V. (2007c). Significs and the origins of analytic philosophy. *Journal of the History of Ideas.*

Pietarinen, A.-V. and L. Snellman (2006). On Peirce's late proof of pragmaticism. In: *Truth and Games* (T. Aho and A.-V. Pietarinen, eds.), pp. 275-288. Acta Philosophica Fennica, Societas Philosophica Fennica **78**, Helsinki.

Saarinen, E. (ed.) (1979). *Game-Theoretical Semantics: Essays on Semantics by Hintikka, Carlson, Peacocke, Rantala, and Saarinen.* D. Reidel, Dordrecht.

Sally, D. (2003). Risky speech: Behavioral game theory and pragmatics. *Journal of Pragmatics*, **35**, 1223-1245.

Szabó, Z. G. (ed.) (2005). *Semantics versus Pragmatics.* Clarendon Press, Oxford.

Turner, K. ed. (1999). *The Semantics/Pragmatics Interface from Different Points of View.* Elsevier, Oxford.

Wittgenstein, L. (2000). *Wittgenstein's Nachlass, The Bergen Electronic Edition* (The Wittgenstein Trustees). Oxford University Press, Oxford.

Index

This page intentionally left blank

This page intentionally left blank