Information Transmission in Communication Games

Signaling with an Audience

by

Farishta Satari

A dissertation submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2013

This manuscript has been read and accepted for the Graduate Faculty in Computer Science in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Rohit Parikh

Date — Chair of Examining Committee

Ted Brown

Date — Executive Officer

Melvin Fitting

Bud Mishra

Stephen Neale

Howard Rachlin

Noson Yanofsky

Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

INFORMATION TRANSMISSION IN COMMUNICATION GAMES

SIGNALING WITH AN AUDIENCE

by

Farishta Satari

Adviser: Professor Rohit Parikh

Communication is a goal-oriented activity where interlocutors use language as a means to achieve an end while taking into account the goals and plans of others. Game theory, being the scientific study of strategically interactive decision-making, provides the mathematical tools for modeling language use among rational decision makers. When we speak of language use, it is obvious that questions arise about what someone knows and what someone believes. Such a treatment of statements as moves in a language game has roots in the philosophy of language and in economics. In the first, the idea is prominent with the work of Strawson, later Wittgenstein, Austin, Grice, and Lewis. In the second, the work of Crawford, Sobel, Rabin, and Farrell.

We supplement the traditional model of signaling games with the following innovations: We consider the effect of the relationship whether close or distant among players. We consider the role that ethical considerations may play in communication. And finally, in our most significant innovation, we introduce an audience whose presence affects the sender's signal and/or the receiver's response.

In our model, we no longer assume that the entire structure of the game is common knowledge as some of the priorities of the players and relationships among some of them might not be known to the other players.

to Mom and Dad

Acknowledgments

I would like to thank my advisor, Rohit Parikh, for scheduling his priorities around my life and all his advice. I would also like to thank the rest of my dissertation committee for their comments: Melvin Fitting, Bud Mishra, Stephen Neale, Howard Rachlin, and Noson Yanofsky. Thanks to Ted Brown for his unlimited support.

Thanks to the following people for acknowledging my existence and/or providing feedback on my work at various stages of development: Steven Brams, Adam Brandenburger, Herbert Clark, Michael Devitt, Juliet Floyd, Michael Franke, Uri Gneezy, Navin Kartik, Prashant Parikh, Steven Pinker, Brian Skyrms, and Joel Sobel.

I am forever in debt to my loving parents, Najib and Farzana, for the sacrifices that they have made, the hardships that they have gone through, and the values that they have taught me. Thanks to my brothers for enriching my life in many ways. To Wahid, thank you for loving me unconditionally and for standing by me every step of the way. Last, but not least, I thank Aaron for finding ways to distract me and give me advice, "*I am going to Disney Store but you go see Parikh. You can't become a doctor if you don't study. Do you want to become a doctor?*" Yes, I do Dr. Aaron Nowrouzie!

Contents

List of Figures

List of Tables

1 Introduction

The creation of *symbolic systems* was perhaps one of the greatest human inventions. Natural languages, non-verbal languages such as written or sign language, mathematical logic, or computer programming languages, they all serve the purpose of creating a repository of information, using objects and events to represent other objects and events forming discrete mental or machine representations. Each of them allows us to represent the world to ourselves and communicate it to others through language.

The formal inquiry of language and meaning is an interdisciplinary field of study that lies at the intersection of psychology, philosophy of language, economics, linguistics, and computer science. Psychology of language is concerned with the psychological and neurobiological factors that enable humans to acquire, use, comprehend, and produce language. In philosophy of language the inquiry into language and the nature of meaning dates back as far as Aristotle. What does it mean to *mean* something? What is the relationship between language and reality? How are sentences composed into meaningful wholes out of the meanings of parts? What is the social aspect of communication between speakers and listeners? And so on. In Economics, researchers study information flow in the market and how decisions are made in transactions by means of information exchange. The dynamics of information asymmetry is studied empirically and using theoretical models. Linguistics is the scientific study of human language; form, meaning, and use.

In the linguistics of both natural and artificial languages, syntax, se-

mantics and pragmatics categorize language characteristics. Syntax is the rules or form of representation that governs the way words are combined to form phrases, and phrases are combined to form sentences in a language, code, or other forms of representation. Semantics is the meaning of such words, phrases, sentences and how meaning attaches to larger chunks of text as a result of the composition from smaller parts. Pragmatics bridges the explanatory gap between sentence meaning and speaker meaning. It is the study of the relationship between the symbols of a language, their meaning, and use in a given context. In short, syntax is about form, semantics about meaning, and pragmatics about meaning that arises from use.

In computer science, an application of mathematical logic, formal languages take the form of character strings, produced by a combination of syntax grammar and semantics. A programming language is equipped with semantics that can be utilized for building programs that perform specific tasks. In computer languages syntax serves as the underlying grammatical structure of a program and semantics reflects the meaning. For example, $x \mathrel{+}= y$ in *Java* and *(incf x y)* in *Common Lisp* are two statements with different syntax but issue the same instruction i.e. arithmetical addition of y to x and storing the result in variable x. The semantic function of a programming language is embedded in the logic of a compiler or interpreter, which compiles or interprets the program for execution based on a mathematical model that describes the possible computations described by the language. An equivalent semantic function, not necessarily with the same representation, can presumably be found in the mind of the programmer. Mathematical models such as Backus Normal Form

BNF and parse trees are used for syntactical representation of programs while models such as *Denotational*, *Operational*, and *Axiomatic* semantics are used to explain code semantics. The role of pragmatics becomes obvious in the context of information exchange over the Internet and the Wolrd Wide Web. The Internet is a decentralized global network of interconnected computers using the standard protocol *TCP/IP* consisting of millions of business, government, private, academic, and other networks carrying information resources and services through interlinked documents. With the advancements in the last decades, we have made information available anywhere and anytime but not necessarily the right information.

Better formal models of communication and information exchange that capture game theoretic and social aspects of information transmission are a crucial step towards the realization of robust multi-agent systems that better understand and satisfy the needs of people and machines alike.

2 Philosophical Background

In the twentieth century, there have been two broad traditions in philosophy of language, the ideal language and the ordinary language traditions.

Ideal language philosophy originated in the study of logic and mathematics. Philosophers believed that for ordinary language to be unambiguous, it must be reformulated using the resources of modern logic. The predominant account in this tradition has been the view that the purpose of a sentence is to state a proposition and thus is true or false based on the truth or falsity of that proposition. In this view, sentences are treated as *propositions*; the semantic content of a sentence, which is either true or false depending on its agreement with reality. Language is then *about* the world and it *references* objects in the world.

Frege's [55] puzzle of identity shows that treating meaning as reference to objects runs into problems i.e. one cannot account for the meaning of certain sentences simply on the basis of *reference*. Where an identity statement like "*the morning star is the morning star*" is trivially true, there is much to be said about a statement like "*the morning star is the evening star.*" The first statement is true in virtue of language alone. However, the second statement has cognitive value. To solve this, Frege suggested that the words or expressions of a language have both a *reference* and a *sense*. Descriptions "*the morning star*" and "*the evening star*" reference the same object (i.e. planet Venus) but express different ways of conceiving it so they have different senses. The sense of an expression accounts for its cognitive significance. When two

objects have the same sense, they reference the same object. Expressions with different senses may reference the same object and we can't determine whether or not they do based on language alone. In other words, that "*the morning star is the evening star*," has to be an astronomical discovery.

Russell [123] developed the theory further but rejected Frege's notion of *sense* replacing it with the idea of a *propositional function*; an expression having the form of a proposition but containing undefined variables that become a proposition when variables are assigned values. He tried to analyze definite descriptors of the form "*The ...*" by distinguishing between logical and grammatical content of the sentence. Consider the statement, "*The present King of France is bald.*" Is it true or false? Russell proposed that when we say, "*The present King of France is bald*," we are implicitly making three separate existential assertions. First, there is an x such that x is a present King of France ($\exists x(Fx)$). Second, for every x that is a present King of France and every y that is a present King of France, x is the same as y ($\forall x(Fx \rightarrow \forall y(Fy \rightarrow y = x))$). Third, for every x that is a present King of France, x is bald ($\forall x(Fx \rightarrow Bx)$). These three assertions together say that the present King of France is bald [1].

Kripke [81] held the view that proper names do not have a sense and articulated his idea using the formal model of possible worlds. For example, take the current president of the United States of America, Barack Obama. When we say, "President of the United States in 2009," first we must state that the name "Barack Obama" is the name of a particular individual. Then we

[1] Also expressed as there is some x such that x is the present King of France, and if anyone happens to be the present King of France, it is x, and x is bald $\exists x(Fx \wedge \forall y(Fy \rightarrow y = x) \wedge Bx)$.

must imagine the possible worlds besides reality e.g. where Barack Obama was never born, did not go to Harvard, or chose a different career, etc. Then it is easy to see that a description like "President of the United States in 2009" does not necessarily describe Barack Obama, because it does not necessarily have the same value in all possible worlds. It only contingently describes Barack Obama. In comparison, he argues proper names like "Barack Obama" will always describe the same things across all possible worlds. Kripke calls terms that have the same reference across all possible worlds *rigid designators*. If we have two designators a and b that are rigid within metaphysical necessity, $a = b \rightarrow \Box(a = b)$ it doesn't follow that they are rigid in epistemic context, $a = b \not\rightarrow K(a = b)$. That is if $a = b$ is true then it is necessarily true (true in all possible worlds) but it doesn't follow that we know that $a = b$.

Idealist philosophers have often argued that two speakers can say the same thing by uttering different sentences, whether in the same or different languages, as long as the logical content of sentences agrees. For example, when a German speaker utters the sentence "*Schnee ist weiss*" and an English speaker utters the sentence "*Snow is white*," they have said the same thing. These two statements have different representation but both state a true fact about snow being white. Davidson proposed that we can give a finite theory of meaning for natural languages using Tarski's approach, the correctness test would be that it generates all the sentences of the form, *"P" is true if and only if P*. For example, because "*snow is white*" is true if and only if snow is white, the meaning of "*snow is white*" is snow is white. However, it is a much more difficult problem. Formal logic is more precise and decisive but constructing

an ideal language using formal devices whose sentences are clear, determinate in truth-value, and free from metaphysical implications is impossible. Even if such a precise language was realized, it is not guaranteed to be fully intelligible.

Ordinary language philosophy attempts at determining meaning in terms of language use; Strawson, Wittgenstein, Austin, and Grice were among the first contributors.

Strawson criticized Russell's characterization of statements where the referenced object doesn't exist as being false. He held the view that a statement like "*The present King of France is bald*," is neither true nor false but absurd. He argued that, if someone said, "*The present King of France is bald*," we would not say his statement is true or false since that question would not arise as there is no King of France[2]. We may think he is under miscomprehension but the statement will not have a truth-value. Strawson believed that use determines the meaning of a sentence.

In his latter work, Wittgenstein introduced the idea of *language games* drawing an analogy between language and playing a game and how both activities are rule governed, "*We can easily imagine people amusing themselves in a field by playing with a ball so as to start various existing games, but playing many without finishing them and in between throwing the ball aimlessly into the air, chasing one another with the ball and bombarding one another for a joke and so on. And now someone says: The whole time they are playing a ball game and following definite rules at every throw*" [156]. We use language following some rules but those rules don't need to be the same rules every-

[2]France is presently a republic and has no king.

where. We take part in a number of *language games* and confusion arises when a statement in one *language game* is interpreted according to the rules of another.

Austin [13] was the first who gave an account of sentence meaning in terms of speakers' actions. He argued that truth-evaluable sentences form only a small part of the range of all sentences and that there are other types of sentences, which perform actions or make the hearer take some action. By saying, "*I take this man as my lawfully wedded husband*," in the course of a marriage ceremony, the speaker is indulging in the act of marriage.

Grice [63] points out that the alleged divergence between formal logical devices such as $\wedge, \vee, \forall x, \exists x, \neg$, and their analogues in natural language arises due to use. He defines the Cooperative Principle and derives his celebrated theory of *implicatures* drawing clear distinction between sentence meaning and speaker meaning.

The ideal language philosophers have attempted to focus on reference or what language could be about, whereas, the ordinary language philosophers have tried to understand use or the communicative function of language. The later is of importance to the development of communication games as it deals with how *information* flows between individuals in a communication setting. Since the focus is on language use and speaker meaning, it is evident that questions arise about what someone knows and since knowledge presumes beliefs, *how do we know what someone believes?* Parikh [97] argues that such questions have been addressed by Ramsey, de Finetti, and Savage in the context of *decision theory* and the foundations of subjective probability. That is

beliefs are revealed by the choices we make, the bets we accept and the bets we refuse and among these choices are the choices of what to *say* and what to assent to. This view of statements as moves in a game gives us the flexibility to focus on the dynamics of *information exchange*, which will be explored in this thesis.

3 Meaning and Truth

Truth-conditional semantics for natural languages explain the meaning of assertions as being the same as, or reducible to, their truth conditions. Davidson [34] attempted at defining a semantic theory for natural languages along the same lines as Tarski's *semantic theory of truth* for formal languages. To understand Davidson's theory, let us first look at Traski's theory.

What is *truth*? As philosophical investigations reveal, this is not a trivial question. There are many theories of truth, among them is the *correspondence theory of truth* which suggests that the truth of a sentence depends on how it relates to reality. The sentence giving the truth condition of a sentence is called a *T-sentence*. A T-sentence takes the form; *"P" is true in language L if and only if P*, where the quoted sentence *"P"* is the name of a sentence in a language L and the unquoted sentence P is the translation.

T-sentences can be problematic as they can produce what is called a *Liar Paradox*. A *liar paradox* is a self-referential statement of the form *"This sentence is false"* and one arrives at a contradiction by reasoning about it. That is trying to assign a truth-value to this statement leads to a contradiction. If "*This sentence is false*" is true, then it is false, which would in turn mean that it is actually true, but this would mean that it is false, and so on. Similarly, if "*This sentence is false*" is false, then it is true, which would in turn mean that it is actually false, but this would mean that it is true, and so on.

Tarski addresses the issue of semantic incoherence in his famous *un-*

definability theorem. Informally, Tarski's undefinability theorem states that for any sufficiently strong formal system, the truth predicate of such a system cannot be defined within the system. He argued that in order to generate linguistic theories free of paradoxes, it is important to distinguish the language that one is talking about *object language* from the language that one is using *metalanguage*. Tarski demanded that a theory of truth must have, for every sentence P of a language L, a *T-sentence* of the form *"P" is true if and only if P*.

For a language L containing connectives $\neg$, $\wedge$, $\vee$ and quantifiers $\forall$ and $\exists$:

1. Negation $\neg$P is true if and only if P is not true
2. Conjunction P$\wedge$Q is true if and only if P is true and Q is true
3. Disjunction P$\vee$Q is true if and only if P or Q is true, or both are true
4. Universal statement $\forall xP(x)$ is true if and only if each object x satisfies "P(x)"
5. Existential statement $\exists xP(x)$ is true if and only if there is an object x which satisfies "P(x)"

   Consequently, the truth condition of complex sentences are built up on these connectives and quantifiers and can be reduced to the truth conditions of their constituents. The simplest constituents are atomic sentences and the truth for an atomic sentence is defined as:
6. Atomic sentence F(x_1, ...,x_n) is true relative to assignment of values

to variables $x_1, \ldots, x_n$ if the corresponding values of variables bear the relation expressed by predicate F.

Davidson proposed that we can give a finite theory of meaning for natural languages using Tarski's approach. To verify correctness we would test if it generates all the sentences of the form "*P" is true if and only if P*. For example, because *"snow is white" is true if and only if snow is white*, the meaning of *"snow is white*" is *snow is white*. Davidson's theory was harshly criticized by some philosophers namely Soames and Dummett. Soames [136] criticized truth-conditional semantics arguing that it is wrong and circular. Truth-conditional semantics gives every necessary truth precisely the same meaning as all of them are true under the same conditions. In other words, the bi-conditional *"if and only if*" ensures only that the left sentence will have the same truth value as the right sentence, therefore allows us to make any substitution of sentences on the right as long as its truth value is identical to the sentence on the left. For example, if *"snow is white" is true if and only if snow is white*, then it is the case that *"snow is white" is true if and only if snow is white and grass is green*, therefore under truth-conditional semantics *"snow is white*" means both that *snow is white* and that *grass is green*. Soames also argues that in specifying which of the infinite number of truth-conditions for a sentence will count towards its meaning, one must take the meaning of the sentence as a guide. However, the theory is to specify meaning with truth-conditions, and now it specifies truth-conditions with meaning, therefore it a circular argument. Dummett [44] objected to Davidson's theory on the basis that such a theory of meaning will not explain what it is a speaker has to know

in order for them to understand a sentence. The theory doesn't account for the learning process. Davidson realized the difficulty that natural languages are rich and include a variety of sentences such as indirect speech such as *"Galileo said that the earth moves,"* adverbial expressions such as *"John walked slowly"* where *"slowly"* modifies *"John walked,"* and non-indicative sentences such as imperatives such as *"Eat your food."* He further developed the *Principle of Charity*, originally introduced by Quine, which assumes that participants of a talk exchange are rational in the sense that they promote agreement which allows them to better understand words and thoughts of other people. In other words, in order to converse with someone you have to attribute to him or her mostly true beliefs.

4 Words as Actions

In his most influential work [13], Austin argues that there are sentences in the English language that are not statements of facts and thus neither true nor false.

I take this woman to be my lawfully wedded wife.

I name this ship the Queen Elizabeth.

I give and bequeath my watch to my brother.

I bet you sixpence it will rain tomorrow.

To utter one of these under the appropriate circumstances is not to describe what one is doing but rather doing the act itself. To utter, *"I take this woman to be my lawfully wedded wife,"* before the registrar or alter, you are not reporting on a marriage but rather indulging in it. Similarly, the other three sentences are not used to describe what one is *doing*, but used to actually *do* an action such as naming, giving, or entering a contract. In these typical cases, the action that the sentence describes is performed by the utterance of the sentence itself. Austin calls these types of sentences *performative sentences* or in short *performatives*. He argues that when something goes wrong in connection with a performative utterance then the utterance is *infelicitous* or *unhappy* rather than being true or false.

4.1 Speech Act

Austin distinguishes between *locutionary*, *illocutionary* and *perlocutionary* acts.

A *locutionary act* describes the linguistic function of an utterance i.e. the actual utterance and its ostensible meaning, comprising phonetic, and phatic acts corresponding to the verbal, syntactic and semantic aspects of any meaningful utterance.

An *illocutionary act* is the semantic of the utterance. The action, which a performative sentence performs when uttered, belongs to an *illocutionary act*. An *illocutionary act* is an act (1) for the performance of which the speaker must make it clear to the hearer that the act is performed, and (2) the performance of which involves the production of what Austin calls *conventional consequences*[3] such as rights, commitments, or obligations.

A *perlocutionary act* is the actual effect a sentence has whether intended or not. It can be thought as an effect of the illocutionary act. It is viewed at the level of its psychological consequences on the hearer or reader, such as persuading, convincing, scaring, enlightening, inspiring, or getting someone to do or realize something.

Ann to Bob: *"Do not go out without an umbrella."*

The utterance itself is a locutionary act with distinct phonetic, syntactic and semantic features. It counts as an illocutionary act of Ann warning

[3]The idea of conventions is further developed by David Lewis and described here under section 'Conventions'.

Bob about the weather, and if Bob indeed takes an umbrella then Ann has succeeded in persuading Bob and thus a perlocutionary act.

Bob to Ann: *"I have a Porsche; would you like a ride sometime?"*

The sentence has an illocutionary act of Bob offering Ann a ride and a perlocutionary act of Bob showing off or impressing Ann.

These different types of acts, in particular illocutionary act, is now widely known as *speech acts*. Austin's work was further developed by Searle [131] who classified illocutionary speech acts into different categories such as *assertives*, *directives*, *commissives*, *expressives*, and *declarations*.

Assertives are speech acts that commit a speaker to the truth of the expressed proposition. *Directives* are speech acts that cause the hearer to take a particular action e.g. requests or commands. *Commissives* are speech acts that commit a speaker to some future action e.g. promises. *Expressives* are speech acts that express the speaker's attitudes and emotions towards the proposition e.g. congratulating or thanking someone. *Declarations* are speech acts that change the reality in accord with the proposition of the declaration e.g. pronouncing someone guilty of a crime.

5 Intention-based Theory of Meaning

Grice's [63] work greatly influenced the way philosophers, linguists, and cognitive scientists think about meaning and communication. His work is a foundation of the modern study of pragmatics drawing clear distinction between speaker meaning, linguistic meaning, and the interrelations between these two phenomena. He examined how in an ordinary conversational situation a speaker S shapes his/her utterances to be understood by a hearer H and how both S and H observe some central principles during the talk exchange. His theory of meaning is one that is intention-based defining linguistic meaning in terms of speaker meaning, "*S meant something by U*" is roughly equivalent to "*S uttered U with the intention of inducing a belief in H by means of the recognition of his intention.*" Grice used Searle's example to make a distinction between linguistic and speaker meaning.

> An American soldier in the Second World War is captured by Italian troops. In order to get the Italian troops to release him he intends to tell them in Italian or German that he is a German soldier. He doesn't know Italian but says the only German line that he knows, *"Kennst du das Land, wo die Zitronen bluhen"* which in German means *"Knowest thou the land where the lemon trees bloom."* However, the Italian troops who do not know this meaning but can figure out the soldier is speaking in German, may reason as follows. *"The soldier just spoke in German. He must intend to tell us that he is a German soldier. Why would he speak in German*

otherwise? It could very well be that he is saying I am a German soldier."

Utterances can be divided into *Indicative* and *Imperative*. The first type of utterances are those that make the hearer *believe* something. The second type of utterances are those that try to make the hearer *do* something. Furthermore, Grice emphasized that not only should the hearer believe/do something by hearing the utterance but also recognize the speaker's intentions. An *utterance* could be any type of sound, mark, gesture, grunt, groan, etc. In other words, an *utterance* is anything that can signal the *intention* of the speaker.

5.1 Natural vs. Non-natural meaning

Grice understood meaning to refer to two rather different kinds of phenomena. *Natural meaning* is supposed to capture something similar to the relation between cause and effect as, for example, applied in the sentence *"Those spots mean measles."* This must be distinguished from what Grice calls *non-natural meaning* as in, *"Those three rings on the bell (of the bus) mean that the bus is full."*

5.2 Cooperative Principle and its Maxims

At the heart of Grice's theory of meaning lie the *Cooperative Principle* and its special *maxims* of conversation. Cooperative Principle is a set of *norms*

expected in a conversation. It mainly consists of four maxims. The *Quantity* maxim requires that a speaker is as informative as required. It relates to the quantity of information to be provided. The *Quality* maxim requires a speaker to tell the truth provable by adequate evidence. The *Manner* maxim requires the speaker to avoid ambiguity or obscurity, be direct and straightforward. Finally the *Relation* maxim requires a speaker's response to be relevant to topic of discussion.

The *Quantity* maxim has two sub-maxims. First the speaker is to make its contribution as informative as is required for the current purpose of the exchange and second, the speaker is not to make its contribution more informative than is required as it will bring confusion, raise side issues, and may mislead the hearer to think there is some particular reason or point in the provision of the excess of information. The *Quality* maxim also have two sub-maxims, namely speaker is not to say what they believe to be false and not say that for which they lack adequate evidence. The *Manner* maxim has various sub-maxims such as speaker is to avoid obscurity of expression, avoid ambiguity, be brief, and be orderly. The *Relation* maxim is to be relevant. Grice doesn't go into the possibility of questions that arise from this maxim. Questions such as what might be the different kinds of relevance or the shift of relevance during the talk exchange. He acknowledges the existence of all sorts of other maxims aesthetic, social, or moral in character such as being polite that may be observed by participants during talk exchanges.

A good question to ask is, why do people observe the Cooperative Principle? Grice assumed that people have learned to do so in childhood. Lewis explains

this in terms of social conventions. Both emphasize that the participants interest must be aligned with a common goal and the existence of some sort of mutual understanding in a talk exchange.

5.3 Conversational Implicature

Grice's Cooperative Principle and its maxims are rational along the following lines. Any one who cares about the central goal of a conversation given suitable circumstances must have an interest in participating in that talk exchange. Any kind of exploitation of these maxims gives rise to a *conversational Implicature*. A conversational Implicature falls under non-conventional Implicatures where the meaning of the utterance is not part of the conventional meaning of the words.

Bob: *I'm low on gas.*

Ann: *There is a station around the corner on Main St.*

Here Ann's utterance does not logically imply that the gas station is open. However, if both Ann and Bob are obeying the Cooperative Principle, Ann's remark is irrelevant unless the station is open. Therefore, Bob can infer from *Manner* and *Quality* maxims that Ann believes she has the evidence for the gas station being open.

A speaker S by utterring P has implicated Q, if

1. S is presumed to be observing the maxims under Cooperative Principle
2. S supposed that (s)he is aware that Q is required in order to make his/her

saying P consistent with this presumption

3. S thinks that it is within the competence of the hearer to work out or grasp that the supposition mentioned in (2) is required

   A hearer H can work out the Implicature by relying on the

4. Conventional meaning of the words used together with the identity of any references that may be involved

5. Cooperative Principle and its maxims

6. Context (linguistic or otherwise) of the utterance

7. Items of background knowledge

8. Fact that all the facts falling under (4), (5), (6), and (7) are available to both parties

It is worth mentioning that participants in a talk exchange may fail to fulfill the above requirements in various ways such as mislead (quietly violating a maxim in various ways), unwillingness to cooperate (opt out from the Cooperative Principle and its maxims or indicate the unwillingness to cooperate), be faced with a clash (unable to fulfill one maxim without violating another), or flout a maxim (fail to fulfill it).

6 Conventions

Hobbes [67] claims that the lives of individuals in the state of nature were *solitary*, *poor*, *nasty*, *brutish* and *short*, a state where *self-interest* and the *absence* of rights and contracts prevented the *social* or society to form. The society was in an anarchic state lacking leadership and its individuals apolitical and asocial. This state of nature was followed by the social contract where individuals came together and ceded some of their individual rights so that others would cede theirs. For example, I give up my right to kill you if you do the same. This resulted in the establishment of *society*, and by extension, the *state*, and a sovereign entity to protect these new rights, which were now to regulate social *interactions*.

Lewis [83] defines *conventions* as a set of generally accepted *social norms*, which are enforced if they sustain and become laws. For example, in some states, driving on the right side of the road, which may have started as convention to avoid collision, has now become a law. But a social norm doesn't need be a law in order to be a convention. Convention is described as a self-perpetuating solution to a recurring co-ordination problem i.e. no one has reason to deviate from it given that others conform. For example, if everyone drives on the right, you have reason to do so too, otherwise you will cause a collision. A convention for the same problem may exist under different form such as Europeans (at least in some countries) drive on the left side and Americans on the right side of the road. A *salient* solution must rely on what Lewis calls *precedence*. If both participants know that a particular

co-ordination problem has been solved in the same manner before by others, both know that both know this, both know that both know that both know this, etc. the solution is common knowledge, then they will easily solve the problem. Other people will in turn notice this and eventually the convention becomes widespread. Lewis formalizes this phenomenon as follows.

6.1 Formal Definition

A regularity R in the behavior of members of a population P when they are agents in a recurrent situation S is a *Convention* if and only if it is true that, and it is *common knowledge* in P that, in any instance S among members of P:

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

Where R' is some possible regularity in the behavior of members of P in S, such that no one in any instance of S among members of P could conform to

both R' and to R.

Lewis analyzing the nature of social conventions in game theoretic context and argues that social conventions such as the driving example, are solutions to *co-ordination problems*. Since participants' interests are fully aligned and the problem is finding a solution, there is no wrong solution so long as all participants pick the same solution. The difficulty lies with finding a *salient* solution since there may be several solutions to the problem.

6.2 Schelling's Focal Point

Schelling's [129] concept of *focal points* is one way to narrow down possible solutions for a coordination problem. Schelling illustrates this with a puzzle asking his students to answer the following questions. *If you had to meet a stranger in NYC, where and when do you meet them?* This is a co-ordination game, where any time in the day and place in the city could be an equilibrium solution. Schelling found that the most common answer was, *noon at Grand Central Station*. Note that there is no payoff related to this selection. Another place in the city, perhaps a public library, could hold the same payoff so long as enough people selected it as the meeting place. But, Grand Central Station's tradition as being a meeting place at the time raised its *salience* and made it a *focal point*.

6.3 Convention and Communication

The idea of linguistic conventions dates back to Plato [107] but David Lewis was the first to provide a systematic theory of how social conventions generate linguistic meaning. Consider the following example, the sexton of the Old North Church and Paul Revere (communicator and his audience) must coordinate to warn the countryside of an assault by the British army.

The sexton acts according to some contingency plan, such as:

R1:

If the redcoats are observed staying home, hang no lantern in the belfry.
If the redcoats are observed setting out by land, hang one lantern in the belfry.
If the redcoats are observed setting out by sea, hang two lanterns in the belfry.

or

R2:

If the redcoats are observed staying home, hang one lantern in the belfry.
If the redcoats are observed setting out by land, hang two lanterns in the belfry.
If the redcoats are observed setting out by sea, hang no lanterns in the belfry.

or

R3: *If the redcoats are observed staying home, hang one lantern in the belfry.*
If the redcoats are observed setting out by land, hang no lantern in the belfry.
If the redcoats are observed setting out by sea, hang two lanterns in the belfry.

There are three more contingency plans with no lantern, one lantern, and two

lanterns, plus any number of further plans involving other actions for example hanging three lanterns, hanging colored lanterns, waving lanterns, hanging a flag, and so on.

Paul Revere acts according to a contingency plan, such as:

C1:

If no lantern is observed hanging in the belfry, go home.
If one lantern is observed hanging in the belfry, warn the countryside that the redcoats are coming by land.
If two lanterns are observed hanging in the belfry, warn the countryside that the redcoats are coming by sea.

or

C2:

If no lantern is observed hanging in the belfry, warn the countryside that the redcoats are coming by sea.
If one lantern is observed hanging in the belfry, go home.
If two lanterns is observed hanging in the belfry, warn the countryside that the redcoats are coming by land.

or

C3:

If no lantern is observed hanging in the belfry, warn the countryside that the redcoats are coming by land.
If one lantern is observed hanging in the belfry, go home.

If two lanterns is observed hanging in the belfry, warn the countryside that the redcoats are coming by sea.

It does not matter what contingency plan is followed so long as the communicator and his audience coordinate on their plans. That is Paul Revere warns the countryside that the redcoats are coming by land if and only if the sexton observes them setting out by land, and that Paul Revere warns the countryside that the redcoats are coming by sea if and only if the sexton observes them setting out by sea. The coordination game for this example is given below.

		Paul Revere			
		C1	C2	C3	...
Sexton	R1	**1, 1**	0, 0	.5, .5	
	R2	0, 0	**1, 1**	.5, .5	
	R3	.5, .5	.5, .5	**1, 1**	
	⋮				⋱

Successful communication is achieved when the sexton and Paul Revere agree on one of the coordination equilibria that occurs along the diagonal, (R1, C1), (R2, C2), (R3, C3), ... and Paul Revere gives the right warning to the countryside.

6.4 Formal Definition of Signaling

Lewis formally defines two-sided signaling (as in the example of Paul Revere and the sexton) in which coordination is needed between a communicator and

his audience.

A two-sided signaling problem is a situation S involving an agent called the communicator and one or more agents called the audience, such that, and it is common knowledge for the communicator and the audience that: Exactly one of several alternative states of affairs $s_1, \ldots, s_m$ holds. The communicator, but not the audience, is in a good position to tell which one it is. Each member of the audience can do any one of several alternative actions $r_1, \ldots, r_m$ called responses. Everyone involved wants the audience's response to depend in a certain way upon the state of affairs that holds. There is a certain one-to-one function F from $\{s_i\}$ onto $\{r_i\}$ such that everyone prefers that each member of the audience do $F(s_i)$ on condition that s_i holds, for each s_i. The communicator can do any one of several alternative actions $\sigma_1, \ldots, \sigma_n$ $(n > m)$ called signals. The audience is in a good position to tell which one he does. No one involved has any preference regarding these actions which is strong enough to outweigh his preference for the dependence F of audience's responses upon states of affairs. Preferred response is the same for all members of the audience.

In Lewis's example, the sexton knows whether the redcoats are *staying home*, *coming by land*, or *coming by sea*. By placing either *zero*, *one*, or *two lanterns* in the belfry, he signals Paul Revere whether to *go home*, *warn people that redcoats are coming by land*, or *warn people that the redcoats are coming by sea*. A *signaling* problem in this sense is a coordination problem, because communicator and his audience must coordinate so that the communicator's signal result in the mutually desired action.

6.5 Meaning and Convention

Lewis argues that the use of language in a population consists of conventions of *truthfulness* and *trust* among members of the population. Given that this convention prevails, speakers who want to communicate have reason to conform to it, which in turn preserve the convention. He proposes that, *a language L is used by a population G if and only if there prevails in G a convention of truthfulness and trust in L, sustained by an interest in communication*, where a speaker is *truthful in L* if and only if she tries to avoid uttering sentences not true in L, and a speaker is *trusting in L* if and only if she believes that sentences uttered by other speakers are true in L. In many respects, Lewis's account of language and convention descends from Grice's theory of speaker meaning.

Austin's work explained that speech is not merely descriptive but can serve as an *action* that implies further action of some kind. He was perhaps one of the first to make this distinction. Lewis emphasized the existence of social conventions at the heart of which lie language and co-operative problem solving. Schelling's *focal points* gave an account of finding the salient solution among several solutions to a coordination problem. But in many cases we must rely on what Lewis calls *precedent* in order to get a salient solution. If both participants know that a particular co-ordination problem has been solved in the same way numerous times before, both know that both know this, both know that both know that both know this, etc. If it is common knowledge between them then they will easily solve the problem. Even more people will

see that they have solved the problem successfully, and thus the convention will spread in society. A convention exists because it serves the interests of people in a society. Needless to say a similar convention may exist that is entirely different. For example, it is more or less arbitrary that one drives on the right in the USA and left in some European countries. Lewis argued that in some sense language is ruled by conventions. Grice presupposed this phenomenon that if the participants in a talk exchange care about the central goal of a conversation, they will obey the Cooperative Principle and its maxims.

This literature is critical to the development of communication models using game theory. Both Lewis and Grice assumed that the participants' interests must be fully aligned in communication; thus limiting the discussion to coordination games. Of course, the case where participants' interests are not aligned is of less interest as no communication can take place. The more interesting case is where participants have partially aligned interests in a communication game. There has been some work done by economists and game theorists in this area, which we'll discuss after a formal overview of *decision theory* and *game theory*.

7 Decision and Game Theory

We play a *game* together with other people whenever we have to *decide* among several *actions* such that the *decision* depends on the *choice* of actions by others and on our *preferences* over the ultimate *result*.

7.1 Decision Theory

A *pure decision* problem is one where the outcome of an action solely depends on the state of the world and not on the actions of other players. A player chooses among several actions based on the state of the world and his preferences over expected outcomes. Preference means if a player can choose between actions a_1 and a_2, and prefer the outcome s_1 of a_1 over s_2 of a_2, then he prefers a_1 over a_2 and will choose a_1.

Suppose Ann would like to buy a pair of shoes and a purse. She prefers a pair of shoes and a purse over a pair of shoes only and a pair of shoes over nothing. Ann's preferences can be ranked as,

$$\text{pair of shoes and a purse} \succ_{Ann} \text{pair of shoes} \succ_{Ann} \text{nothing}$$

Another way to represent this ranking is by assigning numbers to the outcomes of Ann's choices (Table 1) called utilities. It has been shown mathematically that cardinal utility is invariant up to a certain positive affine transformation; utilities are arbitrary as long as they respect the preference orderings that they intended to represent.

Ann's Preferences	
Choice	Utility
Pair of shoes and a purse	2
Pair of shoes	1
Nothing	0

Table 1

If Ann's budget is such that she can purchase a pair of shoes and a purse then she will go with her first preference otherwise her second preference which is just a pair of shoes.

Decision theory is mainly divided into three branches; *decision under certainty*, *decision under risk*, and *decision under uncertainty*. Decision under certainty is the decision situation where a decision maker knows the outcome for each one of his actions. If each action leads to a set of possible outcomes where each outcome occurs with a certain probability and the decision maker knows these probabilities, then the decision situation is referred to as decision under risk. Decision under uncertainty is the situation where no probabilities for the outcomes are known to the decision maker, and further, no reasonable assumptions can be made about such probabilities.

7.2 Game Theory

What distinguishes game theory from decision theory is the fact that in game theory decisions have to be made with respect to the decisions of other players. Game theory has a prescriptive and a descriptive aspect. It can tell us how we

should behave in a game in order to produce optimal results or it can be seen as a theory that describes how players actually behave in a game. A game is a well-defined mathematical object consisting of a set of players, a set of moves or strategies available to those players, and a specification of payoffs for each combination of strategies.

Suppose Ann and Bob have to decide where to go out for the evening. Bob would like to go to a football match while Ann would like to go to a movie. Both would rather spend the evening together than apart.

$$\text{Movie} \succ_{Ann} \text{Football} \succ_{Ann} \text{Alone}$$

$$\text{Football} \succ_{Bob} \text{Movie} \succ_{Bob} \text{Alone}$$

Ann and Bob's utilities are shown in Table 2 and Table 3, respectively.

Ann's Preferences	
Choice	Utility
Movie	2
Football	1
Alone	0

Table 2

Bob's Preferences	
Choice	Utility
Football	2
Movie	1
Alone	0

Table 3

There are two common representations of games in the literature; the *normal form* and the *extensive form* representations.

The normal or strategic form game is usually represented by a matrix which shows the players, strategies, and payoffs. More generally it can be

represented by any function that associates a payoff for each player with every possible combination of actions. In each cell of the matrix, the first number represents the payoff to the row player, and the second number represents the payoff to the column player.

		Bob	
		M	F
Ann	M	**2,1**	$0, 0$
	F	$0, 0$	**1,2**

Figure 1: Strategic representation of battle of the sexes game where Ann and Bob choose between movie (M) or football (F).

The battle of the sexes game (Figure 1) has two equilibria. The equilibria set correspond to the choice of going to a movie together (M, M) or going to a football match together (F, F). If Ann and Bob decide to go to a movie then Ann's receives a payoff of 2 and Bob a payoff of 1. If they decide to go to a football match then Ann receives a payoff of 1 and Bob a payoff of 2. In addition to the two pure strategy equilibria, the game has a third mixed strategy equilibrium where Ann and Bob go to their preferred event more often than the other. In this equilibrium, Ann chooses movie with probability 2/3 and football with probability 1/3 and Bob chooses football with probability 2/3 and movie with probability 1/3[4].

The payoff matrix facilitates elimination of dominated strategies and is often used for his purpose. For example, in the prisoner's dilemma game (Figure 2), one can see that cooperate is strictly dominated by defect. Comparing

[4]The equilibrium is derived as follows. Let p be the probability Ann assigns to Football at equilibrium. Since the two pure strategies of Bob must yield equal expected payoffs, $2p + 0(1-\text{p}) = 0p + (1-\text{p})$ which implies $p = 1/3$. The other calculation is symmetric.

the first numbers in each column ($3 \succ 2$ and $1 \succ 0$) shows that no matter what Carl chooses, Bob can do better by choosing defect. Similarly, comparing the second payoff in each row ($3 \succ 2$ and $1 \succ 0$) shows that no matter what Bob chooses, Carl can do better by choosing defect. Thus, the unique Nash equilibrium of the game is (D, D) where both Bob and Carl choose defect and receive a payoff of 1 each. However, Bob and Carl can receive a better payoff if they both cooperate.

		Carl	
		C	D
Bob	C	$2, 2$	$0, 3$
	D	$3, 0$	**1,1**

Figure 2: Strategic representation of prisoners dilemma game where Bob and Carl choose between cooperate (C) and defect (D).

When a game is presented in normal form, it is presumed that the players act simultaneously or, at least, without knowing the actions of the other. If players have some information about the choices of other players, the game is usually presented in an extensive form.

A tree structure is used for graphical representations of games in the extensive form. This form is useful for the representation of dynamic games; games where there may occur whole sequences of moves by different players with some order such as in a chess game. In a tree, each vertex or node represents a point of choice for a player. The player is specified by a letter listed by the vertex. The lines out of the vertex represent a possible action for the player. The payoffs are specified at the bottom of the tree.

Figure 3 shows a game of perfect information. A, B, M, F stand for

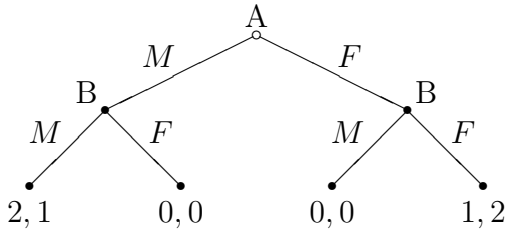


Figure 3: Extensive representation of battle of the sexes game with perfect information.

Ann, Bob, movie, and football respectively. A moves first and chooses either M or F. B sees A's move and then chooses M or B. Suppose that A chooses F and then B chooses F, then A gets 1 and B gets 2. Bob knows Ann's choice prior to making a move.

The extensive form can also capture simultaneous-move games of imperfect information. A dotted line is drawn along two different vertices to represent them as being part of the same information set where Bob does not know which point he is at (Figure 4).

However, both are games of complete information because there is no uncertainty for the players about which game is being played i.e. there is a single initial vertex. In games of incomplete information there are two or more initial vertices and perhaps an initial move by nature. An oval around vertices indicate they are part of the information set of a player in an incomplete game.

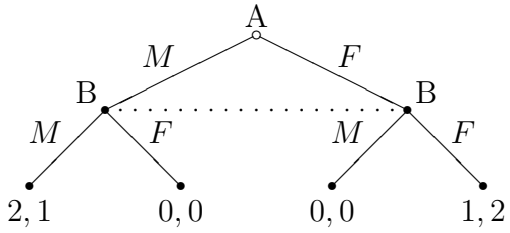


Figure 4: Extensive representation of battle of the sexes game with simultaneous moves.

7.2.1 Classification of Games

Games are often classified into *static*, *dynamic*, *cooperative*, and *non-cooperative* games. In static games, every player performs only one action, and all actions are performed simultaneously. Static games can be represented with a payoff matrix. In a two-player game, one player is called the row player and the other is called the column player. In dynamic games, there is at least one possibility of performing several actions in sequence. These types of games are represented using the extensive form. In a cooperative game, players are free to make binding agreements in pre-play communications. This means that players can form coalitions. In non-cooperative games no binding agreements are possible and each player plays for himself.

7.2.2 Formal Framework

N = 1, ... , n is the set of players who choose actions and have preferences over outcomes. A_i is the set of actions available to player i. An *action profile* $(a_1, \ldots, a_n)$ is an n-tuple of actions where each $a_i \in A_i$ is performed simultaneously. A strategy tells players what to do given background knowledge. It is a function from sequences of previous events or histories to action sets[5]. The binary relation $\preceq_i$ represents *preference* between profiles or payoff functions. The payoff function u_i maps profiles to real numbers. If $(s'_1, \ldots, s'_n) \prec_i (s_1, \ldots, s_n)$ or $u_i(s'_1, \ldots, s'_n) \prec u_i(s_1, \ldots, s_n)$, then player i prefers strategy profile $(s'_1, \ldots, s'_n)$ being played. The payoff profiles $(u_1, \ldots, u_n)$ define the

[5]Actions are used in static games instead.

payoff function U of a game. $U : A \to R^n$ is a function mapping all actions or strategy profiles to payoff profiles. If $S = (s_1, \ldots, s_n)$ is an action, strategy, profile then $S_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$. An action a_1 strictly dominates another action a_2 if a_1 is preferred to a_2 in all possible courses of events.

7.2.3 Nash Equilibrium

Strategic games can be classified according to how much the payoff functions of the players resemble each other. In *zero-sum games* also called *strictly competitive games* the payoff of players sum up to zero; if one wins a certain amount then the other loses it. A *pure coordination game* is the opposite of a zero-sum game where the payoffs of the players are identical.

In a strategic game without uncertainty, what strategy will a rational player choose? One way to answer this question is to say a player may just eliminate all *strictly dominated actions*, and hope to find a single possible move to choose. *Strict strategy domination* is based on players' preferences and is formalized as follows.

Definition 1: Strategy s_i of player i *strictly dominates* a strategy s'_i if and only if for all profiles s it holds that $(s'_i, s_{-i}) \prec_i (s_i, s_{-i})$.

Definition 2: A atrategy s_i of player i *weakly dominates* a strategy s'_i if and only if for all profiles s it holds that $(s'_i, s_{-i}) \preceq_i (s_i, s_{-i})$ and there is a profile s such that $(s'_i, s_{-i}) \prec_i (s_i, s_{-i})$.

Nash equilibrium is a kind of solution concept of a game involving two or more players, where no player has anything to gain by changing only his or

her own strategy from unilaterally. If each player has chosen a strategy and no player can benefit by changing his or her strategy while the other players keep theirs unchanged, then the set of strategy choices and the corresponding payoffs constitute a Nash equilibrium.

Definition 3: A strategy profile s is a *weak Nash equilibrium* if and only if for none of the players i there exists a strategy s'$_i$ such that s $\prec_i$ (s'$_i$, s$_{-i}$) or equivalently if for all of i's strategies s'$_i$ it holds that (s'$_i$, s$_{-i}$) $\preceq_i$ s. A strategy profile is a *strict Nash equilibrium* if $\prec_i$ is used instead or $s_i \neq$ s'$_i$ for the second characterization.

There is another characterization in terms of best responses. A move s$_i$ of player i is a best response to a strategy profile s$_{-i}$. We write s$_i \in$ BR$_i$(s$_{-i}$), if and only if u$_i$(s$_i$, s$_{-i}$) = max$_{s'_i \in S_i}$ u$_i$(s'$_i$, s$_{-i}$).

A strategy profile s is a *Nash equilibrium*, if and only if for all i = (1, ..., n) s$_i$ is a best response to s$_{-i}$ i.e. s$_i \in$ BR$_i$(s$_{-i}$). It is strict if in addition BR$_i$(s$_{-i}$) is a singleton set for all i.

A strategic game with mixed strategies is defined as follows. Let $\Delta(A_i)$ be the set of probability distributions over A$_i$, i.e. the set of functions P that assign a probability $P(a)$ to each action a $\in$ A$_i$ such that $\sum_{a \in A_i} P(a) = 1$ and $0 \preceq P(a) \preceq 1$. Each $P \in \Delta(A_i)$ corresponds to a mixed strategy of player i. A mixed strategy profile is a sequence (P$_1$, ..., P$_n$) for the set of players N = {1, ..., n}. A pure strategy corresponds to a mixed strategy P$_i$where P$_i$(a) = 1 for one action a $\in$ A$_i$ and P$_i$(b) = 0 for all other actions. We can calculate the expected utility of player i given a mixed strategy profile P = (P$_1$, ...,

P_n) and payoff profile (u_1, ..., u_n) by $EU_i(P) = \sum_{a \in A_1 \times \ldots \times A_n} P_1(a_1) \times \ldots \times P_n(a_n) \times u_i(a)$.

It is assumed that rational players try to maximize their expected utilities, i.e. a player i strictly prefers action a over action b exactly if the expected utility of a is higher than the expected utility of b. For mixed strategy profiles $P = (P_1, \ldots, P_n)$, we use the same notation P_{-i} as for pure strategy profiles to denote the profile $(P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n)$ where we leave out the strategy P_i. (P'_i, P_{-i}) denotes again the profile where we replaced P_i by P'_i.

Defition 4: A *weak mixed Nash equilibrium* is a mixed strategy profile (P_1, ..., P_n) such that for all i = (1, ... , n) and P'$_i \in \Delta(A_i)$ it holds that EU_i(P'$_i$, P_{-i}) $\leq EU_i(P)$. A mixed Nash equilibrium is *strict* if we can replace $\leq$ by $\prec$ in the last condition.

A Nash equilibrium such as (a, a) is called *strongly Pareto optimal*, or strongly Pareto efficient; more precisely, a Nash equilibrium s = (s_1,..., s_n) is strongly Pareto optimal, if and only if there is no other Nash equilibrium $s' = (s'_1, \ldots, s'_n)$ such that for all $i = (1, \ldots, n) u_i(s) \prec u_i(s')$. That is a Nash equilibrium is strongly Pareto optimal if and only if there is no other equilibrium where every player is better off.

7.2.4 Common Knowledge and Rationality Assumptions

The classical interpretation of game theory makes very strong assumptions about the rationality of players. First, it is assumed that every player is logically omniscient; they know all logical theorems and all logical consequences

of their non-logical beliefs. Second, they are assumed to always act in their enlightened self interest in the sense of utility maximization. Third, for a concept like Nash equilibrium to make sense in classical game theory, it is assumed that the structure of the game is common knowledge between players.

Mutual knowledge of a proposition α between players is when each player knows α. For example, both Ann and Bob know α. Common knowledge between two players of a proposition α is equivalent to two infinite chains of knowledge of α. Ann knows that Bob knows that Ann knows that ... α and Bob knows that Ann knows that Bob knows that ... α. Of course, human beings are seldom able to go beyond just a few iterations of shared knowledge; they can't explicitly represent the infinite chain of knowledge.

8 Communication Games

There are different models of a two-player communication game in economics and game theory literature. In these models, one player (the sender or agent) tries to communicate some private information (sender's type) to another player (the receiver or principle). These games model the difficulties that arise under conditions of incomplete and asymmetric information. Rasmusen [116] divides communication games into different types. We'll very briefly review them here before discussing cheap talk games in detail.

8.1 Signaling

In *signaling games*, the sender's message is costly and more costly when he lies than tell the truth, but messages need not be truthful. The sender's payoff is affected even if the receiver ignores his message. The sender's type varies from bad to good in these models. If the sender's type is better, it is cheaper for him to send a message that his type is good. For example, a worker has a given skill level and chooses the amount of effort he will exert. If the worker knows this and can acquire credentials to signal his ability to an employer then the problem is signaling.

8.2 Truthful Announcement

In *truthful announcement games*, the sender may be silent or send a message, but the message must be truthful if it is sent. There is no cost to sending the

message, but it may induce the receiver to take actions that affect the sender. If the receiver ignores the message, the sender's payoff is unaffected by the message. An example of a truthful announcement game is when the sender's ability A is uniformly distributed on [0,1], and the sender can send a message Y such as $A > .5$ or $A = .2$.

8.3 Auditing

In *auditing games* the sender's message might or might not be costly and receiver may audit the message at some cost to verify if the sender was lying. An example is lobbying. The lobbyist can tell the truth or lie (in both cases sending a costly message) to the politician. The politician can then investigate the truth of the message at some cost.

8.4 Mechanism

Sender's message might or might not be costly. Before the sender sends a message he commits to a contract with the receiver. Decisions are based on what they can observe and enforcement is based on what can be verified by the courts. A mechanism is chosen before the sender observes the true state (private information) otherwise the choice of mechanism itself may convey some information. For example, if in the *screening game* the receiver commits to his response to a signal it turns into a mechanism game.

8.5 Screening

A *screening game* is closely related to signaling games where rather than choosing an action based on a signal, the receiver gives the sender proposals based on the type of the sender. The sender sends a message in response to an offer by the receiver. For example, the employer offers a wage level first, at which point the worker chooses the amount of credentials he will acquire (education or skills) and accepts or rejects a contract for a wage level.

8.6 Cheap Talk

Farrell and Rabin [49] introduce cheap talk games. In economics, signaling games have an associated cost. However, it is widely believed that most of the information transmission in modern microeconomics is not done through costly signaling systems but through ordinary *cheap talk*. Cheap talk[6] is an incomplete-information game that consists of costless, non-binding, non-verifiable messages that may affect the listener's beliefs but the message itself does not directly affect the payoffs of the game. The receiver, after hearing the message from the sender, must take an action, which decides the payoffs for both players. The game proceeds as follows.

1. Nature decides S's type t (the sender's private information)

2. S observes t and sends a message m to R

[6]The peacock's tail is an example of talk which is not cheap. The tail convinces the hens that the peacock is a worthy suitor but the tail imposes cost by taking up resources and making the peacock easier to catch.

3. R does not know the sender's type t but takes an action a based on his prior beliefs about t and S's message m

4. S's type t and R's action a decides the payoffs for both S and R

A *self-signaling* message is such that the speaker says it if and only if it is true i.e. it is to the speaker's benefit to tell the truth. A *self-committing* message creates an incentive for the speaker to fulfill it. A message that is self-signaling and self-committing seems credible. A credible message is *believable* therefore the receiver can base its decision on it.

8.6.1 Cheap Talk About Private Information

Suppose Ann is a job applicant and Bob a potential employer who wants to hire Ann for one of two positions, demanding and undemanding. Bob will give Ann the demanding job if he believes her ability is high and the undemanding job if he believes her ability is low. Bob does not know Ann's ability. Ann sends a message "High" or "Low" to Bob signaling her ability. Bob then decides which job to give Ann. It is quite obvious that Ann has preferences over Bob's beliefs about her ability as Bob relies on those beliefs to take an action. The normal form game for one version of this example is shown in Figure 5.

Ann's types, "High or "Low" are self-signaling and they coincide with her true type. Therefore, Bob can make his job assignment depend on Ann's message. Since Ann has no incentive to lie, cheap talk conveys all of Ann's private information to Bob.

		Bob	
		D	U
Ann	H	**2,1**	$0,0$
	L	$0,0$	**1,3**

Figure 5: Normal form representation of the game where Ann's type high (H) or low (L) is self-signaling and Bob can make his job assignment based on Ann's message. That is hire Ann for the demeaning job (D) if she sends the message "High" or undemanding job (U) if she sends the message "Low".

Consider a modification of the game. Let's assume that the demanding job pays more and Ann is greedy. So Ann has an incentive to lie and get the demanding job regardless of her ability. The game is shown in Figure 6.

		Bob	
		D	U
Ann	H	$\mathbf{2},1$	$0,0$
	L	$\mathbf{2},0$	$1,3$

Figure 6: Normal form representation of the game where Ann's type "High" or "Low" is no longer correlated with Ann's true type. And cheap talk fails to convey Ann's private information to Bob.

Ann's preferences over Bob's beliefs are no longer correlated with Ann's true type and types "High and "Low" are no longer self-signaling. Due to lack of self-signaling and correlation cheap talk fails to convey Ann's private information to Bob.

In these two situations, correlation between the sender's true type and preference over the receiver's beliefs is either perfect or fails completely. The more interesting situation is of course where Ann and Bob's preferences are partially aligned. Can cheap talk be credible in problems where preferences of the sender and the receiver are partially aligned? Crawford and Sobel [30]

argue that not all games are coordination games and many difficulties with reaching agreements are due to players having different information about preferences. Sharing information helps in reaching potential agreements, but it also has a strategic effect that revealing all information to an opponent is not usually the most advantageous strategy. But even a selfish agent will frequently find it beneficial to reveal some information. They showed that limited common interest might lead to meaningful talk.

8.6.2 Crawford and Sobel's Model

In Crawford and Sobel's model, there are two players, the sender S and the receiver R. The sender observes the value of a random variable m (S's private information or type), whose differentiable probability distribution function, $F(m)$, with density $f(m)$, is supported on [0, 1]. The sender has a twice continuously differentiable von Neumann-Morgenstern utility function $U^s(y, m, b)$, where y, a real number, is the action taken by the receiver upon receiving the sender's signal and b is a scalar parameter used to measure how nearly agents' interests are aligned. The receiver's twice continuously differentiable von Neumann-Morgenstem utility function is denoted $U^R(y, m)$.

The assumptions are that for each m and for $i = R,\ S$, denoting partial derivatives by subscripts in the usual way, $\mathrm{U}_1^i(\mathrm{y,m}) = 0$ for some y, and $\mathrm{U}_{11}^i(.)$ < 0, so that U^i has a unique maximum in y for each given *(m,b)* pair; and that $\mathrm{U}_{12}^i(.) > 0$. The latter condition ensures that the best value of y from a fully informed agent's standpoint is a strictly increasing function of the true value of m. All aspects of the game except m are common knowledge.

The game proceeds as follows. The sender observes her *type*, m, and sends a signal to the receiver; the signal may be random, and can be viewed as a noisy estimate of m. The receiver processes the information in the sender's signal and chooses an action, which determines both players' payoffs. In equilibrium, each agent responds optimally to his opponent's strategy choice, taking into account its implications in the light of his probabilistic beliefs, and maximizing expected utility over his possible strategy choices. Formally, an equilibrium consists of a family of signaling rules for S, denoted $q(n|m)$, and an action rule for R, denoted $y(n)$, such that

1. For each $m \in$ [0,1],

$$\int_N q(n|m)\,\mathrm{d}n = 1,$$

where the Borel set N is the set of feasible signals, and if n* is in the support of q(.| m), then n^* solves

$$\max_{n \in N} U^s(y(n), m, b); and$$

2. For each n, $y(n)$ solves

$$\max_{y} \int_0^1 U^R(y, m)p(m|n)\mathrm{d}m, where$$

$p(m|n) \equiv q(n|m)$ f(m) / $\int_0^1 q(n|t)f(t)\mathrm{d}t$.

The first condition says that the sender's signaling rule yields an expected-utility maximizing action for each of his information types, taking the receiver's

action rule as given. The second condition says that the receiver responds optimally to each possible signal, using Bayes' Rule to update his prior, taking into account the sender's signaling strategy and the signal he receives.

Crawford and Sobel characterized the set of equilibrium outcomes and demonstrated that there is a finite upper bound, N*, to the number of distinct actions that the receiver takes with positive probability in equilibrium, and that for each N = 1, ..., N*, there is an equilibrium in which the receiver takes N actions. In addition, when monotonicity condition holds, for all N = 1, ..., N*, there is a unique equilibrium outcome in which the receiver takes N distinct actions with positive probability, and the expected payoffs for both the sender and the receiver are strictly increasing in N. The equilibrium with N* actions is the most informative equilibrium.

Let us look at the job applicant example in light of the Crawford and Sobel's model. Ann's ability lies on a continuum rather than being binary "High" or "Low." Based on Bob's beliefs about her ability, it will set her wage and make workplace demands on her. If Bob believes Ann has high ability, he will demand more work and pay more. Ann knows her ability, but Bob only has his beliefs about Ann's ability and what Ann says. Suppose Ann's type t is uniformly distributed on [0, 1]. Ann sends a message m and Bob chooses an action a, where a and m are also $\in$ [0, 1]. A message m may be a sentence, *"My type is t."* The payoffs are quadratic loss functions in which each player has an ideal point and wants a to be close to the ideal point. Let $U^{Ann} = -(a-(t+b))^2$ and $U^{Bob} = -(a-t)^2$ be the payoff functions for Ann and Bob, respectively.

At the extreme of payoff function similarity, it is clear what happens. Suppose Bob wants a to be as close to t as possible. If Ann also wants a to be close to t, then she will reveal her true type. This is called a *separating equilibrium*. On the other hand, If Ann wants a to be as big as possible then she will lie. The signal will convey no information and Bob will ignore Ann's message. This is called a *pooling equilibrium*.

Let's say Ann wants to persuade Bob that her ability is somewhat higher than it actually is. However, Ann doesn't want to exaggerate too much. The interesting question is what happens if Ann likes Bob's ideal action to be $t + .1$? So Ann doesnt want a to be too big, but she does want a to be bigger than what Bob would choose if he was fully-informed about true state of the world.

Crawford and Sobel showed that there exists a *partially pooling equilibrium* in which Ann truthfully reports her type by reporting t is in the low interval $[0, x]$ or the high interval $[x, 1]$, say $x = .3$. So in effect, Ann reduces her message space to two messages, *Low* and *High*.

Bob's optimal strategy in a partially pooling equilibrium is to choose his action to equal the expected value of the type in the interval the sender has chosen. Thus, if $m = 0$, Bob will choose $a = x/2$ and if $m = 1$, he will choose $a = (x + 1)/2$. Bob's equilibrium response determines Ann's payoffs from her two messages. The payoffs between which she chooses are; $U^{Ann}_{m=0} = -((t + .1) - \frac{x}{2})^2$ and $U^{Ann}_{m=1} = -(\frac{1+x}{2} - (t + .1))^2$.

There exists a value x such that if $t = x$, Ann is indifferent between

$m = 0$ and $m = 1$, but if t is lower he prefers $m = 0$ and if t is higher he prefers $m = 1$. To find x, equate $U_{m=0}^{Ann}$ and $U_{m=1}^{Ann}$ and simplify to obtain $(t + .1) - \frac{x}{2} = \frac{1+x}{2} - (t + .1)$. We set $t = x$ at the point of indifference, and solving for x then yields $x = .3$.

Thus, the divergence in preferences of the sender and the receiver coarsens the message space. Ann will not send a truthful precise message, but if there is a partially pooling equilibrium, she will send a truthful coarse message. If the true value of t is small, Ann will report the fairly precise information that t lies in [0,.3]. If t is larger, it is harder to induce a truthful message, since Ann has a tendency to exaggerate and report t larger than it is.

If instead of wanting (t + .1) to be the action, the preferences of Ann and Bob diverge more e.g (t + .8), then there would be the uninformative pooling equilibrium. If they diverged less e.g (t + 0.001), then there would exist other partially pooling equilibria that had more than just two effective messages and would distinguish between three or more intervals instead of between just two.

8.6.3 Cheap Talk Equilibria

Every cheap talk game has a *babbling equilibrium* where the sender's message does not affect the receiver's beliefs and the receiver ignores the sender's message. The sender might as well make noises that are not related with her type. In turn, Ann's babbling justifies Bob's strategy of ignoring her message

and assigning the undemanding job, which is his best move given an expected value of 50/50. It is always consistent with rationality to treat cheap talk as meaningless. Farrell and Rabin [49] argue that people don't usually take a destructive attitude, "*I won't presume words don't mean what they have always meant.*" Rather people take the literal meaning as a starting point. The view that cheap talk may be blocked by incredulity but not by incomprehension is called the *rich language assumption*. It assumes that players share a common language and are competent to work out the literal meaning of sentences.

8.6.4 Cheap Talk about Intentions

Can cheap talk be effective in coordination problems? Farrell and Rabin [49] argue that if a message is credible, then cheap talk efficiently resolves coordination problems.

Suppose Bob hired Ann and now they work together. Ann and Bob are planning to have lunch together. Ann leaves the office before Bob who will join her later. Ann says to Bob, *"I'm off to Eatery 2."*

		Bob			
		*Eatery*1	*Eatery*2	*Eatery*3	*Eatery*4
Ann	*Eatery*1	$3,3$	$0,0$	$0,0$	$0,-2$
	*Eatery*2	$0,0$	**3,3**	$0,0$	$0,-2$
	*Eatery*3	$0,0$	$0,0$	$3,3$	$0,-2$
	*Eatery*4	$-2,0$	$-2,0$	$-2,0$	$1,1$

Figure 7: A two-player coordination game.

Here, Ann's message is self-signaling and self-committing and thus credible. If Ann's message is credible then Bob believes it and his best response

will be going to Eatery 2, a Nash equilibrium in this game.

8.6.5 Cheap Talk vs. Conventions

Let's assume Bob did not hear Ann's message about what Eatery she is heading to but knows Ann left for lunch and she is waiting for him. Can Schelling's *focal point* help them coordinate?

Schelling's focal point is best explained with following example, *"Two people planned to meet in New York but forgot to say where. The leading focal point at the time was Grand Central Station. In this situation, each can infer the other person would pick Grand Central Station as it is the natural focal point."*

Suppose Ann and Bob go out for lunch quite often and they have a favorite place where they usually eat. Then Bob can infer where Ann may be waiting for him and try his luck but it is not guaranteed he will find her there. Perhaps, Ann wanted to try another place that day. Going to the usual place is better than no coordination but worse than what they can get by talking.

8.6.6 Coordination Under Conflict

Suppose Ann and Bob are working on a joint project where each prefers the other to do more work. They both reason whether the other player uses high effort or low effort while her/his own best response is low if the other uses high effort. This leads to Nash equilibrium (6, 6), which is Pareto-dominated by (7, 7), where both offer high effort. Can Ann and Bob talk their way out of

this?

		Bob	
		H	L
Ann	H	$7, 7$	$5, 8$
	L	$8, 5$	**6,6**

Figure 8: Strategic representation of a two-player game where Ann and Bob chooses whether to put high (H) or low (L) effort in their joint project.

If Ann says, *"I will put in high effort and I expect you to do the same,"* the message is not self-signaling as Ann likes Bob to put in high effort whatever she plans to do. And it is not self-committing as Ann has no incentive to follow through on her promise. Even if Bob believes Ann's plan to put in high effort, he will have no incentive to put in high effort himself. Whatever they say, low effort remains a strictly dominant strategy. If there is conflict, messages are less likely to be self-signaling or self-committing and cheap talk will be less successful or less informative.

8.6.7 Conflict in Talk

Suppose Ann and Bob have become good friends and would like to spend an evening together. Both would rather spend the evening together than apart. Bob would like them be together at the prizefight, while Ann would like them be together at the opera, and both players can talk in the game. Can they reach an agreement through cheap talk?

If Ann says, *"I'm going to the opera,"* and Bob says, *"I'm going to the opera,"* these messages are self-signaling and self-committing and they

		Bob	
		O	F
Ann	O	$2, 1$	$0, 0$
	F	$0, 0$	$1, 2$

Figure 9: Strategic representation of a two-player game where Ann and Bob choose between opera (O) and prizefight (F).

reinforce each other. It is likely they will continue as in the game of pure coordination. However, if Ann says, *"I'm going to the opera,"* while Bob says, *"I'm going to the fight,"* each message individually is self-signaling and self-enforcing but they conflict. Unless it's common knowledge between them who's in charge, they can't coordinate.

9 Game Theory and Pragmatics

Communication is a goal-oriented activity where interlocutors use language as a means to achieve an end while taking into account the goals and plans of others. Game theory, being the scientific study of strategically interactive decision making, provides the mathematical tools for modeling language use among rational decision makers. In a game, there are at least two players who interact with each other and it results in a certain outcome. Each player has a choice between various courses of action, their strategies. Each player has a preference ordering over expected outcomes. Preferences are usually encoded as numerical values called utilities or payoffs assigned to possible outcomes. One of the objectives of game theory is to derive insights into how rational players ought to behave in a strategic situation. A rational player is said to hold some consistent beliefs about the structure of the game and the strategies of other players, and they will choose their strategy in such a way that their expected utility is maximized. Also, rational players are assumed to be logically omniscient i.e. they take all logical consequences of their beliefs into account in their decisions. It is common knowledge among the players that all players are rational in this sense.

Application of game theory to communication has recently attracted attention for two reasons. First, communication between players may affect the outcome of the game. Second, communication itself can be analyzed as a game.

9.1 Equilibrium Semantics

Prashant Parikh [94] present a new account for language meaning called *equilibrium semantics*. In a game, an equilibrium means balance amongst multiple interacting elements. Parikh argues that for language equilibrium enters through the element of *choice*. The speaker must choose his utterance and the addressee must choose her interpretation and these choices must be in equilibrium for information exchange to take place. A speaker and addressee participate in multiple games at multiple levels in a single utterance so there are multiple equilibria that occur in communication. Thus, not only does the equilibrium involve a balance amongst the choices and strategies available to the speaker and addressee in each game, but also the multiple equilibria are themselves in balance, an equilibrium of equilibria. Situated games of partial information can be used as a mathematical framework to model language games taking into account *choice* and *strategic interaction* as fundamental properties of linguistic and communication systems.

There are the following sets of constraints called *SCIF*: Syntactic (S), Conventional (C), Informational (I), and Flow (F). S is some account of syntax of the language being considered that interacts and is influenced by meaning and plays a role in derivation of content. C is a set of conventional constraints that maps every word into one or more properties or relations - these can be extracted from a dictionary and is independent of context. I maps the properties and relations obtained from the conventional map into certain special situation-theoretic objects determined by S and part of the information space

or ontology relative to a context or utterance situation u. This map is called the informational map and S influences its behavior. Finally, F embodies much of the equilibrium semantics. It is essentially a system of situated games provided a model of utterance situation u, so that together with the sentence and its phrase structure, one can infer its meaning. Equilibrium in semantics is defined in terms of these four sets of constraints being in equilibrium within each constraint and across constraints, both in the context of the system of meaning and in the context of utterance.

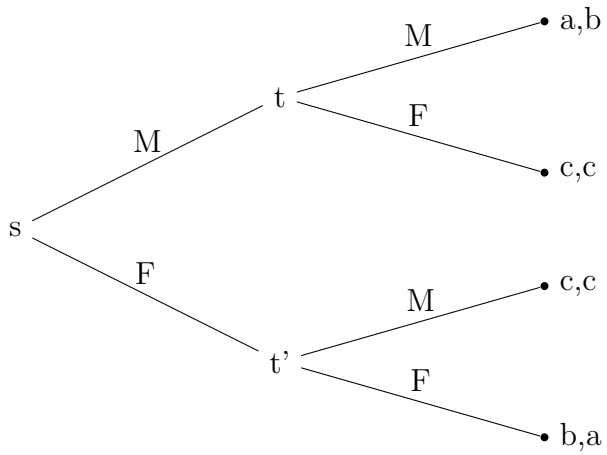


Figure 10: Battle of the sexes game.

Going back to the battle of the sexes example, imagine a situation where Ann and Bob are married and sitting in the living room negotiating their plans for the evening i.e. whether to go to a movie or a football match together. The extensive form for this game is shown in Figure 10. Assume the movie *Emma* is playing in the theaters. Ann and Bob's daughter Emma, who

happens to have the same name as the movie, is playing with her siblings in the living room. In this context Ann utters, "*Emma is playing.*" This sentence when spoken is ambiguous. On the intended reading, Ann is noting that the movie "*Emma*" is showing at some theater. On another reading, it could be an observation that their daughter Emma is playing. All Ann and Bob know is the first interpretation is more likely based on the fact that they are discussing what to do this evening. A game can be constructed to model this communication scenario between Ann and Bob.

Let's call this utterance situation u. Since disambiguation is a selection of one meaning from many, we need to lay out all possible meanings of the sentence uttered by Ann. For simplicity, let δ = *Emma* [7], we need to find out what the possible meanings of δ are in situation u. The *Conventional Constraint (C)* is a map from a word ω to one or more conventional meanings P^{ω}. The *Information Constraint (I)* takes properties associated with a word and maps them into contextually appropriate possible meanings. Let σ and σ' stand for the possible meanings of δ in utterance situation u; where σ means film and σ' means Ann and Bob's daughter. The *Flow Constraint (F)* is given with the extensive form game of partial information; s is an initial situation represented by a node that contains the setting u together with Ann's intention to convey σ. In s, Ann can utter the ambiguous word δ = Emma, this action is being represented by the relevant arrow issuing from this situation. If she does indeed utter δ, the resulting situation is t, where Bob has to choose an interpretation of δ in u. Each action corresponding to two

[7] One can also represent the sentence as a conjunction of words, we'll use a single word for simplification.

possible interpretations, σ and σ' are represented by corresponding arrows.

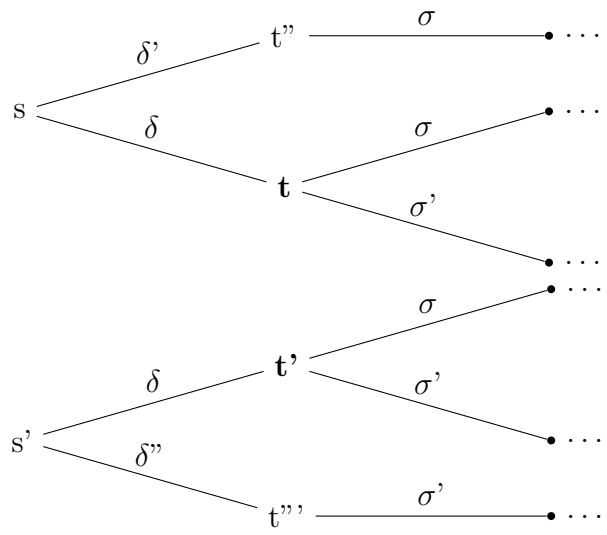


Figure 11: A lexical game.

Since δ is ambiguous, there is an alternative counterfactual situation s' that also contains u, together with the alternative possible intention to convey σ' and $u = s \cap s$'. In s' also Ann can utter δ and this results in t', where Bob again has the same two choices of interpretation. {t, t'} forms an information set for Bob because he is not able to distinguish between the two situations.

δ' and δ" stands for an alternative locution that the speaker might have uttered but chose not to e.g. δ' could be "*The film Emma*" and δ" could be "*Our daughter Emma.*" Since these two are unambiguous, there is just one interpretation that Bob can choose, either σ or σ'.

Additionally probabilities can be assigned to express the likelihood of one utterance meaning over another. Let ρ and ρ' stand for probabilities

that Ann is conveying σ or σ' in s or s'. Since Ann's intention is to convey information about the movie *Emma*, both Ann and Bob can infer that $\rho > \rho$' given the context of the utterance. Finally, the payoffs for Ann and Bob can come from the embedding situation u as well as from the language and can depend on a variety of factors such as their beliefs, desires, hopes, fears, and on the language and it's rules. Thus the payoffs are a complex resultant of positive and negative factors. The payoffs vary greatly between players based on the situation they are in and their varying characters and these assignments at a given situation will decide the Nash equilibria for the game, which will give the intended meaning for the utterance δ in situation u.

9.2 Gricean Meaning and Game Theory

Stalnaker [138] connects Grice's work with game theory using the dynamics of best responses in cheap talk games. Stalnaker defines credibility of messages as follows.

1. A message is prima facie rational (*pf rational*) for player S of type t, if and only if S prefers that R believe the content of the message, S prefers that R believe the message rather than remain in his prior belief state which is assumed to uniformly distributed.

2. The definition of credibility in terms of *pf* rationality is that a message is *credible* if and only if it is *pf* rational for some types, and only for types for which it is true.

3. It is common belief that the content of any credible message that is sent by S is *believed* by R.

4. The structure of the game is common belief, and it is common belief that both players are rational, that they make choices that maximize their expected utility.

		Bob			
		a_1	a_2	a_3	a_4
Ann	t_1	$5, 5$	$10, 10$	$0, 0$	$0, 0$
	t_2	$5, 5$	$0, 0$	$0, 6$	$1, 8$
	t_3	$5, 5$	$0, 0$	$6, 6$	$0, 0$

Figure 12: Normal representation of the game between Ann and Bob where Ann sends a cheap talk message signaling her type t_1, t_2, or t_3 to Bob and Bob takes an action a_1, a_2, a_3, or a_4 based on his beliefs about Ann's type and Ann's message.

In the game shown in Figure 12, if Ann is of type t_2, then her first choice is that Bob get no information and remains in the prior belief state because that would motivate him to choose a_1. But this option is not available since it is clear that the message *"My type is* t_1*"* is a *credible* message that Ann would be *rationally* required to send if and only if she was of type t_1. Therefore, Bob can infer that Ann is not t_1 if he does not get that message. In this case, sending no message would induce the belief that Ann is either t_2 or t_3. And if Bob didn't know which of the two it was, it would result in action a_3, which is a worst outcome for t_2. But if t_2 is able to reveal her type, Bob will instead choose a_4. And if Ann is of type t_2, she would prefer this to a_3. So the message, *"My type is* t_2*,"* is *pf rational* for t_2, since Ann prefers that Bob believe that message to the feasible alternatives to believing it. Since this message is *pf*

rational only for t_2, it is *credible*. The definitions ensure that Ann will reveal her actual type if she is t_1 or t_2, and that Bob will believe her and respond appropriately.

The example shows that sending no message may reveal information, whether the sender wants to reveal it or not. It is also true that sending a credible message may reveal more information than is contained in the explicit content of the message. Sometimes a message that is credible in one model of a given game is not credible in other models of the same game. Let's look at the game shown in Figure 13. Assume that there are just two available messages: *"My type is* t_1*"* or *"My type is not* t_1*."*

		Bob		
		a_1	a_2	a_3
Ann	t_1	$5, 5$	$0, 0$	$0, 0$
	t_2	$5, 5$	$0, 6$	$0, 0$
	t_3	$5, 5$	$6, 6$	$8, 8$

Figure 13: Normal representation of the game between Ann and Bob where Ann sends a cheap talk message signaling her type t_1, t_2, or t_3 to Bob and Bob takes an action a_1, a_2, or a_3 based on his beliefs about Ann's type and Ann's message.

The second message is *pf rational* for t_3, and not for t_1 or t_2. So it is *credible*, but will not be sent by t_2. The first message is not credible, since if Ann is of type t_2, the message would be false, but she might have an incentive to send it, and tempted even more if it is required that one of the two messages be sent. Here we have a case where the meaning of the messages diverges from what the messages literally say. Even though the first message literally means that *"My type is* t_1*,"* it will manifestly express Ann's intention to induce the belief that she is either t_1 or t_2, and will succeed in doing this. It will not

credibly communicate its literal content, but it will credibly convey something weaker. And since it will be mutually recognized that the second message will be sent only by t_3, it will induce the stronger belief that it is manifestly intended to induce, that *"My type is t_3.*

		Bob				
		a_1	a_2	a_3	a_4	a_5
Ann	t_1	$-5, 9$	$-5, 0$	$5, 8$	$0, 3$	$0, 6$
	t_2	$-5, 0$	$-5, 9$	$0, 3$	$5, 8$	$0, 6$

Figure 14: Normal representation of the game between Ann and Bob where Ann sends a cheap talk message signaling her type t_1 or t_2 to Bob and Bob takes an action a_1, a_2, a_3, a_4, or a_5 based on his beliefs about Ann's type and Ann's message.

Assume that Ann is of type t_1 in the game shown in Figure 14. Ideally, Ann would like to convince Bob to choose a_3, giving her a payoff of 5 rather than 0, which is what she would get if she did nothing to change Bob's prior 50/50 beliefs. If she could somehow change Bob's belief to 2/3, rather than 1/3, in the hypothesis that she is of type t_1 , then Bob would make this choice. But what can Ann say to accomplish this? Stalnaker suggests that Ann might try revealing some, but not all, of the evidence that she is of type t_1 , or she might say something that could be taken to be evidence for this, but that might mean something else. She might say something that Bob already knows to be true, but that might give some support to the conjecture that Ann said it because she is of type t_1. But if Bob fully believes she is of type t_1, in which case he would choose a_1, giving Ann a payoff of -5, and given that it is common knowledge that Ann knows whether she is of type t_1 or type t_2, Ann would be taking a risk if she made such an attempt.

Application of Game Theory to linguistics has attracted attention by other researchers. Parikh and Ramanujam [100] present a knowledge based model of communication where meaning of messages are given in terms of how it affects the knowledge of other agents involved in the communication. Jäger, Benz, Rooji [71][91] connects Gricean ideas to game theory and characterize players' moves in terms of their best responses to each other in a game setting. Jäger [71] shows the existence of Nash equilibrium in communication games.

10 Deception in Games

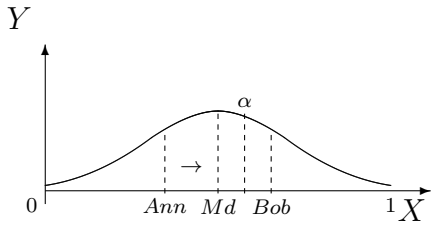
In classical game theory, it is assumed that each person acts selfishly to obtain the highest possible well-being for himself and is unconcerned about the well-being of others. It is widely accepted that we live in a world of deception where people lie in everyday conversations whenever the outcome from lying outweighs the outcome from telling the truth.

10.1 Politics

Consider an *election game* with two phases, the state primaries and the general election. All players in the game belong to one of two parties, Democratic or Republican. In the state primaries, candidates must beat other candidate from their own party to become a nominee and proceed to the general election where they must beat the opposite party nominee to become the president. All players in the game act *rationally* to satisfy their goals. Rationality is defined along the following terms. A rational voter's goal is to pick a candidate closest to its favorite position. A rational candidate's goal is to choose a position that maximizes the total number of votes (s)he receives, considering the voter's rationality.

Brams [21] argues that given a two-candidate game and a distribution of favorite positions, the best position the candidates can choose is the median in the sense that if one candidate is at the median and the other not, then the candidate occupying the median position wins. If both occupy a median

position, then this should result in a tie, although, typically this will not happen as other factors like race, gender, or a candidate's previous record may come in.



Let X represent *voters' favorite positions* and Y the *number of voters*. Then if Ann and Bob are the two presidential candidates, by choosing Md (the median) Ann beats Bob as the area between $[0, \alpha] > [\alpha, 1]$.

The presidential election game illustrate an important phenomenon. In a signaling game, where there is potential for information transmission, the sender can manipulate information without being detected to control the receiver's decision. In this sense, the sender has an invisible power over the receiver's beliefs.

Economists have done experiments to see how lying varies based on different factors.

10.2 Lying Aversion

Gneezy [62] conducts some experiments in order to empirically study the effect of consequences on behavior. He runs three treatments of a two-player experiment where there are only two possible outcomes, A_i and B_i, in each

treatment $i = 1, 2, 3$. The actual choice between the options is made by player two (the receiver) and only player one (the sender) is informed about the monetary consequences of each option. The only information player two has about the payoffs prior to making her choice is the message that player one decides to send. This message could either be *"Option A_i will earn you more money than option B_i"* or *"Option B_i will earn you more money than option A_i."* In all three treatments, option A_i gives a lower monetary payoff to the sender and a higher monetary payoff to the receiver than option B_i and the receiver does not know this. Therefore, if sender sends the second message it can be considered as telling a lie, whereas sending the first message can be considered as telling the truth. The different monetary allocations (in dollars) in the three treatments are as listed below, where a pair (x, y) indicates that the sender would receive x and receiver would receive y.

$A_1 = (5,6)$ and $B_1 = (6,5)$;

$A_2 = (5,15)$ and $B_2 = (6,5)$;

$A_3 = (5,15)$ and $B_3 = (15,5)$.

Gneezy compares people's behavior in two different settings; a *deceptive game* where a person can tell the truth and obtains an allocation that is more equitable and generous to the subject he is matched with, or a *dictator game* where a person lies and obtains a selfish allocation. The reason for setting up a dictator game in addition to a deceptive game is to determine the extent to which the results of the deceptive games reflect an aversion to lying as opposed to preferences over monetary distributions. Gneezy uses the control dictator

game in which player one has the role of dictator and chooses between two the options, while player two has no choice. Again, three treatments were run, corresponding to exactly the same options of the three treatments of the deception games.

Gneezy's results showed that a significant fraction of people display an aversion to lying or deception. The fraction of subjects who chose the selfish allocation in the dictator game were higher than the fraction who made the same choice in the deception game by lying. Whether a sender would lie or tell the truth depends on what beliefs he holds about his partner's responses to his message. His results suggested that people generally expect their recommendations to be followed, i.e. they expect their partner to choose the option that they say will earn the partner more money. In this context, lies are expected to work. Gneezy also showed that people not only care about their own gain from lying, they are sensitive to the harm that lying may cause the other side. Fewer people lied when the monetary loss from lying was higher for their partner, but the monetary gain remained the same for them. Similarly, fewer people lied when their own monetary gain decreased, while the loss for their partner remained the same.

10.3 Social Preferences and Lying Aversion

Harkens and Kartik [69] reinterpret the evidence on deception presented by Gneezy. They present their own hypothesis, *"People are one of two kinds: either a person will never lie, or a person will lie whenever she prefers the*

outcome obtained by lying over the outcome obtained by telling the truth. This implies that so long as lying induces a preferred outcome over truth telling, a person's decision of whether to lie may be completely insensitive to other changes in the induced outcomes, such as exactly how much she monetarily gains relative to how much she hurts an anonymous partner." It is believed to be an important hypothesis to test since if it is right people can be categorized as one of two types: either they are *ethical* and never lie, or they are *economical* and lie whenever they prefer the allocation obtained by lying. Harkens and Kartik claim that conditional on preferring the outcome from lying, a person may be completely insensitive to how much he gains or how much his partner loses from the lie. That is people's social preferences influence whether they actually prefer the outcome from lying relative to truth-telling, independent of any aversion to lying.

In order to test their hypothesis they ran new but similar experiments to Gneezy at the Universitat autonomy de Barcelona in Spain where subjects were college students. They had all subjects play both the deception game and the dictator game unlike Gneezy's experiment where subjects played only one or the other game. And both games were played with same set of monitory payoffs, but each player was matched with a different partner for each. They believed it was important to have a within subject design to directly compare any subject's behavior in the deception game with her preference over allocation as revealed by her choice in the dictator game. They also conducted the experiment using the strategy method for player two in the deceptive game; rather than telling receiver what the message sent by sender is, they asked the

receivers which option they would pick contingent on each of the two possible messages from the sender. This was to directly observe a receiver's strategy. Finally they conducted two different treatments

$A_4 = (4,12)$ and $B_4 = (5,4)$;

$A_5 = (4,5)$ and $B_5 = (12,4)$.

Treatment 4 is similar to Gneezy's treatment 2 in the sense that option B entails a small gain for player one (sender/dictator) and a big loss for player two, relative to option A. Treatment 5 is substantially distinct from any of Gneezy's treatments because option B results in a big gain for player one and only a small loss for player two. If lying induces outcome B whereas telling the truth induces outcome A, as is suggested by Gneezy's data, and if the decision whether to lie or not depends on the relative gains and losses even conditional on preferring the outcome from lying, then one would expect to find that the proportion of lies among the selfish subjects in treatment 5 is significantly higher than in treatment 4.

Their results confirmed that the proportion of selfish subjects in treatment 5 was significantly higher than in treatment 4. It also showed that the proportion of lies in the deceptive game was significantly lower than the proportion of selfish choices. Additionally, they found that subjects in Spain were less willing to follow the recommendations they received. Instead, recommendations were often ignored or even inverted. Senders seem to have been aware of the possibility that lies would often not be believed and not work. Their data did not reject their hypothesis but confirmed Gneezy's results on the

existence of lying aversion.

10.4 Social Ties and Lying Aversion

Chakravarty et al [23] run experiments to see the interaction between social ties and deceptive behavior with a modified sender and receiver game in which a sender obtains a private signal regarding the value of a state variable and sends a message related to the value of this state variable to the receiver. The sender is allowed to be truthful or to lie. The receiver can take no action, which eliminates strategic deception. Additionally, subjects (senders) are not restricted to choose between truth telling and a unique type of lies but are allowed to choose from a distinct set of allocations that embodies a multi-dimensional set of potential lies. They implement two treatments: one in which players are anonymous to each other (strangers); and one in which players know each other from outside the experimental laboratory (friends). They find that individuals are less likely to lie to friends than to strangers; and that they have different degrees of lying aversion and that they lie according to their social preferences.

Aoki et al [8] studies the effect of anonymous vs. non-anonymous interaction. They investigate lying behavior and the behavior of people who are deceived by using a deception game in both anonymity and face-to-face treatments. Subjects consist of students and non-students to investigate whether lying behavior is depended on socioeconomic backgrounds. To explore how liars feel about lying, they give senders a chance to confess their behavior to

their counter partner for the guilty aversion of lying. Their results showed that the frequency of lying behavior for students was higher than that for non-students at a payoff in the anonymity treatment, but that was no significant difference between the anonymity and face-to-face treatments. Lying behavior was not influenced by gender. Frequency of confession was higher in the face-to-face treatment than in the anonymity treatment. And the receivers who were deceived were more likely to believe a sender's message to be true in the anonymity treatment.

11 Research Questions

The models of communication reviewed in the forgoing are based on the classical interpretation of game theory, which make strong assumptions about the rationality of players. First, it is assumed that every player is logically omniscient i.e. they know all logical theorems and all logical consequences of their non-logical beliefs. Second, they are assumed to always act in their self-interest maximizing utility. Third, for the concept of Nash equilibrium to work, it is assumed that the form of the game is common knowledge between players. Each player relies on the rationality of others without doubt and relies on other players relying on her rationality and so on. The models are often oversimplified and fail to adequately describe real-world communication dynamics. Valid questions arise as to whether these assumptions are realistic and the models reasonable.

11.1 Rationality Assumptions

Traditionally, reasoning has been thought of as conforming to rules and accepted procedures. How well someone engages in reasoning has been viewed as a major factor in the extent to which the person is rational. Psychologists have attempted, in a number of different experiments [6], to determine whether or not people are capable of rational thought. In a majority of these experiments subjects made inferences that did not logically follow from the premises. For example, when subjects were told to assume, *"All A are B,"* and then asked wither it followed that *"All B are A"* must be true, false, or could be either.

A majority of subjects did not approximate a classical interpretation of the quantifiers. In similar studies, subjects concluded, *"Some A are not B"* from the premise *"Some A are B"* ([6], p. 230).

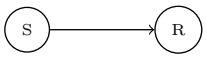
The premise that human beings are rational-utility maximizing individuals is subject to significant qualification as well. Fukuyama ([56], p. 19) explains that the most basic definition of utility is a narrow one associated with Jeremy Bentham (1748 - 1832) who defines utility as the pursuit of pleasure and avoidance of pain. People want to be able to consume the largest quantity of the good things in life. However, there are numerous occasions when people pursue other goals than utilities. They have been known to run into burning houses to save others, die in the battle, or throw away careers so that they can commune with nature somewhere in the mountains. People dont just think utility but also have ideas that certain things are just and unjust, and their choices follow accordingly.

Common knowledge is another rigid assumption that is up for question. Common knowledge is different from mutual knowledge. Mutual knowledge of a proposition α between two players is when each player knows α, whereas common knowledge between two players of a proposition α is equivalent to two infinite chains of knowledge of α; all know that they know that they know α, and so on ad infinitum. With finite memory and processing capabilities, human beings do not go beyond 2-3 levels.

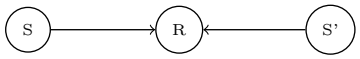
11.2 Oversimplified Model

A large proportion of the literature models signaling between two players; the sender and the receiver. This oversimplification has restricted our view to a narrow one. In a two-player game, the number of states, acts, and signals are often assumed to be the same. There are obviously other possibilities such as extra signals, or too few signals, or not enough acts. All these possibilities raise interesting questions. Needless to say that even adding a third player who is an audience to a two-player signaling game changes the game dynamics.

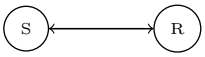
Skyrms[134] argues that there are other possible cases for a signaling game. In the simplest possible case, one sender sends signals to one receiver.



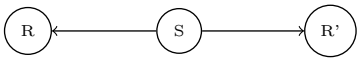
Another simple topology involves multiple senders and one receiver. For example, two senders may observe different partitions of the possible states and sends signals to one receiver.



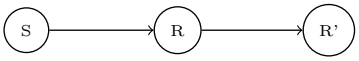
Suppose nature flips a coin and presents the receiver with one or another decision problem. The receiver sends one of two signals to sender. The sender selects one of two partitions of the state of nature to observe. Nature flips a coin and presents the sender with the true state. The sender sends one of two signals to the receiver. The receiver chooses one of two acts. Here a question and answer signaling system can guarantee that the receiver always does the right thing. This is a case where information flows in both directions.



A sender may send information to several receivers. Consider the case, where a third individual is eavesdropping. In a more demanding setup, the sender sends separate signals to multiple receivers who then have to perform complementary acts for everyone to get paid. For instance, each receiver must choose one of two acts, and the sender observes one of four states of nature and sends one of two signals to each receiver. Each combination of acts pays off in exactly one state.



Senders may form chains where they pass information from one to the next. In one scenario, the first individual observes the state and signals the state, and the second observes the signal and signals the third, which must perform the right act to ensure a common payoff.



There is no requirement that the second individual sends a message that has the same content as the original signal that she received. In this case, the second sender might function as a translator from one signaling system to another.

11.3 Avoiding Difficult Problems

Deception is a real problem in signaling games. Deception is when the sender systematically manipulates the signal to his benefit or to the detriment of the receiver. It is different from misinformation where a signal is sent as a result of a mistake.

In any game where there is potential for information transmission, a sender can manipulate information without being detected to control a receivers decision for his own selfish interest. A pure utility-based approach to signaling assumes players will always act selfishly and deceive others as long as they get a higher payoff. However, empirical studies show that people don't often take a deceptive attitude [62] and they seem to show an aversion to lying. The current models of signaling do not account for these results.

To be able to describe the dynamics of information exchange in signaling games, we need to model communication in a less idealized way. This thesis focuses on the topology where the sender sends information to the receiver in the presence of an audience. In our work, we will address some of the questions raised in this section.

12 Hypothesis Development

A few years ago, a colleague struggling to debug his Java code approached me for help. He was in the middle of describing the bug when a manager passed by, at which point, he shifted the topic of his conversation to techniques for optimizing SQL queries.

My colleague was killing two birds with one stone, seeking my help in fixing his buggy Java code while implying his SQL expertise to the manager. Here the manager was indirectly involved in the conversation and whether he can be considered an eavesdropper depends on his intentions.

In an explicit case of eavesdropping, an audience secretly listens to the private conversation of others without their consent. Eavesdropping is not limited to the traditional communication methods but also other forms of communication such as telephone, email, instant messaging, etc. that are considered private.

12.1 Virtual Communication

For better or worse, the Internet and social media have changed communication forever. E-mails, texts, blogs, Facebook, Twitter, etc. have made virtual communication today's reality. Virtual communication has opened the door to billions of people creating, replicating, and sharing information every second, minute, hour of the day across the entire globe; a self-reinforcing cycle that is leading to a tsunami of bytes submerging our world.

12.1.1 Social Networks

There was a time when people would say, "*you are what you eat*," but nowadays, "*you are who you know online*." In this new era, we have taken our social lives online. Social networking sites such as Facebook, Twitter, and LinkedIn (to name a few) have crossed borders enabling people to create online communities.

Facebook, an online social networking site, connects people with friends and others who work, study, or live around them. People use Facebook to share photos, videos, and keep up with each other. As of December 31^{st} 2011, Facebook reported 845 million active users worldwide, more than 100 billion friend connections, 250 million photos uploaded per day, 2.7 billion likes and comments per day, and a revenue of $3.71 billion in 2011, up from $1.97 billion in 2010. On February 1^{st} 2012, Facebook filed for a $5 billion initial public offering [4]. Facebook's future looks promising as more and more people join the network. Facebook shareholder and portfolio manager of Firsthand Technology Value Fund, Kevin Landis, told The New York Times [45], "*Facebook will have more traffic than anyone else, and they'll have more data than anyone else.*"

As of June 2011, Twitter reported over 300 million users [145] and revenue of $140 million in 2010 [85]. TechCrunch has projected that at the end of 2013; Twitter will have 1 billion users, $1.54 billion in revenue, 5,200 employees and $111 million in net earnings [9]. Twitter has evolved the way we use language, people communicating at each other rather than communicating with

each other. Twitter is not considered to embody two-way discussions. However, one thing is for sure that it has made information flow faster than ever. In the Twitter echo system, it takes few users with large number of followers to share something and with a click those followers share the information with their followers and it is all over the Internet. Twitter has been used as a tool in citizen uprisings and fundraising efforts for crisis situations.

When people post millions of tweets every day, within that stream are some valuable pieces of information that can only come together when we act in aggregate. A recent study [137] by medical researchers at Harvard showed that Twitter was substantially faster at tracking the spread of cholera in Haiti following the earthquake in 2010 than any traditional diagnostic method. By using information from Twitter, researchers were able to pinpoint outbreaks of the deadly disease more than two weeks before they were identified by traditional methods. At the other extreme, Twitter is the ground where Internet hokum can grow beyond prevention. The spread of apparently impressive or legitimate but actually untrue and nonsense content cannot be overlooked.

LinkedIn, a business-related social networking site, reported 150 million users in 2010 and the site's revenue as $243 million [2]. The site has focused on providing professionals with a means to manage their professional identity, engage with their professional network, and access insights into opportunities. LinkedIn has taken professional networks online, providing access to people, jobs, and opportunities.

12.1.2 The Inevitable Audience

These social networking services do not charge a fee and neither do they produce anything. They just enable you to *communicate* with each other and form social networks similar to what you have in real life. By doing so, you *signal* information, whether explicitly or implicitly, about who you are, what you like, who are in your social circle, and the cycle continues through others replicating, distributing, and creating feedback. The theory pioneered by American economist Paul Samuelson, called *revealed preferences*, says that one can know what is useful by what people reveal to be useful by their choices.

Though not very obvious to an individual, the aggregate has created great opportunity for certain groups. The New York Times reports, "*According to Facebook, in December 2011, an advertiser could reach an estimated audience of more than 65 million United States users in a typical day on Facebook, compared with American Idol reaching an audience of 29 million people with its 2011 season finale.*" [3]

Have you noticed the same advertisement pop up on different websites you visit? An advertisement tailored to your preferences timed so well that you couldn't possibly resist. Serendipity? No. There are a number of technology companies collecting data on people's online activity. The process starts with dropping *cookies* in user's browser, segmenting users based on collected data, and serving them customized ads.

Micro targeting is not unique to commercial advertising. On February 21^{st} 2012, The New York Times [5] reported that political campaigns are now

building customized ads based on online data. The campaigns are aiming ads at potential supporters based on where they live, the websites they visit, and their voting records.

> *In recent primaries, two kinds of Republican voters have been seeing two different Mitt Romney video ads pop up on local and national news websites. The first, called "It's Time to Return American Optimism," showed the candidate on the campaign trail explaining how this was an election "to save the soul of America." It was aimed at committed party members to encourage a large turnout. The second video ad, geared toward voters who have not yet aligned themselves with a candidate, focused more on Mr. Romney as a family man. Versions of the two ads were seen online in Florida, Iowa, New Hampshire and South Carolina.*

This type of micro targeting uses the same techniques that commercial advertisers use for customized ads. For example, serving up hotel ads to people who had shopped for vacations recently. Except, here it allows campaigns to put specific messages in front of specific voters. "*Two people in the same house could get different messages,*" Mr. Moffatt said. "*Not only will the message change, the type of content will change.* [5]

Social networks are used by some lenders to evaluate loan requests [73]. The New York Observer reports that, a new wave of startups is working on algorithms gathering data for banks from the web of associations on the Internet known as "*the social graph*," in which people are "*nodes*" connected to each

other by "*edges*." From the perspective of the banks, "*birds of a feather flock together*." If your friends are upstanding citizens who pay off their loans, you will be, too. And vice versa, if you're responsible then your friends are too and they can be approached as potential clients. Although these algorithms maybe a few years away from being used by major banks, smaller institutions such as micro lender Lenddo are already using an algorithm based on input from a person's various social networks. When you register with a bank using such a system, you would be required to verify your network logins, for example, Facebook. Information from your accounts would be fed into the algorithm and the bank would make a decision based on your online data. Information that will contribute to systematic discrimination where certain segments of the population could be refused loans or charged higher rates based on racial, religious, sexual, or other prejudice.

"*A picture is worth a thousand words*." Employers review publicly available social network profiles to learn more about job candidates. But often users set their profiles private making it available only to people in their network. Recently, a few companies and government agencies have gone beyond just glancing at your social networking profiles. They have asked for username and password to gain full access to the individual's digital history. Employers have asked job candidates to provide their Facebook login information during interviews. "*Bassett, a New York City statistician, had just finished answering a few character questions when the interviewer turned to her computer to search for his Facebook page. But she couldn't see his private profile. She turned back and asked him to hand over his login information.*" In another

instance: "*When Collins returned from a leave of absence from his job as a security guard with the Maryland Department of Public Safety and Correctional Services in 2010, he was asked for his Facebook login and password during a reinstatement interview, purportedly so the agency could check for any gang affiliations.*" [150] Sears is one of the companies offering an opportunity for future jobs by letting applicants logging into the Sears job site through Facebook. This allows Sears to draw information from candidate's profile, such as friend lists. You may opt out of such an interview or process but "*if you need to put food on the table for your three kids, you can't afford to stand up for your belief.*" [150]

On June 5^{th} 2013, Edward Snowden, a technical specialist who has contracted for the NSA and works for the consulting firm Booz Allen Hamilton, revealed large-scale surveillance of Internet user data by the National Security Agency, in a program known as PRISM [59]. The New York times [147] claims that some of the leading companies, including Microsoft, Google, Yahoo, Apple and Facebook, apparently made it easier for the National Security Agency to gain access to their data. On June 9^{th}, 2013, Jameel Jaffer writes in The New York Times [70].

> *"The Guardian revealed that the government has directed Verizon Business Network Services to hand over an array of sensitive information about every domestic and international phone call made by its customers in the United States over a three-month period. The directive, sanctioned by the secretive court that oversees government surveillance in some national security cases, requires Ver-*

izon to tell the government who made each call, whom they called, when they made the call, how long the call lasted, and (maybe) where the parties to the call were located ... As if that weren't enough, The Guardian and The Washington Post also revealed last week that the N.S.A. has secured direct access to the major Internet companies' central servers. There seems to be some confusion about precisely what the N.S.A. is doing with that access, but The Washington Post reports that the agency is collecting information about surveillance targets believed (with 51 percent certainty) to be outside the United States and about people one and two degrees removed from these targets. So the N.S.A. might focus initially on, say, a British journalist working at Der Spiegel, collecting all of her e-mail communications as well as all uploaded videos, photos, Web surfing data, social media posts and then collect the same information about all of the contacts in the journalist's address book and then about all of the contacts in their address books."

Mr. Jaffer argues that congress should have limited the NSA's authority to monitor the communications of innocent people.

12.1.3 Critical Mass

The virtualization of communication has led to an interesting phenomenon, which is best explained with Schelling's [128] ant colony example.

It is not believed that any ant in any ant colony knows how the

ant colony works. Each ant has certain things that it does, in coordinated association with other ants. But no single ant designed the system and no individual ant knows whether there are too few or too many ants exploring for food or rebuilding after a thunderstorm or helping to carry in the carcass of a beetle. Each ant lives in its immediate environment and responding to signals of which it does not know the origin.([128], p. 21-22)

Why are millions of people joining the pool? It is well established that individuals are increasingly influenced by the opinions of others. We're more likely to do things when we know our friends approve of. Schelling [128] calls this *critical mass*.

You sometimes double park if it looks as though everyone else is double parked, stay inline if everyone else is staying inline, but if people surge toward the ticket window, you are alert to do the same. What is common to all these situations is the way people's behavior depends on how many are behaving a particular way. ([128], p. 93)

To the individual engaging in virtual communication, the effect of building a relationship can be as meaningful as building one in the real life. Virtual reality has psychological effects on people. In a different context, Adler and Satari [7] examine the effectiveness of virtual reality simulations to the treatment of phobias and anxiety disorders. Research shows that virtual reality can be as powerful as traditional treatment methods, where patients suffering from

fear of flying, fear of driving, acrophobia, social phobia, and eating disorders showed long term improvements after treatment.

Wood and Smith [157] quote Ellen Ullman, who on an occasion, found herself up one night and decided to send a text message to a colleague. After reading her message, he wrote back to inquire why she was up so late. The two exchanged cordial messages, but the next day at the office, Ullman was unsure about how to approach him. They had been friendly with one another on the Internet, yet in the office, she felt tension. Ullman questioned, "*In what way am I permitted to know him? And which set of us is the more real: the sleepless ones online, or these bodies in the daylight*?"

Ullman's experience was before the social networks era. Today we are moving towards a society where virtual personas are weighted higher than anything else. I worked at a company that went through a merger. We had a visitor Y from the new company for an all day meeting. During the lunch break, Y talked about his co-workers and in particular, X. X was described not by his professional expertise but how he didn't fit in this world. *"X is the most weird guy you'll ever meet.*" Why? Apparently, X did not have cell phone, never used IM, etc. The mocking went on for a while when someone in the room suggested looking up X on Google. All eyes on the projector and VOILA! X is on Facebook! "*He is on Facebook, ah so he is not as weird as everyone thought he was,*" said Y.

"*I am on Facebook, therefore I am.*" I wonder what Rene Descartes might have to say about this!

12.1.4 The Fourth Revolution

Floridi [53] argues that we are in the middle of a fourth revolution, an *information* revolution. There have been three scientific revolutions before this, which have had great impact on changing our understanding of the external world and ourselves.

Nicolaus Copernicus (1473-1543) theory of the *heliocentric cosmology* displaced the Earth and humanity from the center of the universe. Charles Darwin (1809-1882) with his theory of *evolution* showed that all species of life have evolved over time from common ancestors through natural selection, therefore displacing humanity from the center of the biological kingdom. Sigmund Freud (1856-1939), acknowledged that the mind is also *unconscious* and subject to the defense mechanism of *repression*. Thus we are not immobile, at the center of the universe, we are not unnaturally separate and diverse from the rest of the animal kingdom, and we are very far from being standalone minds entirely transparent to ourselves.

The credit for the fourth revolution goes to Alan Turing (1912-1954). Since 1950s, computers have had an influence on changing not only our interactions with the world but also our self-understanding. Floridi argues that we are no longer standalone entities, but rather interconnected informational organisms that he calls *inforgs*, sharing with biological agents and engineered artifacts a global environment ultimately made of *information*, called the *infosphere*.

12.2 Relationships and Trust in Communication

Joan Silk [149] argues that we cooperate because it contributes to a public good.

> ... altruistic social preferences are a precondition for the kinds of effective collaborative that humans are so good at. It makes it look as if our joint endeavors are mutualistic stag hunts, when in fact we are often in situations in which our own interests and the interests of the group are imperfectly aligned. I don't give to public radio because my $50 contribution is necessary in order for me to listen to it. I give to public radio because I feel that it is the right thing to do because it contributes to a public good.

It is well established by anthropologists that by nature human beings are communal. Tomasello and his colleagues [148] describe in what way humans are considered more intelligent than other animals. They ran an array of cognitive tests to adult chimpanzees and orangutans (two of our closest primate relatives) and to two years old human children. As it turned out, the children were not more skillful overall. They performed about the same as the apes on the tests that measured how well they understood the physical world of space, quantities and causality. However, the children performed better only on tests that measured social skills such as social learning, communicating, and reading the intentions of others.

Language plays an important role in building relationships. Pinker [105] argues that relationships are defined by language. There are essentially three human relationships across cultures, as proposed by anthropologist Alan Fiske; *dominance*, *communality*, and *reciprocity*. Dominance is the type of relationships with some sort of top down hierarchy. Communality is the type of relationships that involve kinship and mutualism. Reciprocity is the type of relationships that involve business like exchanges.

Behavior that is acceptable in one relationship type can be anomalous in another. For example, there can be awkward moments in workplace when an employee doesn't know whether to address their supervisor as by their first name or to invite them for a drink after work. Pinker says that this is because of the ambiguity whether their relationship is governed by dominance or communality.

Two kinds of communal relationships of friendship and sex give rise to the anxiety of dating. Say Bob wants to invite Ann to his place after a date, he uses indirect speech, "*Would you like to come up and see my etchings?*" instead of a more direct one such as, *"Would you like to come up and have sex?"* Pinker argues that an obvious indirect message merely provide individual knowledge where as direct speech provides mutual knowledge and relationships are maintained or nullified by mutual knowledge of the relationship types.

So when Bob says, "*would you like to come up and see my etchings*" and Ann says "*no*," then Ann knows that she turned down a sexual overture and Bob knows that she turned down a sexual overture but does Ann know that Bob knows? Ann could be thinking maybe Bob thinks that she is naive.

Does Bob know that Ann knows that he knows? He could be thinking that maybe Ann is thinking that he is dense. Since, there is no common knowledge, they can maintain the friction of friendship.

Ferrazi [50] provides insight into how building relationships can help individuals reach their full potential. He claims that the path to both personal and professional success is through creating an inner circle of trusted people that he calls "*lifeline relationships*." These are relationships that can offer you feedback, encouragement, and mutual support to help discover a more successful individual in you.

Garfield [57] considers the element of relationship in digital marketing. During the Vancouver Olympics in 2010, in order to promote their same old antiperspirant/deodorant products, Secret started a movement with *"Let Her Jump."* They wanted to get women ski jumping into the Olympics. *"We believe in the equality of the genders and that all people should be able to pursue their goals without fear.*" It was extremely successful. The *Let Her Jump* video was viewed more than 700,000 times. Among the viewers, 57 percent reported their impression of the brand had improved and 85% reported the brand helped them feel more confident. They also saw an increase in purchase intent for women by 11% and teens by 33% from Facebook fans and 50% jump for those who viewed the video. Secret sales increased by 8% despite cutting TV ad spending by 70%.

What is interesting is that the product itself did not change; the company merely sent a new message and positioned itself in relation to millions of women who could potentially announce their affection for the brand. Thus

creating a cycle effect that improved Secret's trust relationship with its consumers.

Oxford dictionary [1] defines the word *trust* as,

> *Trust [noun] firm belief in the reliability, truth, or ability of someone or something; relations have to be built on trust; they have been able to win the trust of the others.*
>
> - *acceptance of the truth of a statement without evidence or investigation: I used only primary sources, taking nothing on trust*
> - *the state of being responsible for someone or something: a man in a position of trust*
> - *[count noun] literary a person or duty for which one has responsibility: rulership is a trust from God*

Trust is the foundation on which all relationships are built. Maintaining trust helps sustain a relationship and violating it leads to friction, which is difficult to repair. Trust starts among family members, expands to friends, and others overtime. Trust is what allows us to have meaningful relationships with one another, within and between all three relationship levels, *dominance*, *communality*, and *reciprocity*. Trust is the glue that holds together social groups such as families, friends, communities, organizations, companies, and nations.

Fukuyama [56] explains that for a country to grow economically, its people must strive for something bigger than self-interest. Trust among individuals is what holds them together in a society and gives rise to middle organizations in between family and government. Fukuyama provides historical references and categorizes China, Italy, France and Korea as low trust societies; Japan, Germany and the United States as high trust societies. In low-trust societies, individuals rely on the extended family to build commercial, social and political networks. The trouble with the extended-family approach to economic development is that all families will soon run out of bloodline managerial, scientific, literary or artistic talent. Countries of High-trust societies form volunteer and meritocratic organizations that expand in scope and efficiency to reach optimum economies of scale. These commercial and non-profit organizations, which are not dependent on family ties, create a network of efficiencies that benefit commerce, media communication and social change.

Trust plays a crucial role in communication; without trust we cannot converse without wondering if other is lying or not. Trust is when we believe the other is telling the truth not based on their message but based on our perception of their character. Trust is knowing the unknown, believing the unseen, giving and receiving without a second thought. Trust means that we can act while taking something for granted. Language helps build and maintain trust and the level of trust in turn decides the integrity of messages. When we trust someone, it means we have no doubt our my mind about his or her integrity. This is when communication channel opens and private information not only gets transmitted but also believed. I may not have the empirical

observation that "*the morning star*" is the same as the "*evening star*" (and we rarely do), but when it comes from a trusted source, we believe it.

Current models of information exchange that are based on game theory are over idealized often limiting the context to two players and making assumptions that are unrealistic in real life. We strongly believe that relationships lie at the heart of communication and *trust* is the heuristic decision rule that allows us to deal with complexities that would require unrealistic effort if we had to decide rationally. It is the heuristic rule that helps us converse with each other. When there is trust, people with opposing preferences can have meaningful talk and share information, even if they don't agree with each other on every single issue. Where there is no trust, communication turns into a transaction, everyone looking out for their own self-interest. This is where deception lives.

12.3 Knowledge in Communication

Communication requires awareness of self knowledge and knowledge of others. In Speech Acts, the choice of what to say depends on knowledge of what the speaker knows about the hearer. For example, *"I have a Porsche, would you like a ride sometime?"* The sentence on the surface acts as an offer while an effect maybe to impress someone. For it to have the intended effect, the speaker has to know that material objects can impress the hearer. Grice's Cooperative Principle and its maxims requires awareness of knowledge. For example, "Make your contribution as informative as is required but not more informa-

tive than is required," depends on an understanding of what the hearer knows already and what information needs to be communicated. Knowledge states directly affects information transmission in communication. It is this awareness of self-knowledge and knowledge of others that enables us to converse with each other in a meaningful way.

In standard theories of language use, the speaker's main purpose is issuing an utterance to get her addressee to recognize her intentions. The question these theories addresses is how does the speaker design her utterances to achieve their goal. The context is limited to communication between two people. However, merely adding an audience to the conversation changes the dynamics. If the speaker does not know that there is an audience, he can continue with the conversation as normal. If the speaker knows an audience is present, he may formulate and execute his utterances differently. Clark et al. [130][27][28][26] argue that there are four attitudes a speaker may take towards an overhearer; namely indifference, disclosure, concealment, and deception. Say Ann is the speaker, Bob is the addressee, and Carl is an overhearer. If Ann is indifferent to Carl understanding what she says to Bob, she can refer to the subject of her conversation by name say *"Derek"* as she normally would with Bob. But if Ann wants to be certain that Carl too can identify Derek, then she may need to expand on or change that reference, "Derek Aitken from Denver". If Ann wants to conceal Derek's identity from Carl, she might say, *"The man we talked about last night."* She may even want to disguise Derek's identity to make Carl think she is referring to someone else.

Our attempt in the foregoing was to provide real-world examples and

point out the urgency of building formal models to study the dynamics of information exchange in more complex settings such as that of information exchange in the presence of an audience. Also, we have argued that notions such as relationship, trust, and knowledge must be considered into building an effective model of signaling.

13 Signaling with an Audience

Grice's work [63] which introduced game-theoretic ideas into reasoning about communication greatly influenced the way philosophers, linguists, and cognitive scientists think about meaning and communication. His work is a foundation of the modern study of pragmatics drawing a clear distinction between speaker meaning, linguistic meaning, and the interrelations between these two phenomena. He examined how in an ordinary conversational situation a speaker, S, shapes his/her utterances to be understood by a hearer H and how both S and H observe some central principles during the talk exchange. His theory of meaning is one that is intention-based, defining linguistic meaning in terms of speaker meaning, "S meant something by U" is roughly equivalent to "S uttered U with the *intention* of inducing a belief in H by means of the recognition of his intention".

At the heart of Grice's theory of meaning lie the *Cooperative Principle* and its special *maxims* of conversation. The Cooperative Principle is a set of *norms* expected in a conversation. It mainly consists of four maxims. The *Quantity* maxim requires that a speaker is as informative as required. It relates to the quantity of information to be provided. The *Quality* maxim requires a speaker to tell the truth provable by adequate evidence. The *Manner* maxim requires the speaker to avoid ambiguity or obscurity, be direct and straightforward. Finally the *Relation* maxim requires a speaker's response to be relevant to topic of discussion.

A good question to ask is, why do people observe the Cooperative

Principle?

Grice assumes that people have learned to do so in childhood. Lewis [83] explains it in terms of social conventions. The work of Lewis emphasizes the existence of social conventions at the heart of which lie language and co-operative problem solving.

For example, the sexton of the Old North Church and Paul Revere must coordinate to warn the countryside of an assault by British army. The sexton knows whether the redcoats are staying home, coming by land, or coming by sea. By placing either zero, one, or two lanterns in the belfry, he signals Paul Revere whether to go home, warn people that redcoats are coming by land, or warn people that the redcoats are coming by sea.

A signaling problem in this sense is a coordination problem, because communicator and his audience must coordinate so that the communicator's signal results in the mutually desired action. Both Grice and Lewis emphasize that participants' interests must be aligned with a common goal and the existence of some sort of mutual understanding in a talk exchange.

But, not all communication is confined to people whose interests are identical. Communication takes place between buyer and seller, or between a suitor and a person of interest, or even between two politicians from nations with opposite interests. Such semi-adversarial communication has been studied by economists and even by philosophers and linguists.

We have argued that the current models of signaling are over-simplified and lack the machinery to explain real-world dynamics of information ex-

change. In particular, a two-player signaling game fails to apply to the problem we have identified that with the emergence of virtual communication we constantly share information in the presence of an audience. Moreover, everyday communication is often automatic and based on our perception of others and heuristic rules we develop over time for sharing information. An effective model of signaling games must account for knowledge, relationships, and ethics in communication.

We extend the ideas of Grice on cooperative communication and the ideas of Crawford, Farrell, Rabin, Sobel, and Stalnakar on communication with partially overlapping interests. In our model, we introduce a third player in the two-player signaling game. The audience may or may not have a move in the game. However, it is clear that the existence of an audience may affect the sender's signal and/or receiver's action depending on the dynamics of the game. Needless to say, the audience may benefit from observing the signal even if he does not make a move in the game. Additionally, we allow for the players to act based on their mental models of how they perceive their relationships with the other players dropping some of the common knowledge assumptions along the way. We distinguish between surface and net utilities to account for the results from empirical studies.

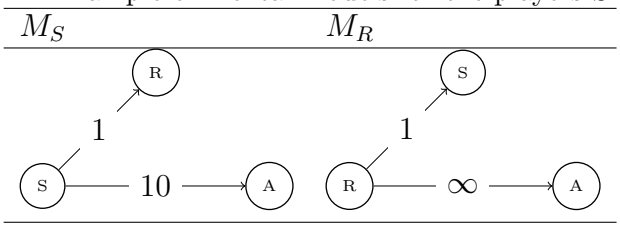
13.1 Abstract Framework

13.1.1 Quantifying Relationships and Trust

We have argued that relationship and trust among individuals play an important role in communication; without trust we cannot converse without wondering whether the other is telling the truth. As we start trusting individuals, we form meaningful relationships; the closer a relationship the more freely we can share information.

One way to formalize these notions is to consider how closely a player perceives himself in relation to the other players. This is a subjective measure i.e. a player perceiving himself close to another player does not necessary mean that it be mutual. We can think of it as players' mental models [8], a diagram of some sort, that gets updated based on experience.

Table 1: Example of mental models for the players S and R.



A Weighted Directed Graph can be used to represents players' mental models. Each player has his/her own mental model[9] in which they have an edge going to other players. The number on the edge is a measure of perceived

[8] Some background material on the theory can be found in Appendix B.

[9] We do not necessarily intend the mental model to be in the head of the agent. We could think of it as part of our representation of the agent.

closeness (or distance) with or trust in another player. A smaller distance means the player perceives himself closely related to the other player. A larger distance means the player doesn't perceive himself closely related to the other player.

Players form their mental models overtime based on their experiences with other players. A naive or trusting player can start off assigning relatively smaller distances on the edges going to other players and as betrayed increment the distances. A calculating player can start off not trusting other players and assign larger distances to edges going to other players but as he starts trusting decrement the distances. Players can form levels of mental models i.e. not only a mental model of one's relationship with others but also a model of others' mental models.

On September 6, 2012, Sen. John Kerry talks about Mitt Romney's stance on political issues at the DNC [151].

> *It isn't fair to say Mitt Romney doesn't have a position on Afghanistan. He has EVERY position! He, he was against setting a date for withdrawal, then he said it was right, and then he left the impression that maybe it was wrong to leave that soon. He said, it was tragic to leave Iraq and then he said it was fine. He said we should've intervened Libya sooner then he ran down the hallway to run away from the reporters who were asking questions. Then he said, the intervention was too aggressive and then he said, the world was a better place because the intervention succeeded; talk about being*

for it before you were against it. Mr. Romney, Mr. Romney, Mr. Romney, here is a little advice.. before you debate Barack Obama on Foreign policy you better finish the debate with yourself.

Sen. Kerry is speaking to the Democratic group of voters (receiver) but he is aware that his speech is airing on TV and the Republican group of voters (audience) is watching. He is taking advantage of the *common ground* he shares with the receiver (i.e. they belong to the same group and thus they have mutual knowledge, beliefs, and assumptions) and not telling the receiver anything new. However, his signal is directed to the audience. He may get cheers and applause from the receiver but the effect of his signal on the audience is more lucrative as it can turn into a potential vote for Obama. Sen. Kerry is sending a signal to the audience about Mitt Romney, attacking Mitt Romney's character accusing him of dishonesty. Why is this so important to his speech? Well, if someone is not trusted with their words, how can they be trusted with decisions concerning a nation. What Sen. Kerry is attempting to do with his signal is to challenge trust that the Republican group (audience) may have in Mitt Romney.

For simplicity, we assume one level of mental models and require players to only consider their subjective measure in the calculation of utilities. Although, in real-world communication, we not only account for those we are directly relate to but also their relationships to people who may be strangers to us. Suppose two friends Ann and Beth are meeting for dinner. The two are extremely close and can talk and laugh for hours. Usually they share everything with each other. Ann arrives at the restaurant and finds Beth and her

mother who has decided to join them for dinner. They spend a good half an hour talking about the menu then food. But there won't be any talk about boys. In this case, Beth's mother is an audience to Ann and Beth's conversation. Ann is close to Beth but not Beth's mother. Technically, there is an edge from Ann to Beth with a small distance but the sum of distances on the edges from Ann to Beth and from Beth to Beth's mother has a larger distance. In this case, Ann has to think twice before sharing any private information.

While relationships are individualistic, notions such as fairness, ethics, etc are social and also play a role in communication.

13.1.2 Surface vs. Net Utilities

We distinguish between two types of utilities; *surface* and *net* utilities. Players' surface utilities are given in the game. Net utilities are subjective and calculated by adjusting surface utilities based on factors important to each player. The resulting game is a transformation of the original game, which may or may not be common knowledge among players.

The idea of discounting utilities based on social distance has been empirically examined by Jones and Rachlin [113]. Their results showed that the amount of money a person was willing to forgo in order to give a sum of money to another person decreased as a hyperbolic function of the perceived social distance between them.

Consider the ultimatum game where Ann is given \$10 to divide between her and Bob. She divides it into u_i for herself and u_j for Bob where $u_i + u_j = 10$.

Here Ann is player i and Bob is player j. If Bob agrees to her division then that is what they both get. If not, neither gets anything. We will discuss how the game is played differently based on net utilities.

The extensive form representation of this game is given in Figure 15.

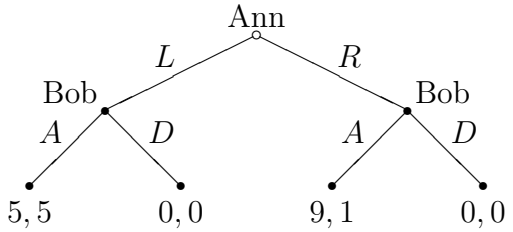


Figure 15: Extensive form representation of an ultimatum game where Ann either offers a fair (L) or unfair (R) proposal and Bob can accept (A) or reject (D).

Under a strictly utilitarian view, if Ann is rational, then doing backward induction, she should give Bob the lowest possible amount and keep the rest for herself. In the example above, Ann may split the ten dollar bill giving Bob a dollar and keeping nine dollars for herself. If Bob is rational, he would accept the dollar as it is better than nothing. There are experimental studies that show that Bob would reject such an unfair allocation. For example, Roth et al. [122] run an experiment comparing related two-person bargaining and multi person market environments in Israel, Japan, US, and Yugoslavia, market outcomes converged to equilibrium everywhere, and there were no payoff-relevant differences among countries. However, bargaining outcomes were everywhere different from the equilibrium predictions and substantial differences were observed among countries due to cultural differences.

Let u_i and u_j be the surface utilities for player i and j, and Δ_{ij} be a

measure of relationship between i and j as perceived by player i [10]; we can define player j's contribution to player i's utility as, $\frac{u_j}{k \times \Delta_{ij}}$ where $k > 0$ is a constant measuring the degree of social discounting or how much one cares about relationships in general; a larger k would describe more selfish or less altruistic choices. For simplification, we'll drop the constant k and assume that it is reflected in the value of Δ_{ij}. The resulting function $g(j, i) = \frac{u_j}{\Delta_{ij}}$ is player j's utility from player i's perspective.

Secondly, we define the fairness correction as c $\times |u_i - u_j|$ where $c > 0$ is a constant measuring how much the player values social norms; here a larger c would mean the player cares more about fairness. While relationship is subjective and depends on individuals, fairness is social and varies by culture. In other words, the amount of fairness created by u_i and u_j depends on the social norms surrounding fairness. In our definition, we have assumed that the society accepts an even allocation.

Let u_i and u_j be the surface utilities for player i and j, $g(j, i)$ be the net utility to player j from player i's perspective i.e. contribution to the other player, and $h(i, j)$ be the social measure for fairness [11]. We define player i's net utility as

$$f(i, j) = u_i + g(j, i) - h(i, j)$$

In the ultimatum game above, Ann and Bob have utilities from L and R. Ann's net utility from choosing L is

[10] Δ_{ij} need not be the same as Δ_{ji} but typically we expect them to be the same or close. If i dislikes j then Δ_{ij} could as well be negative but we will not look into this case.

[11] The payoff to j occurs as a benefit to i and lack of fairness occurs as loss.

$\mathrm{u}^L_{Ann} + \frac{u^L_{Bob}}{\Delta_{AnnBob}} - (c \times |u^L_{Ann} - u^L_{Bob}|)$

Say $\Delta_{AnnBob} = 1$, that is Ann perceives her relationship with Bob to be close. Since the allocation is even, the fairness term will be zero.

Ann's net utility is $5 + \frac{5}{1} - (c \times |5 - 5|) = 10$. So Ann's net utility is 10 instead of face value utility of 5.

Similarly, Ann's net utility from choosing R is

$\mathrm{u}^R_{Ann} + \frac{u^R_{Bob}}{\Delta_{AnnBob}} - (c \times |u^R_{Ann} - u^R_{Bob}|)$

Ann's net utility is $9 + \frac{1}{1} - (c \times |9 - 1|)$. If Ann is not ethical (say c = 0) then her net utility is 10. However, if she is even a little ethical (say c=$\frac{1}{2}$) then her net utility is $9 + 1 - \frac{8}{2} = 8$. Thus Ann is better off choosing L.

Bob does not choose the allocation but has the option to accept or reject Ann's proposal. Let's say Ann has proposed the allocation (9, 1). Let $\Delta_{BobAnn} = 1$ so Bob perceives himself closely related to Ann. Now if Bob puts high value on fairness (say c = 2) then his net utility is

$u^R_{Bob} + \frac{u^R_{Ann}}{\Delta_{BobAnn}} - (c \times |u^R_{Bob} - u^R_{Ann}|)$

Bob's net utility is $1 + \frac{9}{1} - (2 \times |1 - 9|) = 1 + 9 - 16 = -6$. In this case Bob will reject.

13.1.3 Knowledge, Relationships, and Ethics in Signaling Games

In a signaling game, players may not only be interested in the objective reality but also in each others' knowledge. Players can have knowledge of state of the world and each other's knowledge in various ways; the presence of an audience

in turn can be a fact that can lead to different states of knowledge.

- Audience is eavesdropping and neither the sender nor the receiver knows about his presence
- Audience's presence is known to the sender only
- Audience's presence is known to the receiver only
- Audience's presence is known to both the sender and the receiver
- Audience's presence is common knowledge among the sender, the receiver, and the audience

Common knowledge is the highest possible level of knowledge and there could be all kinds of other complex cases depending on levels of knowledge.

Let p be a proposition that says, *"Carl is eavesdropping."* Both the sender and the receiver may know that p is true but neither may know if the other knows or the fact that p is true can be common knowledge between the sender and the receiver but not the audience. Sender's private information can be represented by another proposition. Let q be a proposition, *"It is raining now."* Perhaps the sender happens to just come from outside and knows that q is true but the receiver who is indoors doesn't have access to this information. Other factors such as those of credibility can also be described in this manner. Let r be a proposition, *"The sender never lies."* Then if the sender sends a signal to the receiver informing her that q is true and the receiver knows that r is true, she may decide to take an umbrella when she steps out. It is clear

that there could be complicated knowledge states among players. Formalisms that support such representation and reasoning are called logics of knowledge or epistemic logics[12].

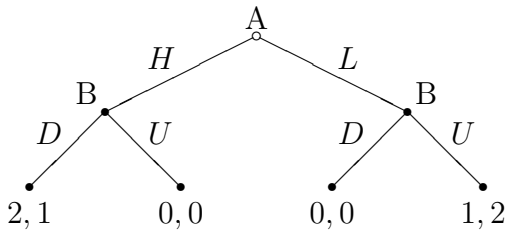


Figure 16: Extensive representation of a two-player signaling game between Ann (A) and Bob (B). Nature chooses Ann's true ability H or L. Ann knows her ability but not Bob. Ann sends one of two messages, "High" or "Low" to signal her ability H or L to Bob who decides whether to hire Ann for the demanding (D) job or the undemanding (U) job.

Suppose Ann is a job applicant and Bob a potential employer who wants to hire Ann for one of two positions; demanding and undemanding. Bob will give Ann the demanding job if he believes Ann's ability is high and the undemanding job if he believes Ann's ability is low. Ann knows her ability but Bob does not. Ann has a choice to send message "High" or "Low" to signal her ability to Bob, who then decides to hire Ann for either the demanding job or the undemanding job. Ann's true ability and Bob's action determine the payoffs for both players. The extensive representation of the game is shown in Figure 16.

Let p be the proposition, *"Ann's ability is low."* We will assume that it is common knowledge between Ann and Bob that Ann knows whether p i.e. $CK_{A,B}(K_A(p) \vee K_A(\neg p))$. In other words, two world connected by an A arrow will never differ in the truth value of p. Also, each player i's action follows

[12]Some background material about epistemic logic is in appendix A.

from formulas beginning with K_i. So for Ann, all formulas starting with K_A.

Case 1: Consider the structure of possible worlds shown in Figure 17.

At w the following are true:

(1) p

(2) $K_A(p)$

(3) $K_B(p)$

(4) $CK_{A,B}(p)$

Figure 17: The content of the world w is $\{p\}$ which is common knowledge between Ann and Bob.

In this case, Ann's ability is low and this fact is common knowledge between Ann and Bob. Since Bob knows Ann's true ability, he will choose undemanding regardless of Ann's message. Therefore, Ann might as well send a truthful message "Low." Ann and Bob's payoffs are 1 and 2 respectively.

Case 2: Consider the structure of possible worlds shown in Figure 18.

At $w1$, the following are true:

(1) $\neg p$

(2) $K_A(\neg p)$

(2) $\neg K_B(\neg p)$

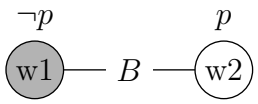


Figure 18: B is Bob's accessibility relation. The content of the worlds $w1$ and $w2$ are $\{\neg p\}$ and $\{p\}$ respectively.

Say $w1$ is the true state of the world. Then it is not the case that Ann's ability is low. Since Ann's ability is high, she has no incentive to lie so she will send the message "High" to signal her true ability. Bob does not know Ann's ability. If Bob is trusting, he may believe Ann's message and hire her for the demanding job. In which case, Ann receives a payoff of 2 and Bob a payoff of 1.

Figure 19 is a modification of the game shown in Figure 16. As before, nature chooses Ann's ability. Ann can send a message "High" or "Low" to signal her ability to Bob. Bob chooses whether to hire Ann for the demanding job, undemanding job, or not hire her. Ann's true ability together with Bob's action decides the payoff for both players.

Let's examine how players' knowledge states may alter the outcome of this game.

Case 3: Consider the structure of possible worlds shown in Figure 20.

At $w2$, the following are true.

(1) p

(2) $K_A(p)$

(3) $\neg K_B(p)$

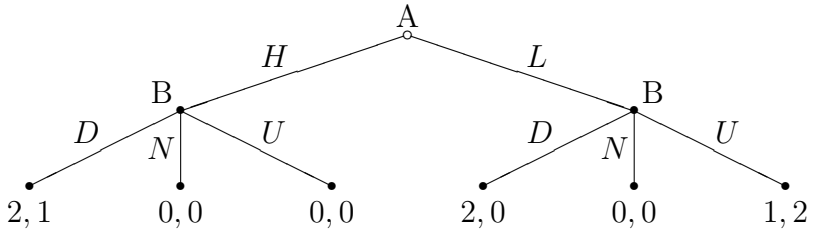


Figure 19: Extensive representation of a two-player signaling game between Ann (A) and Bob (B). Nature chooses Ann's ability. Ann can send one of two messages, "High" or "Low" to signal her ability H or L to Bob, who decides whether to give Ann the demanding (D) job, give Ann the undemanding (U) job, or not hire (N) her. Ann has an incentive to lie.

(4) $K_A(\neg K_B(p))$

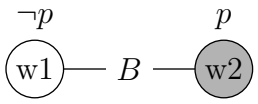


Figure 20: B is Bob's accessibility relation. The content of the worlds $w1$ and $w2$ are $\{\neg p\}$ and $\{p\}$ respectively.

Say $w2$ is the true state of the world. Ann's ability is low. Ann knows that her ability is low. Ann also knows that Bob does not know her true ability. Will Ann lie to Bob? Ann can get a higher payoff by sending a dishonest message "High" signaling to Bob that her ability is high. If Bob is trusting and believes Ann's message, he will choose demanding; Ann will receive a payoff of 2 and Bob a payoff of zero. She may lie and get away with it.

Case 4: Consider the structure of possible worlds shown in Figure 21.

At $w1$, the following are true:

(1) p

(2) $K_A(p)$

(3) $K_B(p)$

(4) $\neg K_A K_B(p)$

In fact, Bob knows (1) through (4). Also Ann may consider it possible that Bob knows p.

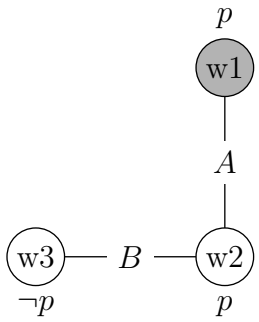


Figure 21: A and B are Ann and Bob's accessibility relations. The contents of the worlds $w1$, $w2$, and $w3$ are $\{p\}$, $\{p\}$, and $\{\neg p\}$ respectively.

Say $w1$ is the true state of the world. Ann's ability is low. Ann knows her ability is low. Bob know Ann's true ability. Ann does not know that Bob knows that her ability is low. Ann may send the message "High" since she can get a higher payoff if Bob acts based on her message. Bob may hire Ann for the undemanding job. In which case, Ann receives a payoff of 1 and Bob a payoff of 2. However, if Bob is annoyed by the fact that Ann lied to him, he may decide not to hit Ann. In which case, Ann and Bob receive a payoff of zero each.

Case 5: Consider the structure of possible worlds shown in Figure 22.

At $w1$, the following are true:

(1) p

(2) $K_A(p)$

(3) $K_B(p)$

(4) $K_A K_B(p)$

(5) $K_A(\neg K_B K_A K_B(p))$

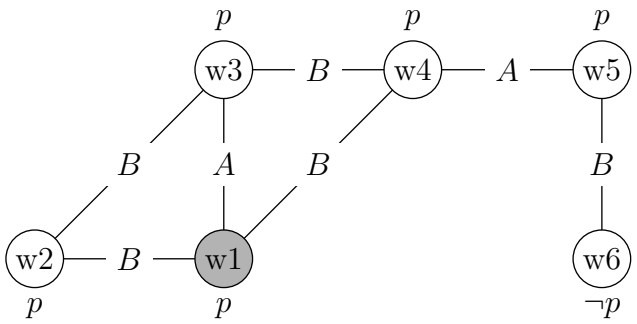


Figure 22: A and B are Ann and Bob's accessibility relations. The content of the worlds $w1$, $w2$, $w3$, $w4$, $w5$ and $w6$ are $\{p\}$, $\{p\}$, $\{p\}$, $\{p\}$, $\{p\}$, and $\{\neg p\}$ respectively.

Ann may send an honest message based on (1) to (4) i.e. (5) is not required. However, it may be the case that Ann may take advantage of the fact that Bob doesn't know whether Ann knows that Bob knows that Ann's ability is low. Therefore, Ann may send an honest message to impress Bob with her honesty.

Suppose Carl is an audience to Ann and Bob's conversation. Assume Carl knows Ann's true ability. The game proceeds as before, except there is a potential move by Carl who may choose to reveal or withhold information to

Bob about Ann's ability. Ann's ability together with Carl and Bob's actions determine the payoff for all three players.

Let p be the proposition, *"Ann's ability is low"* and q the proposition, ''*Carl is present.*" We are interested in the case where Ann's ability is low i.e. p is true and Ann is greedy. In other words, Ann prefers a highly paying job over a lower paying job and a lower paying job over not being hired regardless of her ability.

Case 6: Let's consider the structure of possible worlds shown in Figure 23 [13].

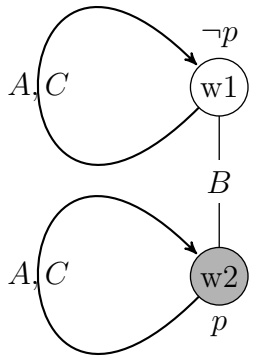


Figure 23: A, B, and C are Ann, Bob, and Carl's accessibility relations. The content of the worlds $w1$ and $w2$ are $\{\neg p\}$ and $\{p\}$ respectively.

At $w2$, the following are true:

[13]We are putting self loops at $w1$ and $w2$ to indicate that Bob is the only one who doesn't have knowledge of the full situation.

(1) p

(2) $\neg K_B(p)$

(3) $CK_{A,B,C}(q)$

(4) $CK_{A,C}(p \wedge \neg K_B(p))$

There are three outcomes that are of interest to us:

(O_1) Ann tells the truth by sending the message "Low" and Bob hires Ann for the undemanding job

(O_2) Ann lies by sending the message "High," Carl doesn't reveal Ann's true ability, and Bob hires Ann for the demanding job

(O_3) Ann lies by sending the message "High," Carl reveals Ann's true ability, and Bob does not hire Ann

Ann prefers O_2 to O_1 to O_3. However, Ann's choice is only between saying "High" or saying "Low." If she says "High" then what happens next depends on what Carl does. Ann's own action will depend on what she anticipates Carl will do. Carl can either reveal the value of p to Bob or not [14]. Also, whether Carl would reveal Ann's true ability to Bob depends on whether Carl values his relationship with Ann over ethics or vice versa. If Carl cares about his relationship with Ann more than what he believes is the right thing to do, he may keep quiet and let Bob hire Ann for the demanding job. If Carl is ethical and this fact takes precedence over his relationship to Ann, Carl will reveal Ann's true ability to Bob. In which case, Bob would not hire Ann.

[14]Since Bob doesn't know Ann but have worked with Carl, we assume he will take Carl's words over Ann.

Thus, if Ann's relationship to Carl is close then doing backward induction on Bob and Carl's moves, she will say "High." If Ann perceives her relationship to Carl to be distant, then she cannot hope for outcome O_2 and must choose between O_1 and O_3. Since her payoff in O_1 is higher, she decides to tell the truth. Table 2 shows the payoff for these different cases.

Ann	Bob	Carl
1	2	0
2	0	(a) 0 or (b) -1
0	0	(a) -5 or (b) 1

Table 2: Ann, Bob, and Carl's payoffs from outcomes O_1, O_2, and O_3.

If the outcome is O_1, then Ann and Bob's payoffs are 1 and 2 respectively. Carl has a payoff of zero as he has no moves. If the outcome is O_2, then Ann's payoff is 2 and Bob's payoff is 0. Carl's payoff is either zero if he is close to Ann or -1 if he is not. If the outcome is O_3 then Ann receives a payoff of zero and Bob a payoff of 0. Carl's payoff is -5 if he is close to Ann or 1 if he is distant.

In the example above, the audience may have an explicit move where he can reveal the sender's type to the receiver. Let's look at another example where the audience's presence alone may change the strategy the sender plays even though the audience has no move in the game.

Suppose Bob, an automobile salesman, is selling a used car to a customer Carl. Bob knows that the car is unreliable but Carl does not. Bob wants to earn a commission by selling the car and Carl wants to get the best deal. Nature chooses the type of car which is either reliable or unreliable, Bob

sends a message "Reliable" or "Unreliable," Carl chooses to buy or not buy the car. Nature's move together with Carl's action decided the payoff for both Bob and Carl. The payoff matrix is shown in Figure 24.

		Carl	
		B	N
Bob	R	$1, 1$	$0, 0$
	U	$1, -1$	$0, 0$

Figure 24: Normal form representation of the game where nature decided car type which is either reliable (R) or unreliable(U). Bob sends a message "Reliable" or "Unreliable" to Carl who decided whether to buy (B) or not buy (N) the car.

Let r be the proposition, *"The car is reliable."*

Case 7: Consider the structure of possible worlds shown in Figure 25.

At $w1$, the following are true:

(1) $\neg p$

(2) $K_B(\neg p)$

(2) $\neg K_C(\neg p)$

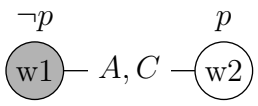


Figure 25: A and C are Ann and Carl's accessibility relation. The content of the worlds $w1$ and $w2$ are $\{\neg p\}$ and $\{p\}$ respectively.

Say $w1$ is the true state of the world. Then it is not the case that the car is reliable. Bob knows that the car is unreliable but Carl does not. Since Bob could get a higher payoff by selling an unreliable car to Carl, he may lie

to Carl and send the message "Reliable" to potentially induce a belief in Carl that the car is reliable. If Carl is trusting, he will buy the car. In which case, Bob receives a payoff of 1 and Carl a payoff of -1.

Suppose Ann, who is Bob's mother, is an audience in the game. Ann disapproves of Bob cheating and Bob knows this.

Case 8: Consider the structure of possible worlds shown in Figure 25.

Say $w1$ is the true state of the world. Will Bob send the message "Reliable" or "Unreliable" to Carl? Both Ann and Carl don't know whether the car is reliable or not. Bob wants to earn his commission and he may reason, *"What my mother doesn't know won't hurt her."* Bob may send the deceitful message "Reliable" to Carl. If Carl is trusting and believes Bob's message, he may decide to buy the car.

Case 9: Consider the same structure of the possible worlds as shown in Figure 26.

At $w1$, the following are true:

(1) $\neg p$

(2) $K_B(\neg p)$

(3) $K_A(\neg p)$

(4) $\neg K_C(\neg p)$

The car is unreliable. Both Ann and Bob know that the car is unreliable but Carl does not. What message will Bob send to Carl? If Bob is close to his mother and cares about her feelings, he may accept monetary loss in order

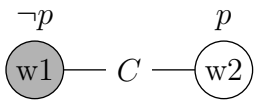


Figure 26: C is Carl's accessibility relations. The content of the worlds $w1$ and $w2$ are $\{\neg p\}$ and $\{p\}$ respectively.

to please Ann. He may send a honest message "Unreliable." If on the other hand, he doesn't care about his mother's disapproval, he may send the message "Reliable."

For cases 9 through 11, consider the structure of possible worlds shown in Figure 26 and say $w1$ is the true state of the world.

Case 9: Assume Ann likes her son to make a commission more than what she thinks is the right thing to do and Bob knows this. What message will Bob send to Carl? He may play his strategy as in the case where his mother was not present. That is try to deceive Carl into buying an unreliable car.

Case 10: Assume Bob and Carl are close friends. What message will Bob send to Carl? Here Ann's presence and whether she knows that the car is reliable or not doesn't come into the picture. Bob may consider his friend's loss and adjust his strategy accordingly.

Case 11: Consider the following situation. Dan, who is Bob's boss, is a second audience to Bob and Carl's conversation. Here Bob's strategy not only depends on his relationship with his mother but also his relationship with his boss. Sending the message "Unreliable" would please Ann while sending the message "Reliable" would please Dan. What message would Bob send to

Carl? He may choose to send a message such as, "the car has been recently painted[15]," leaving Carl to calculate whether the car is mechanically good. Saying something that is positive but not strong enough would not create as much anger in the boss and would not hurt his mother's feelings.

We have shown through a series of examples that three important factors lead to sender playing a different strategy from what he would normally play in the game; players' knowledge states, relationships, and trust. In all these cases, the original payoff matrix is transformed taking into consideration such factors.

13.2 Formal Model

A signaling game with an audience is a communication game between the sender S and the receiver R in the presence of an audience A. The game is characterized by a set of players $\mathcal{P}$, a set of payoff matrices $\mathcal{M}$, a set of worlds $\mathcal{W}$, a set of signals $\mathcal{F}$, a set of actions $\mathcal{A}$, a set of mental models $\mathcal{R}$, a semantic interpretation function $\rightsquigarrow_s$, a pragmatic interpretation function $\rightsquigarrow_p$, and utility functions $\bar{\mu_S}$ and $\bar{\mu_R}$. We assume all sets $\mathcal{P}$, $\mathcal{M}$, $\mathcal{W}$, $\mathcal{F}$, $\mathcal{R}$, and $\mathcal{A}$ are finite.

The game proceeds as follows.

1. Nature chooses $w \in \mathcal{W}$

[15]This is analogous to Grice's example where a professor writing a letter of recommendation says that the candidate has excellent handwriting without saying anything more and leaving the recipient to conclude that the candidate is weak.

2. S observes w but R does not [16]

3. S sends a signal $f \in \mathcal{F}$ to R

4. R chooses an action $a \in \mathcal{A}$ based on the S's signal f

5. The actual world w and receiver's action a determine the payoff for both players

6. All of the above takes place in the presence of an audience A

$\mathcal{P} = \{S, R, A\}$ is the set of players. Both S and R are active players in the sense that they have an explicit move in the game. The audience has no move in the game but his presence may affect the sender's signal and/or the receiver's action. The structure of the game is common knowledge among players.

The semantic interpretation function $\rightsquigarrow_s \in \mathcal{F} \mapsto \phi \subseteq \mathcal{W}$ maps signals to sets of worlds and a signal $f_1 \rightsquigarrow_s \{w_1\}$ says that the conventional meaning of f_1 is $\{w_1\}$. It is also possible to have a signal $f_{12} \rightsquigarrow_s \{w_1, w_2\}$ where the meaning of f_{12} is $\{w_1, w_2\}$. Signals' conventional meaning is common knowledge among players if they share a common language. Signals do not necessarily have an associated cost [17]. The pragmatic interpretation function $\rightsquigarrow_p \in \mathcal{F} \mapsto \varphi \subseteq \mathcal{W}$ also maps signals to sets of worlds but unlike the semantic interpretation function $\rightsquigarrow_s$, the pragmatic interpretation function $\rightsquigarrow_p$ may not

[16] It is possible for A to partially observe w.

[17] Michael Franke had a good remark that if the audience's presence affects whether the sender tells the truth or not (as in the automobile salesman example) then lying may affect the sender's net utility and is therefore costly.

be common knowledge[18]. It could very well be that $f_1 \rightsquigarrow_s \{w_1\}$ and the pragmatic interpretation the receiver chooses is $f_1 \rightsquigarrow_p \{w_1, w_2\}$. In the case where the receiver is close to the sender and fully trusts her, the literal and pragmatic meaning of the receiver will choose may coincide i.e. $f_1 \rightsquigarrow_s \{w_1\}$ and $f_1 \rightsquigarrow_p \{w_1\}$. In such cases, we can say the sender's signal is believed and the receiver may act based on the sender's signal.

In addition to introducing an audience into the two-player signaling games, we are also adding a new concept which is that of mental models. *Mental model* theory was developed by Johnson-Laird and Byrne [75][74]. The theory explains reasoning in terms of models of possibilities where each mental model represents what is common to a possibility. A mental model is a kind of internal representation of external reality that people use for cognition, reasoning, and decision-making[19][20].

Let $\mathcal{R}$ be the set of mental models representing players' perceived relationships with each other. We formally represent a mental model $r \in \mathcal{R}$ as a weighted directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{D}, \mathcal{L})$ where $\mathcal{V}$ denotes the set of vertices, $\mathcal{E} = \{< i, j > | i, j \in \mathcal{V}\}$ denotes the edge set and $< i, j >$ is an ordered pair of vertices, $\mathcal{D} = \{\Delta_{i,j} \in \mathbb{N} | i, j \in \mathcal{V} \wedge < i, j > \in \mathcal{E}\}$ denotes the set of distances on edges, and $\mathcal{L} : \mathcal{E} \mapsto \mathcal{D}$ is a mapping function which assigns distances to edges.

The function $\mathcal{L}$ closely relates to players' perceived relationships. A

[18]While the sender knows how his signal will be literally interpreted, he does not know the pragmatic interpretation the receiver will choose. However, it is possible for the sender to have guesses.

[19]Some background material can be found in Appendix B.

[20]Our notion of mental models may not be exactly the same as theirs but nonetheless there are similarities.

smaller $\Delta_{i,j}$ would means that player i perceives his relationship to player j to be close and a larger $\Delta_{i,j}$ would mean that player i perceives her relationship to player j to be distant. A naive or altruistic player may start off assigning smaller distances on edges going to other players and increment it as betrayed. A calculating or selfish player may start off assigning larger values on edges going to other players and decrement it as he starts to form closer relationships. Relationships are not necessarily symmetric i.e. $\Delta_{i,j}$ need not be the same as $\Delta_{j,i}$ for $i, j \in \mathcal{P}$. It may be possible to have multiple levels of mental models where players not only have a model of their perceived relationships to the other players but also have a model of other players' perceived relationships, etc. Since our players have limited processing capabilities, we limit the level of mental models to at most two. Players can act based on their perception of how they relate to other players but may also consider other players' perceived relationships. For simplicity, we assume direct relationships where each player consider one level of depth in the graph starting from their own vertex. Although, technically indirect relationships, such as S relates to A through R, can be accounted for by a graph traversal and aggregation of distances on the edges.

Definition 1 Let $\bar{\mu_i} \in \mathcal{M} \times \mathcal{W} \times \mathcal{A} \mapsto \mathbb{R}$ be the surface utility for player $i \in \mathcal{P}$. Given the surface utility $\bar{\mu_j}$ for players j, a measure of relationship $\Delta_{i,j}$ between i and j as perceived by player i, we define player j's utility from player i's perspective as

$$\beta(j, i) = \frac{\bar{\mu_j}}{\Delta_{i,j}}$$

Definition 2 Given surface utilities $\bar{\mu_i}$, $\bar{\mu_j}$, and $\bar{\mu_k}$ for $i, j, k \in P$, $\Delta_{i,j}$, and $\Delta_{i,k}$, we define player i's net utility[21] as

$$\mu_i = \bar{\mu_i} + \beta(j, i) + \beta(k, i)$$

It is not necessary for the audience to have a surface utility in the game. However, the audience has a net utility in the game which is calculated by adding $\beta(S, A)$ and $\beta(R, A)$.

The space of pure sender strategies $\mathcal{S} = \mathcal{W} \mapsto \mathcal{F}$ is the set of functions from worlds to signals. The space of pure receiver strategies $\mathcal{R} = \mathcal{F} \mapsto \mathcal{A}$ is the set of functions from signals to actions.

$\mathcal{M}$ contains one ore more matrices. The game matrix $m_{CK} \in \mathcal{M}$ is the surface matrix that is common knowledge between players. In addition to m_{CK}, there may be two transformed matrices $m_S \in \mathcal{M}$ and $m_R \in \mathcal{M}$ from the sender and the receiver's perspectives. These matrices are the result of S and R correcting their surface utilities in m_{CK} taking into consideration their first-level mental models of how they perceive to be related to other players. It is not necessary for m_S and m_R to be common knowledge. In m_{CK}, m_S, and m_R the rows are worlds and the columns are actions.

Since the receiver doesn't know the actual world w and his strategy is from signals to actions, a conversation needs to happen where the receiver is converting his transformed matrix m_R to m_R^σ where rows are signals and columns are actions. In other words, the receiver has to map the sender's

[21] We can consider social norms into the calculation of net utilities but we leave it out of the definition as it greatly varies by culture.

signals to sets of worlds and calculate his payoff.

Suppose m_{CK} has three rows w_1, w_2, and w_3 and the sender sends f_1 if the world is w_1, f_2 if the world is w_2 and f_3 if the world is w_3. The receiver uses $f_1 \rightsquigarrow_s \{w_1\}$, $f_2 \rightsquigarrow_s \{w_2\}$, and $f_3 \rightsquigarrow_s \{w_3\}$ to convert his transformed matrix m_R^σ. In this case, the new matrix has the same number of rows as m_{CK} where w_1, w_2, and w_3 are replaced by f_1, f_2, and f_3 respectively. Of course, there are more complicated cases such as one where given three rows in m_{CK}, the sender is using two signals $f_1 \rightsquigarrow_s \{w_1\}$ and $f_{12} \rightsquigarrow_s \{w_1, w_2\}$. Thus, the receiver's transformed matrix m_R^σ will have two rows instead of three i.e. f_1 and f_{23}. For the signal f_{23}, the receiver has to calculate his utility from w_2 and w_3 in m_R. If utilities are cardinal and worlds w_2 and w_3 are equally likely then the receiver may use Stalnakar's approach of taking an average of his utilities. The receiver may take the minimum of the payoffs from w_1 and w_2 if he is using Maximin or he may use some other strategy.

The sender doesn't necessarily know m_R^σ but he can have guesses $m_{SR}^{\sigma_1}$ ... $m_{SR}^{\sigma_k}$ which are matrices from the receiver's perspective as imagined by the sender. Thus the effective mental model of the sender for the receiver will be one or more matrices whose rows are signals and columns are actions. The receiver's actual matrix m_R^σ may be among these which the sender considers possible. To fix thought we assume $k = 1$. You can think of it as a *theory of mind* that the sender ascribes to the receiver to explain the receiver's behavior and predict his action. It is important to note that the sender's signal is based on how she thinks the receiver will interpret her signal and what the receiver will do if her signal is interpreted in a certain way. She may ask, *"What is*

the receiver thinking and how will my signal be interpreted?" The issue of how the signal is actually interpreted by the receiver does not arise at the time the sender is contemplating of sending her signal. Once the sender sends a signal, it is up to the receiver to decide which action to choose.

Let us now define players' best strategies given this apparatus.

Definition 3: Let $r^* \in \mathcal{R}$ be a receiver strategy and m_R^σ be the receiver's transformed matrix where rows are signals and columns are actions. Then r^* is a *best response* of the receiver to a strategy $s \in \mathcal{S}$ of the sender if and only if $r^* \in BR_R(s)$ where

$$BR_R(s) = \arg\max_{r \in \mathcal{R}} [\mu_R^\sigma(m_R^\sigma, s, r)]$$

Definition 4: Let $w \in \mathcal{W}$ be the actual world, $s^* \in \mathcal{S}$ be a sender strategy, m_S be the sender's transformed matrix, and m_{RS}^σ be the receiver's transformed matrix as imagined by the sender. Then s^* is a *best strategy* by the sender if and only if $s^* \in BS_S(w)$ where

$$BS_S(w) = \arg\max_{s \in \mathcal{S}} \left[\mu_S(m_S, w, \arg\max_{r \in \mathcal{R}} [\mu_R^\sigma(m_{RS}^\sigma, s, r)])\right]$$

One could extend the above definition to one where the sender imagines more than one matrices from the receiver's perspective. The sender has a belief

that when she sends a signal $f \in \mathcal{F}$, the receiver will take an action $a \in \mathcal{A}$ which will give her a payoff from m_S. Since she knows the actual world, she will choose that signal f which would give her the "best" payoff given the receiver's strategy. The payoff that the sender gets from m_S depends on what she thinks the receiver will do once she sends a signal $f \in \mathcal{F}$. Say the sender has two possible signals f_1 and f_2 for w_1 and w_2 respectively, and she imagines two matrices m^1_{SR} and m^2_{SR} from the receiver's perspective. Then for each of her two strategies i.e. sending signal f_1 or sending signal f_2, she considers how the receiver will act in m^1_{SR} and m^2_{SR}. Suppose the outcomes from sending the signal f_1 are x and y and the outcomes from sending the signal f_2 are x' and y' from m^1_{SR} and m^2_{SR} respectively. Then the sender has two sets of possible payoffs for each of her signals f_1 and f_2. The sender is uncertain about which matrix the receiver is using. The sender may compare these two sets of payoffs and pick that signal which is "best" for her given the receiver's strategy. The sender may use Minimax, Maximin, or some other strategy to calculate her best response to the receiver's strategy.

13.3 Examples

Let us revisit my colleague example that started it all. We'll formalize and explain it in terms of our model.

Suppose Ann and Bob work together and they report to Carl, who is an IT manager. Bob has trouble debugging a Java program he has written and approaches Ann for help. Ann is an expert and wants to help her colleague.

The surface matrix m_{CK} (Figure 27) is common knowledge between Ann and Bob.

		Ann	
		H	N
Bob	J	$3, 1$	$0, -1$
	S	$0, 0$	$0, 0$

Figure 27: Normal form representation of the game m_{CK} where Bob needs help with Java code (J) or show off his expertise in Sql (S). Bob sends the message "Java" if he needs help with Java or the message "Sql" if he wants to show off his Sql expertise. Ann decides whether to help (H) or not help (N) Bob.

Let nature's move be J i.e. Bob needs help with Java code. This is Bob's private information which Ann does not have. Bob sends the message "Java" to Ann asking for her help. Ann wants to help her colleague and she chooses H. Bob gets a payoff of 3 and Ann a payoff of 1.

Now imagine the following scenario. Bob needs help with his Java code and is about to send the message "Java" to Ann asking for her help when Carl becomes an audience in the signaling game between Bob and Ann. Let us say, Bob knows of Carl's presence but Ann does not. Bob sends the message "Sql" instead of message "Java" to Ann. Why did Bob send a different message? Bob is now playing from his transformed matrix m_{Bob} (Figure 28) which he may have computed taking into consideration his relationship with the manager. Perhaps he thinks if he asks for help while Carl is present, Carl would come to know that he is weak in Java. Bob is playing his best strategy using his transformed matrix while Ann is playing the game using the surface matrix. Bob sends the message "Sql". Ann chooses not to help (N). Since the actual world and the receiver's action determine players' payoffs, Bob receives a payoff

of 0 from m_{Bob} and Ann receives a payoff of -1 from m_{CK}.

		Ann	
		H	N
Bob	J	$-3, 1$	$0, -1$
	S	$0, 0$	$0, 0$

Figure 28: Normal form representation of the transformed game m_{Bob} from Bob's perspective when Carl is an audience.

Let us say that Ann sees Carl right after Bob starts bragging about his Sql expertise. Now Ann has her own theory of the situation. She has her own transformed matrix due to Carl's presence. Let m_{Ann} (Figure 29) be Ann's transformed matrix in the presence of Carl. Ann gets a higher payoff from m_{Ann} helping Bob while Carl is watching. Perhaps Ann thinks Carl will be impressed with her helping Bob.

		Ann	
		H	N
Bob	"$Java$"	$3, 2$	$0, -1$
	"Sql"	$0, 2$	$0, 0$

Figure 29: Normal form representation of the transformed game m_{Ann} from Ann's perspective when Carl is an audience.

Bob chooses that strategy which is "best" using his transformed matrix m_{Bob}. In this case, the signal "Sql" would get him a payoff of -1 if Ann believes it and chooses not to help. A payoff of -1 is better than a payoff of -3 so Bob sends the message "Sql." Ann chooses the strategy which is "best" using her own transformed matrix m_{Ann}. Ann can get a higher payoff by showing Carl that she is helping Bob. Her payoff from helping Bob is 2 independent of Carl's signal. Therefore, Ann chooses to help. Since the payoffs are determined by

the actual world and Ann's action, Bob gets a payoff of -3 from his transformed matrix m_{Bob} and Ann receives a payoff of 2 from her transformed matrix m_{Ann}.

If Bob is intelligent, he can predict Ann's behavior by imagining Ann's transformed matrix. Let m_{AnnBob} be the transformed matrix as imagined by Bob from Ann's perspective. For simplicity, let's say m_{AnnBob} is the same as Ann's transformed matrix m_{Ann}. Then Bob can try to guess what is Ann's strategy for each of his signals. Bob can ask, *If I send the signal "Java" Ann's best response is H and the payoff from m_{Bob} is -3. If I send the signal "Sql," Ann's best response is H, which would give me a payoff of -3. She will choose H regardless of my message and I will get a payoff of -3 as my payoff is determined by actual world and Ann's action. So I might as well send the signal "Java." At least I will have my broken code fixed* [22].

This example clearly shows how players may be playing the same game using different matrices than what is common knowledge. It is possible for Ann and Bob's transformed matrices to be the same and if it were common knowledge the game reduces to one where players strategize choosing their best strategy given other's strategy as in the original game.

Let us re-examine some examples from the literature in terms of our model. Skyrms[134] provides an example of deception among non-human species.

> *Fireflies use their light for sexual signaling. In the western hemisphere, males fly over meadows, flashing a signal. If a female on*

[22]Of course, one could also consider the case where the actual world is "Sql" and Bob doesn't need help but Ann is tempted to show to the manager that she is helping Carl.

the ground gives the proper sort of answering flashes, the male descends and they mate. The flashing "code" is species-specific. Females and males in general use and respond to the pattern of flashes only of their own species. There is, however, an exception. A female firefly of the genus Photuris, when she observes a male of the genus Photinus, may mimic the female signals of the males species, lure him in, and eat him. She gets not only a nice meal, but also some useful protective chemicals that she cannot get in any other way.

		M	
		I	NI
F	G	$2, 2$	$0, 0$
	NG	$0, 0$	$0, 0$

Figure 30: The surface matrix m_{CK} is common knowledge between Photinus male and female. The sender (F) sends a signal "Go" or "No Go" corresponding to the worlds go (G) or no go (NG). The receiver (M) chooses to interact (I) or not interact (NI). The dominant strategy for the male firefly is to choose I if the world is G.

Figure 30 shows the signaling game between female and male fireflies of the genus Photinus. We'll call this game, love to death.

		M	
		I	NI
F	"Go"	$2, 2$	$0, 0$
	"No Go"	$0, 0$	$0, 0$

Figure 31: Photinus male firefly's transformed matrix m_M where rows are possible Photinus female signals and columns are his actions.

The game matrix m_{CK} is common knowledge between the sender and

		M	
		I	NI
F	G	$2,2$	$0,0$
	NG	$0,0$	$0,0$

Figure 32: Photinus female firefly's transformed matrix m_F where rows are worlds and columns are actions which is identical to m_{CK}.

		M	
		I	NI
F	"Go"	$2,2$	$0,0$
	"No Go"	$0,0$	$0,0$

Figure 33: Photinus male firefly's transformed matrix m_{MF} as imagined by Photinus female firefly.

the receiver. Signals have pre-defined meaning[23] which is common knowledge between the sender and the receiver i.e. "Go" $\rightsquigarrow_s \{G\}$ and "No Go" $\rightsquigarrow_s \{NG\}$. The receiver reasons[24] as follows, *If I receive a signal "Go" from the sender then it is the case that the world is G. If I receive a signal "No Go" from the sender then it is the case that the world is NG. My Best response to "Go" would be to interact as it will give me a higher payoff.* In effect, the receiver is transforming the surface matrix into m_M (Figure 31) which associates signal/action pairs

[23]Signals with pre-defined meaning are signs if they are also common knowledge.

[24]We do not intend to attribute to insects a faculty for reasoning but rather explain their action based on our observation of their behavior.

		M	
		I	NI
F	G	$2,-10$	$0,0$
	NG	$0,0$	$0,0$

Figure 34: Photuris female firefly's transformed matrix $m_{F'}$ where rows are worlds and columns are actions.

to payoffs. In the normal case, where the female firefly intends to mate with the male, the sender's transformed matrix m_F (Figure 32) is identical to m_{CK}. Additionally, the sender imagines the receiver having a transformed matrix m_{MF} (Figure 33) and that the receiver has a best response for each of her signals. The sender reasons as follows, *The signal "Go" will give me a higher payoff in m_F. If I send the signal "Go," the receiver will choose to interact since that action is his best response in m_{MF} to my signal "Go." Therefore, I will send the signal "Go".*

Now let us look at the case where the female firefly of the genus Photuris wants to deceive the male firefly of the genus Photinus. What sets this apart from the normal case is the fact that the sender's transformed matrix $m_{F'}$ (Figure 34) is different from the surface matrix m_{CK}. Here the sender receives the same payoff of 2 but gets a meal instead of a mate.

The receiver has transformed m_{CK} into m_M where rows are signals and columns are actions. He is using the pre-defined meaning of the signals to guess which world he is in and act accordingly. His best strategy given m_M and the signal "Go" is to interact (I) with the female. The sender also has her own net matrix $m_{F'}$ which is different than the surface matrix m_{CK} that is common knowledge. The sender also has her mental model of what matrix the receiver is using. Let m_{MF}, which is the same as m_M, be the receiver's transformed matrix as imagined by the sender. The sender may guess the possible receiver actions for each of her messages. The sender may reason as follows, *If I send the signal "Go", the receiver will choose "I" using m_{MF} thinking that he will receive a payoff of 2. If the receiver chooses action I, my*

payoff from $m_{F'}$ is 2 and the receiver's payoff is -10 but the receiver' doesn't know this. I will send the signal "Go". Photuris female receives a payoff of 2 as in the case of Photinus female and Photinus male gets eaten thus a payoff of -10. Photinus male dies happy thinking he's receiving a payoff of 2.

The love to death game clearly shows how the sender and the receiver are acting based on different matrices. Photuris female is making use of language with established meaning i.e. the meaning of signals have been established by the females of Photinus who want to mate rather than eat Photinus male!

Let's look at some examples where Gricean Implicature is affected by the presence of an audience.

Ann pays Bob a visit. Bob wants to offer Ann tea or coffee but doesn't know her preference. Bob asks Ann whether she likes tea or coffee. Ann signals her preference with a signal "Tea" for tea and "Coffee" for coffee. There are two possible worlds w_1 and w_2. In w_1 Ann prefers tea and in w_2 she prefers coffee. Bob has a choice between two actions, a_1 and a_2, offering tea or coffee to his guest. Figure 35 shows the surface matrix m_{CK} for the game.

		Bob	
		a_1	a_2
Ann	w_1	$1, 1$	$0, 0$
	w_2	$0, 0$	$1, 1$

Figure 35: Normal form representation of the game m_{CK} where Ann and Bob's preferences are aligned and Bob makes his action dependent on Ann's Cheap Talk message.

Here Ann and Bob's preferences are aligned i.e. Ann likes to drink

		Bob	
		a_1	a_2
Ann	"Tea"	$1, 1$	$0, 0$
	"Coffee"	$0, 0$	$1, 1$

Figure 36: Normal form representation of the transformed game m_{Bob} from Bob's perspective where rows are possible sender signals and columns are actions.

coffee and Bob wants to treat his guest well. If w_2 is the true state of the world, Ann will send message "Coffee" and Bob will take action a_2 using m_{Bob} (Figure 36) and offer Ann coffee. Ann gets a payoff of 1 from m_{CK} and Bob gets a payoff of 1 from m_{Bob}.

Suppose Carl is present and an audience to Bob and Ann's conversation. Say Carl likes Ann, and just the day before, the following conversation took place between Ann and Carl.

Carl: *Would you like to go out for coffee?*

Ann: *I don't like coffee.*

By sending the signal "Coffee" to Bob, Ann is observing the Cooperative Principle with Bob but implicature arises between Ann and Carl. Carl may think that Ann intends for him to know that she is definitely not interested and that may be Ann's true intention. However, if Ann cares about Carl's feelings, she may send the message "Tea" instead.

Let us look at the case where Ann doesn't want to be rude to Carl or hurt his feelings. Let m_{Ann} (FIgure 37) be Ann's transformed matrix in Carl's presence.

Let the matrix that Ann imagines from Bob's perspective m_{BobAnn} be

		Bob	
		a_1	a_2
Ann	w_1	$1, 1$	$0, 0$
	w_2	$0, 0$	$-1, 1$

Figure 37: Normal form representation of the transformed game m_{Ann} from Ann's perspective.

the same as m_{Bob}. Ann reasons as follows, *If I send the message "Coffee," Bob will take action a_2 using m_{BobAnn} and my payoff from m_{Ann} for world w_2 and a_2 is -1. If I send the message "Tea," Bob will choose action a_1 and my payoff from m_{Ann} is 0 and Bob's payoff from m_{BobAnn} is 1. A payoff of 0 is better than -1. I will send the message "Tea."*

Let $m_{Ann'}$ (Figure 38) be Ann's transformed matrix in the case where Ann intends for Carl to know that she is definitely not interested.

		Bob	
		a_1	a_2
Ann	w_1	$1, 1$	$0, 0$
	w_2	$0, 0$	$2, 1$

Figure 38: Normal form representation of the transformed game $m_{Ann'}$ from Ann's perspective.

Here Ann gets a higher payoff by send the message "Coffee" to Bob. She not get a cup of coffee but also gets her message a cross to Carl.

Let us modify Searle's example that Grice [63] used to distinguish between literal and pragmatic meaning of a sentence. We'll add an audience and re-examine it in terms of our formal model.

An American soldier in the Second World War is captured by

Italian troops. In order to get the Italian troops to release him he intends to tell them in Italian or German that he is a German soldier. He doesn't know Italian but says the only German line that he knows, *Kennst du das Land, wo die Zitronen bluhen* which in German means *Knowest thou the land where the lemon trees bloom*. However, the Italian troops who do not know this meaning but can figure out the soldier is speaking in German, may reason as follows. *The soldier just spoke in German. He must intend to tell us that he is a German soldier. Why would he speak in German otherwise? It could very well be that he is saying I am a German soldier.*

Here, the sentence uttered by the American soldier doesn't literally mean but implies that the American soldier is German. As one can see, the fact that the Italian troops do not know the literal meaning of the sentence the American soldier uses with the intention of inducing a belief in them that he is German is crucial to the reasoning on both parts.

The surface matrix m_{CK} is shown in Figure 39.

		T	
		R	D
S	A	$1, -1$	$-1, 1$
	G	$1, 1$	$-1, -1$

Figure 39: Normal form representation of the surface matrix m_{CK} between the American soldier (S) and the Italian troops (T). The rows are states of the worlds American (A) and German (G). The columns are actions the Italian troops can take i.e. release (R) or detain (D).

This is an example where $\rightsquigarrow_s$ is defined but not common knowledge. The Italian troops can only guess what language the sentence belongs to but not what it means. So the function $\rightsquigarrow_p$ maps a German sounding sentence to G and an English sentence to A. It seems natural to think that the pragmatic semantic function is common knowledge in this case as the American soldier's reasoning would only work if he knew that the Italian troops are using "G" $\rightsquigarrow_p \{G\}$.

The receiver's transformed matrix m_T is shown in Figure 40.

		T	
		R	D
S	"E"	$1, -1$	$-1, 1$
	"G"	$1, 1$	$-1, -1$

Figure 40: Normal form representation of the transformed matrix m_T for the Italian troops where the rows are signals and columns are actions.

The American soldier knows the Italian troops' transformed matrix and make use of it to get himself released. The American soldier reasons as follows, *If I send the message "G," the Italian troops may release me but if I send the message "E" (or any other English sentence for that matter) then the Italian troops may detain me. I get a higher payoff from uttering the only German sentence that I know. Let me utter that sentence.*

Suppose we add an audience to the game between the American soldier and the Italian troops. Say the audience speaks German and the Italian troops know this but the American soldier does not. The Italian troops can find out the literal meaning of the German sentence by asking the audience. Once the audience informs the Italian troops of the literal meaning of the sentence, the

Italian troops' transformed matrix changes to $m_{T'}$ shown in Figure 41.

		T	
		R	D
S	"E"	$1, -1$	$-1, 1$
	"G"	$1, -1$	$-1, 1$

Figure 41: Normal form representation of the transformed matrix $m_{T'}$ where the rows are signals and the columns are actions.

The American soldier will utter the German sentence considering m_{CK} and m_T. The Italian troops who now know that the solder is not German will choose to detain him. The Italian troops receive a payoff of 1 from $m_{T'}$ and the American soldier a payoff of -1 from m_{CK}.

We have accounted for the results from empirical studies in our formal model. The definition of net utilities where each player considers the benefit or loss to other players based on their perceived relationships provides the mechanism for addressing questions that the existing signaling models fail to answer, such as deception. A number of empirical studies [80][69] suggest that people have an aversion to lying. People don't lie if the loss to the other player is greater than their own gain and people lie less often to friends than strangers.

Let us re-examine the job applicant example in terms of our model. The matrix for one version of the game is shown in Figure 42.

Say Ann's ability is low. Ann has an incentive to lie where by sending the message "High," she can get a payoff of 2 if Bob believes her message and hires her for the demanding job. Let's look at how Ann may play a different

		Bob	
		D	U
Ann	H	$2, 1$	$0, 0$
	L	$2, 0$	$1, 2$

Figure 42: Normal form representation of the game where Nature chooses Ann's type high (H) or low (L). Ann sends the message "High" or "Low" to Bob. And Bob decides whether to hire Ann for the demanding (D) or undemanding job (U).

strategy based on her net utility.

Let $\Delta_{Ann,Bob}$ be a measure of relationship between Ann and Bob as perceived by Ann. Say $\Delta_{Ann,Bob} = 100$, Ann perceives her relationship to Bob to be distant. Then Ann's net utility from (L, D) is 2 and her net utility from (L, U) is 2.02. As Ann's net utilities are not affected much by her relationship to Bob, she may lie and send the message "High" to Bob.

If on the other hand, $\Delta_{Ann,Bob} = 1$, Ann perceives her relationship to Bob to be close, then her net utility from (L, D) is 2 but her net utility from (L, U) is 3. Ann gets a higher net utility by not lying to Bob and she will send the message "Low." and being honest to Bob.

The game shown in Figure 43 is a modification of the game in Figure 42.

		R	
		D	U
S	H	$2, 1$	$0, 0$
	L	$2, -10$	$1, 2$

Figure 43: Normal form representation of the game where Nature chooses Ann's type high (H) or low (L). Ann sends the message "High" or "Low" to Bob. And Bob decides whether to hire Ann for the demanding (D) or undemanding job (U).

As before Ann's ability is low but she has an incentive to lie Bob. Say $\Delta_{Ann,Bob} = 100$, Ann perceives her relationship to Bob to be distant, then Ann's net utility from (L, D) is 1.9 and her net utility from (L, U) is 1.01. If on the other hand, $\Delta_{Ann,Bob} = 1$, Ann perceives her relationship to Bob to be close, then her net utility from (L, D) is -8 and her net utility from (L, U) is 3. In either case, Ann's net utility is higher if she is honest to Bob and she will send the message "Low."

Let us assume Ann and Bob are distant and their net utilities are close or the same as their surface utilities. We'll examine how Carl's presence may affect Ann and Bob's strategies. Consider the surface matrix m_{CK} shown in Figure 44. As before, Ann's ability is low and she has an incentive to lie. Carl is an audience to Ann and Bob's conversation. Assume Carl knows Ann's ability.

		Bob	
		D	U
Ann	H	$2, 1$	$0, 0$
	L	$2, 0$	$1, 2$

]Normal form representation of the game m_{CK}.

If Ann knows whether Bob and Carl are close or distant, she may calculate Bob's transformed matrix. However, if she is not sure of their relationships then she may imagine two matrices from Bob's perspective $m^{\sigma_1}_{BobAnn}$ (Figure 45) and $m^{\sigma_2}_{BobAnn}$ (Figure 46). In $m^{\sigma_1}_{BobAnn}$, Bob and Carl are close and in $m^{\sigma_2}_{BobAnn}$, Bob and Carl are distant.

In $m^{\sigma_1}_{BobAnn}$, Ann imagines Bob and Carl being close. Ann's payoff from

		Bob	
		D	U
Ann	"High"	$-2, 1$	$0, 0$
	"Low"	$2, 0$	$1, 2$

Figure 45: Normal form representation of the game $m^{\sigma_1}_{BobAnn}$ where Ann thinks Bob and Carl are friends and suspects that Carl would reveal her true ability to Bob.

		Bob	
		D	U
Ann	"High"	$2, 1$	$0, 0$
	"Low"	$2, 0$	$1, 2$

Figure 46: Normal form representation of the game $m^{\sigma_2}_{BobAnn}$ where Ann thinks Bob and Carl are distant and suspects that Carl would not reveal her true ability to Bob.

the signal "High" is -2 as Carl may reveal her true ability to Bob. In $m^{\sigma_2}_{BobAnn}$, Ann imagines Bob and Carl as being distant and her payoff is 2 as before. This is an interesting case where Ann's temperament may affect what signal she sends. So while contemplating if she should send the signal "High," if Ann is risk averse, she would not lie to Bob as she could end up with a negative payoff. However, if she is aggressive she may lie to Bob anyway taking the risk of Carl revealing information about her ability to Bob.

14 Conclusion

The computer that was originally built for computing numbers has evolved into a device for computing with all types of information, words, numbers, graphics, and sounds. Thus, with the commoditization of computers and invention of the Internet, the computer has turned into a communication device, transmitting information between people. It has become the new medium for signaling. As information travels faster, the world seems smaller, and our understanding of the external world and self is evolving. Our traditional notions of identity, reality, truth, information, knowledge, and communication are changing. All of these are important issues that need attention but addressing them all is beyond the scope of this thesis.

We live in a digital era, where every action is recorded, transmitted, replicated, and shapes who we are. We constantly exchange information in the presence of an inevitable and often unnoticed audience. In this thesis, we have discussed real world problems associated with signaling in the presence of an audience, the limitations of current game theoretic models, and the urgency of building better models to capture the dynamics of information exchange in communication.

Communication is a goal-oriented activity where interlocutors use language as a means to achieve an end while taking into account the goals and plans of others. Game theory, being the scientific study of strategically interactive decision-making, provides the mathematical tools for modeling language use among rational decision makers. When we speak of language use, it is

obvious that questions arise about what someone knows and what someone believes. Such a treatment of statements as moves in a language game has roots in the philosophy of language and in economics. In the first, the idea is prominent with the work of Strawson, later Wittgenstein, Austin, Grice, and Lewis. In the second, the work of Crawford, Sobel, Rabin, and Farrell.

We have argued that existing models of signaling are over-idealized and fail to explain the dynamics of information exchange in communication. In particular, we have argued that the two-player signaling game doesn't apply to the research problem we have identified, where the sender sends information to the receiver in the presence of an audience. We have also argued that relationships among players lie at the heart of communication and trust is the heuristic decision rule that allows us to deal with complexities that would require unrealistic effort if we had to rationally decide. It is the heuristic rule that helps us converse with each other.

In this thesis, we have brought together ideas from philosophy of language, game theory, psychology, logic, and computer science. We have extended Grice and Lewis' ideas on cooperative communication and the ideas of Crawford, Farrell, Rabin, Sobel, and Stalnakar on communication with partially overlapping interests. We have supplemented the traditional model of signaling games with the following innovations: We have considered the effect of the relationships, whether close or distant, among players. We have considered the role that ethical considerations may play in communication. We have shown that communication requires awareness of self knowledge and knowledge of others. Finally, in our most significant innovation, we have introduced

an audience in a two-player signaling game whose presence affects the sender's signal and/or the receiver's response.

In our model, we no longer have assumed that the entire structure of the game is common knowledge as some of the priorities of the players and relationships among some of them might not be known to the other players.

15 Appendix

15.1 Language of Knowledge

The language of Logic of Knowledge consists of a set of finitely many individuals $I = \{1, \ldots, n\}$. The language is that of propositional calculus augmented by modal operators K_i, for each i ∈ I, as follows:

a) Atomic formulae P=$\{p_1, \ldots, p_m, \ldots\}$ is a set of variables of the propositional calculus; they are to be interpreted as "primitive" facts.

b) Connectives C=$\{\neg, \wedge\} \cup \{K_i : i \in I\}$ is the set of connectives. The K_is are modal operators; $K_i\varphi$ intuitively means: "agent i knows φ." $\varphi \vee \psi$ is equivalent to $\neg(\neg\varphi\ wedge\neg\psi)$ according to DeMorgan's Law. We define an abbreviation $L_i(\varphi)$ which is equivalent to $\neg K_i(\neg\varphi)$. $L_i(\varphi)$. $L_i(\varphi)$ intuitively means: "agent i thinks φ is possible."

c) Well formed formulae WFF is the set of formulae defined as: If $p_j \in P$, then $p_j \in WFF$. If $\varphi, \psi \in WFF$, then $\neg\varphi \in WFF$ and $(\varphi \wedge \psi) \in WFF$. If $\varphi \in WFF$ and $i \in I$, then $K_i(ărphi) \in WFF$. That is a sentence in the language of knowledge is either an atomic formulae p or an expression of the form $\neg\varphi$, $\varphi \wedge \psi$, and $K_i(\varphi)$ where φ and ψ are recursively built sentences.

The notion of knowledge we want to capture is axiomatized by the following set of axioms (called LK5 system):

A1. All tautologies of propositional logic

A2. $K_i\varphi \wedge \mathrm{K}_i(\varphi \rightarrow \psi) \rightarrow K_i\psi$

A3. $K_i\varphi \rightarrow \varphi$

A4. $K_i\varphi \rightarrow \mathrm{K}_i\mathrm{K}_i\varphi$

A5. $L_i\varphi \rightarrow \mathrm{K}_i\mathrm{L}_i\varphi$

R1. $\varphi, \varphi \rightarrow \psi \vdash \psi$

R2. $\varphi \vdash K_i\varphi$

A1 and R1 respectively are the axioms and the modus ponens rule of propositional logic. A2 states that an individual's knowledge is closed under implication, that is, if an individual i knows a formula, then he also knows all its logical consequences. A3 states that individuals know only things that are true. A4 and A5 state that individuals are in introspective; if an individual knows a formula, then he knows of knowing it. There are no universal consensus on assuming the introspection axioms A4 and A5.

The above axiomatization parallels modal logic. In fact, upon reading K_i as the necessity operator, and L_i as the possibility operator, we obtain the axiom system $S5$. This is the reason why our logic of knowledge has been called $LK5$. The parallel modal logic can go further, if we drop axiom scheme A5, the resulting logic (called $LK4$) corresponds to $S4$. Finally, taking out also axiom scheme $A4$, we obtain a system (called LK) that corresponds exactly to system T.

Logical Omniscience: Axioms A2 together with R2 above raises the so-called problem of "logical omniscience." They force a view of individuals as perfect reasoners: adding a theorem ξ implies that $K_i\xi$ also becomes a theorem, and hence it is impossible to have $\xi \wedge \neg K_i\xi$. All individuals then

know all valid formulas and also all their logical consequences. This does not seem to be a realistic model for dealing with everyday reasoning. Even if ξ is valid, we may fail to know ξ.

Suppose we drop R2 from the above axiomatization, and add the axiom of logical omniscience:

A6. If $\vdash \xi$ then $\vdash K_i\xi$

If $\phi \vdash \xi$ then $K_i\phi \vdash K_i\xi$

And also add the axiom scheme:

A7. If ψ is an axiom according to A1-A6, then so is $K_i\psi$ for each i.

Then in the new system all the old theorems are preserved, but now $\xi \wedge \neg K_i\xi$ is consistent. This is so because the new system still preserves the necessitation rule, but it states that necessitation is reasonable only for those formulae φ's which are logically true, or at least true on the whole model.

15.2 Models of Knowledge

We need a semantics in order interpret sentences about knowledge. A semantics consists of an idealized model of the world and an account of when a sentence in the logic is true in the model. Two commonly used models of knowledge are Information and Kripke Structures; the former uses partitions and the later accessibility relations to model knowledge.

Information Structures: An information structure of N players is a pair $(W, (P_i))$ where W is the set of states and P_i is a function that assigns to each

state w a non-empty subset of states $P_i(w)$ for each player i where $i \in N$. At state w, player i considers the states in $P_i(w)$ possible and excludes the states outside $P_i(w)$. We can impose some conditions on information structure:

1. $w \in P_i(w)$ (Players considers the true state possible)

2. If $w' \in P_i(w)$ then $P_i(w') \subseteq P_i(w)$

3. If $w' \in P_i(w)$ then $P_i(w') \supseteq P_i(w)$

These three conditions together are equivalent to saying that the information structure is partitional.

Let $(W, (P_i))$ be an information structure. We say that the event $E \subseteq W$ is known at state w by player i if $P_i(w) \subseteq E$. The statement "player i knows E" is then identified with all the states in which E is known: $K_i(E) = \{w : P_i(w) \subseteq E\}$. Using this definition and the assumption given above, we can derive the following properties about a player's knowledge:

I1. $K_i(E) \subseteq E$ (using 1)

I2. $K_i(E) \subseteq K_i(K_i(E))$ (using 2)

I3. $\neg K_i(E) \subseteq K_i(\neg K_i(E))$ (using 3)

Kripke Structures: We can interpret above logical system using models with possible worlds which intuitively says that besides the current state of affairs, there are other possible states of affairs (i.e. other possible worlds) for individual i; individuals may be unable to distinguish the true world among all possible worlds. An individual is said to know a formula ψ if ψ is true in all the worlds possible for him. Nested modal operators are allowed and intuitively

$K_iK_j \ldots (\varphi)$ means "agent i knows that agent j knows that ... that φ is true." In order to give a semantics to the logic of knowledge, we need a formal way of representing worlds and possibility relations (one for each individual) defined between them; Kripke structures are a good formal tool.

A Kripke structure M, over a set of atomic propositions P, is a (n+2) tuple $\langle W, \pi, R_1, \ldots, R_n \rangle$ where:

- W is a set of states (also called possible worlds);
- $\pi : W \rightarrow 2^P$ is the interpretation function which assigns a truth value to every atomic proposition at every state $w \in W$; $\pi(w, p_i) \in \{1, -1\}$ for each state $w \in W$ and atomic proposition $p_i \in P$.
- $R_i \subseteq W \times W$ is a binary relation (known as accessibility relation) for agent $i \in I$. R_i is read "v is accessible from w for agent i" or "v is $i - accessible$ from w." $(w, v) \in R_i$ holds if and only if agent i cannot distinguish the state of affairs w from the state of affairs v. In other words, if w is the actual state of the world, then agent i would consider v as a possible state of the world.

$(M, w) \models \varphi$ denotes the notion that the formula φ is satisfied by the Kripke structure $M = \langle W, (R_i), \pi \rangle$ at state w. If φ is atomic, $(M, w) \models \varphi$ iff π assigns true to φ at state w. For the test of the formulas, the satisfaction relation $\models$ is defined inductively as follows:

- $(M, w) \models \neg\varphi$ iff $(M, w) \not\models \varphi$

- $(M, w) \models \varphi \wedge \psi$ iff $(M, w) \models \varphi$ and $(M, w) \models \psi$

- $(M, w) \models \varphi \vee \psi$ iff $(M, w) \models \varphi$ or $(M, w) \models \psi$

- $(M, w) \models \varphi \rightarrow \psi$ iff $(M, w) \not\models \varphi$ or $(M, w) \models \psi$

- $(M, w) \models K_i(\varphi)$ iff for all $v \in W$ such that wR_iv, we have $(M, w) \models \varphi$

We could also derive additional properties about knowledge in Kripke structures by imposing some constraints on agent i's accessibility relation R_i. Letting R_i be an equivalence relation ensures that everything known by i is true, and that i knows his own internal knowledge, If R_i is reflexive, transitive, and symmetric (an equivalence relation) we obtain the following for all $w \in W$ and for every formula φ for agent i:

K1. $(M, w) \models K_i(\varphi) \rightarrow \varphi$

K2. $(M, w) \models K_i(\varphi) \rightarrow K_i(K_i(\varphi)))$

K3. $(M, w) \models \neg K_i(\varphi) \rightarrow K_i(\neg K_i(\varphi)))$

These three properties K1-K3 in Kripke structures correspond to I1-I3 in information structures respectively. Kripke structures can be represented by labelled graphs, whose nodes are the states in W, and two nodes w and v are connected by an edge labelled i iff $(w, v) \in R_i$.

16 Appendix B

Beliefs are the products of reasoning and beliefs guide actions. Actions are expected to reach goals if beliefs that guide them are true. Both induction and deduction supply reason to believe each seeks to preserve the truth of its premises while extending them to new truths acquired as beliefs.

16.1 Rational Thought

What is reasoning? Reasoning is the set of processes that enables human beings to go beyond the information given, make sense of things, establish or verify facts, and form beliefs. It is a way by which thinking comes from one idea to a related idea. Adler [6] explains reasoning as a transition in thought, where some beliefs or thoughts provide the ground or reason for coming to another.

From her beliefs that

(1) Either Bob is a tea drinker or a coffee drinker.

and

(2) Bob does not drink tea.

Ann infers that

(3) Bob drinks coffee.

Reasoning in an argument is valid if the argument's conclusion must be true when the premises (reasons given in support of the conclusion) are

true. So assuming Ann bases her inference on the deductive relationship (1) and (2) to (3), her argument is valid since (1) and (2) imply (3). And (3) is a logical consequence of (1) and (2). In reaching (3) Ann comes to a new belief even though its information is entailed by (1) and (2). This is called *deductive* reasoning.

Unlike a deductive argument, an *inductive* argument provides for new beliefs whose information is not entailed by the beliefs from which it is inferred.

(4) Ann brought her book to the class every day of the semester.

So probably

(5) Ann will bring it to the next class.

Inductive reasoning is based on previous observations and the premises only render the truth of the conclusion more probable than in their absence. In inductive reasoning the truth of the premises does not guarantee the truth of the conclusion. So regardless of the above example being a good inductive argument, premise (4) can be true and conclusion (5) false. Therefore, the argument is invalid.

How does reasoning develop? According to Piaget, the twentieth century Swiss psychologist, development of human reasoning occurs in stages. There are four stages identified with Piaget's theory of *cognitive development* of reasoning [10].

The first stage occurs between birth to two years of age and is called the *Sensori-motor*. In this stage, children learn to differentiate self from objects. They start to recognize self as an agent of action and begin to act intentionally

e.g. pulling an object and shaking a rattle to make noise. They realize that things continue to exist even when no longer present to the sense.

The second stage occurs between the ages of two to seven years and is called *Pre-operational*. In this stage, they start to use language and to represent objects by images and words. Thinking is still egocentric so they have difficulty taking the viewpoint of others. They start to classify objects by a single feature. For example, grouping together all the red blocks regardless of shape or all the square blocks regardless of color.

The third stage occurs between the age of seven to eleven years and is called *Concrete-operational*. They start to think logically about objects and events. They can classify objects according to several features and can order them in series along a single dimension such as size.

The fourth and final stage occurs after eleven years of age and is called *Formal-operational*. In this stage, individuals can think logically about abstract propositions and can systematically test hypotheses. They become concerned with the hypothetical, the future, and ideological problems.

16.2 Theories of Reasoning

Psychologists have attempted to study and explain how people reason. Which cognitive processes are engaged in reasoning? How do cultural factors affect the inferences people draw? Can reasoning be modeled computationally? Can animals reason the way human beings do? Researchers have been determined to find which particular formal logic is laid down in the mind and which

rules of inference are used in its mental formulation. In parallel, computer scientists have developed programs that prove arguments based on formal rules of inference. As a result, the research on reasoning has accumulated numerous experimental results and models of human reasoning process.

A majority of these theories fall under logic-based, mental models, and heuristic approaches. Logic-based approaches to deduction have been criticized for being too narrowly focused on classical logic. Probabilistic approaches and mental model theory both provide an alternative to logic-based models. However, they too have their shortcomings.

At the heart of psychological studies on human deductive reasoning lie the topics of selection, suppression, and syllogism.

Selection was originally devised by Wason [153] and has ever since become one of the well studied puzzles in the psychology of reasoning. In Wason's selection task, subjects are presented with a rule and they have to select cases in order to make judgments either about compliance of the cases or about the truth of the rule.

There are different flavors of the selection task and one version is show in Figure 47. In this version, subjects are shown a set of four cards. Each card has a number on one side and a letter on the other side.

The visible faces of the cards show A, B, 2, and 3; subjects are asked which card(s) should be turned over in order to test the truth of the claim that

(6) if a card has an A on one side then it has a 2 on the other side.

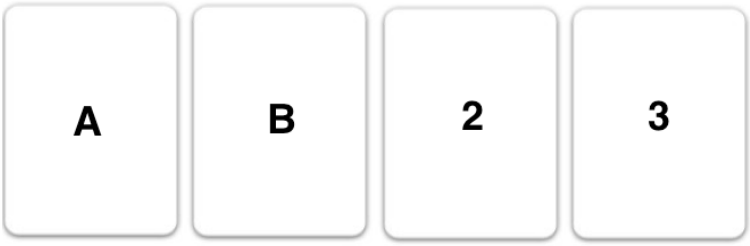


Figure 47: Wason's Selection Task.

Wason discovered that individuals unfamiliar with logic almost always selected the wrong card. For anyone with some formal training in logic, the correct response should be obvious. If you turn over the card showing A and find a number other than 2, then the claim is false. Similarly, if you turn over the card showing 3 and find an A on its other side, the claim is also false. Hence, one needs to select the card showing A or 3. However, subjects rarely select the card showing 3 and often choose the card showing A and maybe 2. If you select the 2 card, then nothing on its other side can show that (6) is false.

The next topic that has been of interest to psychological experiments of reasoning has been *suppression* of modus ponens inferences. It has been argued that background knowledge leads to suppression [22]. Subjects presented with a condition like

(7) If Ann has an essay to write then she studies late in the library.

and the premise

(8) Ann has an essay to write.

make the inference that

(9) Ann studies late in the library.

However, this inference is *suppressed* when there is an additional conditional such as

(10) If the library stays open then Ann studies late in the library.

The other prominent topic that has got quite a bit of attention on psychological studies of reasoning is *Syllogism*. Syllogistic inference is a form of reasoning with *quantifiers* where the conclusion is inferred from two or more premises.

For example,

(11) All men are mortal.

(12) Bob is a man.

Therefore,

(13) Bob is mortal.

The syllogistic language is confined to four sentence types.

1. *All A are B* (universal affirmative)
2. *Some A are B* (particular affirmative)
3. *No A are B* (universal negative)
4. *Some A are not B* (particular negative)

In a majority of experiments on syllogistic reasoning, subjects are given two premises and asked either to choose from a list of possible conclusions or

say if any conclusions followed from the premises. Researchers have also used evaluation tasks, asking subjects to decide whether a given argument is valid or not.

Newstead [88][89] was among the first to study subjects' interpretations of syllogistic inferences, making a connection to Gricean theory of *Implicatures*. His results show that subjects often make inferences that does not logically follow from the premises. For example, when subjects were told to assume, *All A are B*, and then asked wither it followed that *All B are A* must be true, false, or could be either. A majority of subjects did not approximate a classical interpretation of the quantifiers. In similar studies, subjects concluded, *Some A are not B* from the premise *Some A are B*.

This kind of Gricean interpretation is also observed in experiments where subjects were given the premises

(16) Some A are B.

(17) Some B are C.

who concluded that

(18) Some A are C.

The above argument is similar to saying, *some cats are black and some black things are dogs, therefore some cats are dogs.*

In almost all the empirical studies subjects depart from the answer that the experimenter had derived when translating the argument into a logical system and assessing its correctness within the system. This has raised concerns

over the method and whether human beings use logical models or something else when making deductive inferences. This question has been at the center stage for evolutionary psychologists.

Rips[121] argues that changing the deductive rules of a logical system can alter arguments that are deductively correct and psychologists have overlooked this variety assuming a single standard for deductive correctness.

A proof as a finite sequence of sentences $(s_1, s_2, \ldots, s_k)$ in which each sentence is either a premise, an axiom of the logical system, or a sentence that follows from preceding sentences based on specified rules. An argument is deducible in the system if there is a proof whose final sentence, s_k, is the conclusion of the argument.

Consider a system that includes modus ponens among its rules.

(19) If Bob deposits \$1.50 cents then Bob will get a coke.

(20) Bob deposits \$1.50.

(21) Bob will get a coke.

Based on modus ponens rule, (21) is true if the premises (19) and (20) hold and the above argument is deducible in the system. However, Rips claims that blindly applying rules to a problem will not lead to a proof in an acceptable amount of time as some rules can produce infinite sets of irrelevant sentences. Therefore heuristics are important to consider.

Rips presents a theory of sentential reasoning and provides an implementation called PSYCOP (short for Psychology of Proof).

In his theory, Rips merges ideas from logic and computer science. From logic he borrows the idea of suppositions i.e. reasoning involves suppositions or assumptions and people tend to entertain a proposition temporarily in order to trace its consequences. From computer science he adopts the concept of subgoals. People are able to adopt on a temporary basis the desire to prove some proposition in order to achieve a further conclusion. In his view, suppositions are roughly like *provisional beliefs*, and subgoals are roughly like *provisional desires*. According to Rips, beliefs and desires about external states guide external actions while provisional beliefs and provisional desires guide internal actions in reasoning.

His basic inference system consists of a set of deduction rules that construct a proof in the systems working memory. Upon presenting the system with a group of premises, it will use the given rules to generate proofs of possible conclusions. The system first stores the input premises in working memory. It then applies the rules on memory contents in order to determine whether any inference is possible. If so, the newly deduced sentence is added to memory. It then scans the updated configuration, makes further deductions, and so on until a proof has been found or no further rules remain.

The implementation PSYCOP is developed using Prolog program for personal computers. The program model has a standard memory architecture that is divided into long term and working memory with later having smaller capacity. While evaluating an argument, the program begins by applying its forward rules to the premises until no new inferences are forthcoming. It then considers the conclusion of the argument, checking to see whether the

conclusion is already among the assertions. If so the proof is complete, if not, it will treat the conclusion as a goal and attempt to apply the backward rules.

Johnson-Laird [86] argues that empirical studies analyzing everyday arguments have proven that it is extremely difficult to translate arguments into formal logic. Unlike logic, the interpretation of sentences in daily life is often modulated by knowledge.

For example,

(22) If Bob is in Rio de Janeiro then he is in Brazil.

and

(23) Bob is not in Brazil.

then

(24) Bob is not in Rio de Janeiro.

Based on their background knowledge that Rio de Janeiro is in Brazil, subjects inferred (24).

Therefore, a good theory of reasoning must allow for such effects. He argues that the system for interpreting sentences cannot work in truth functional way and must take meaning and knowledge into account.

An alternative to pure logic based and heuristic approaches is the theory of *mental models* or *model theory.* The model theory was originally developed by Johnson-Laird and Byrne [75747574] and is built on the assumption that reasoning is about possibilities. Human beings have difficulty thinking about more than one possibility at a time. Working memory, which holds models in

mind, is limited in its capacity. Therefore, reasoning that is based on models of possibilities, where each mental model represents what is common to a possibility, seems reasonable.

For example, when Ann says

(25) My house is in the middle of the street.

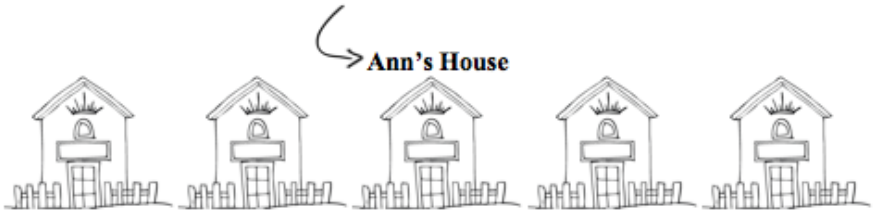


Figure 48: An diagram compatible with statement (25).

We construct a mental model of a single possibility even though the proposition expressed by (25) could be true in many ways. Thus (25) maps to a scene (Figure 48) where Ann's house is roughly in the middle of the street rather than toward one end or the other.

It is well established that humans beings cannot hold an infinitude of possibilities while working out an argument. Mental models lighten the load on working memory by representing less information. The mental model of (25) captures what is common to different possibilities keeping in mind that human beings tend to think about possibilities one model at a time.

He argues that semantic and pragmatic modulation affect the interpretation of sentences so they cannot be treated as strictly truth functional.

For example, consider the following premises

(26) The cup is to the right of the saucer.

(27) The spoon is to the left of the saucer.

A diagram of the possibility compatible with premises (26) and (27) is shown in Figure 49.

Figure 49: A diagram compatible with statements (26) and (27).

The diagram shows that the cup is to the right of the spoon, and this conclusion follows from the premises but it is not asserted in them. In this case, the the diagram has a spatial interpretation i.e. the position of objects in the diagram corresponds to the scene.

An interesting question that arises is, *how does the principle of truth fit in this theory*? Johnson-Laird suggests that the right way to think about the principle of truth is to think of mental models representing only those states of affairs that are possible given an assertion. Mental models represent clauses in the premises only when they are true in all possibilities. Additionally, if individuals retain mental footnotes about what is false then they can flush out mental models into fully explicit models representing both what is true and what is false. The model theory does not abandon logic entirely but relates to logic in the sense that an inference is valid if there are no counter examples to its conclusion. A disadvantage of this model is over-simplification of possibilities. This relates to Schelling's [129] concept of *focal points* which

is a way to narrow down possible solutions in a coordination problem.

So what is the nature of mental representations underlying deduction; is it rules or is it models?

Stenning and Lambalgen [6] argue that the search for a human reasoning mechanism through the tasks of selection, suppression, and syllogism has employed a narrow hypothesis testing methodology. It has ignored the support available from modern logical semantic and pragmatic methods and instead targeted its criticism on an inappropriate classical logic. Rejecting logic has led to attempts to re-invent it producing some hard to interpret systems. They argue that Psychologists have focused their research in the wrong direction; great emphasis has been given on studying representation but the field has pretty much ignored interpretation.

They argue that the mental processes evoked in these experiments are interpretative processes; the processes of reasoning to interpretation.

Most of the experiments carried out by psychologists force interpretation in a vacuum. Wason's selection task is an interesting example where recent experiments on subjects reveal that the underlying problem is due to removing the normal cues on which the choice of interpretation depends. People find Wason's selection task much easier if it is placed in a social context.

Consider a different version of the selection task (shown in Figure 50). You are at a bar and your job is to ensure that people obey the rule

(29) If a person is drinking beer then (s)he must be at least 18 years old.

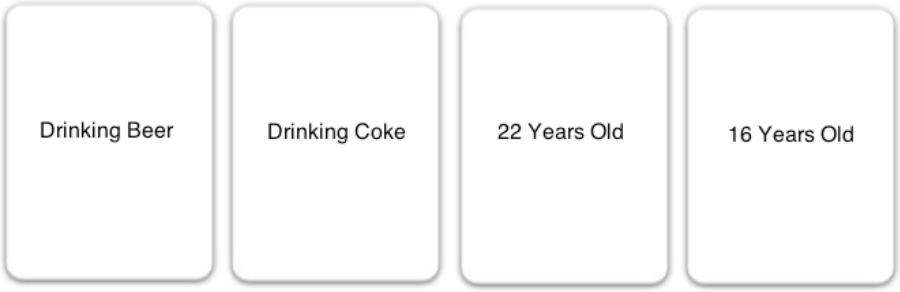


Figure 50: A different version of Wason's Selection Task.

In this version (Figure 50), subjects are shown a set of four cards. Each card has the person's age on one side and what they are drinking on the other side. The visible faces of the cards show *Drinking Beer*, *Drinking Coke*, *22 Years Old*, and *16 Years Old*. Subjects are asked which card(s) should be turned over in order to test the truth of (29). That is, which card(s) should be turned over in order to determine whether or not they are breaking the rule?

The results show that subjects tend to select the correct cards i.e. the cards showing *Drinking Beer* and *16 years old*.

To take interpretation seriously one must take individual differences seriously. Subjects do different things in experiments and this point has been overlooked. Human reasoners take their knowledge into account and often go beyond the information given (i.e. step into inductive reasoning). Stenning and Lambalgen believe the only way out of the confusion is to take interpretation seriously and separate semantics from representational issues.

Geurts [61] focuses his studies on syllogism and argues that despite decades of psychological research on syllogistic reasoning and numerous experimental results, the empirical base has been narrow. He argues that any

psychological account of syllogistic reasoning needs to follow from an adequate theory of interpretation.

Theories about syllogistic reasoning proposed over the years run into problems with certain extensions of the syllogistic language. Geurts claims that current approaches to syllogistic reasoning are based on representational models which encode quantified statements in terms of individuals. These representations are limited in dealing with statements e.g., *Most A are B*, *At least three A are B*, etc. Scientists in the field have not done any studies on cardinal quantifiers (e.g. *five*, *at least six*, *at most seven*, etc), the role of negation in syllogistic reasoning, arguments with multiple quantifiers, and so on.

For example,

(30) At least half of the foresters are vegetarians.

This states that the set of foresters who are vegetarians is not smaller than the set of foresters who aren't. And since first order predicate logic allows us to talk about individuals, it is not expressive enough for representing a sentence like (30).

A system of inference that deals with quantifiers in terms of arbitrary individuals cannot handle arguments such as,

(31) All vegetarians are teetotallers.

(32) Most foresters are vegetarians.

Therefore,

(33) Most foresters are teetotallers.

Even if a quantifier is expressible in predicate logic, the representations involved may not be suited for psychological purposes.

For example,

(34) At least two foresters are teetotallers.

can be expressed in predicate logic as,

(35) $\exists x\ \exists y$ [x≠y & forester(x) & teetotaller(x) & forester(y) & teetotaller(y)]

This is a rather cumbersome representation. Since predicate logic doesn't offer the means for talking about sets, it requires the introduction of two individual variables and specification that their values are distinct and that both variables stand for a forester as well as a teetotaller.

Geurts claims that the current models of syllogistic reasoning are all ad-hoc from the point of view of language understanding. They are incapable of capturing non-standard quantifiers because in predicate logic one cannot talk and reason about sets. Therefore, it is impossible to represent proportional quantified, such as *most* and *at least half of*, etc. Solving a syllogistic argument calls for an interpretation of quantified sentences.

Mental model theory developed by Johnson-Laird et al., runs into the same problems as logic-based theories because again quantified propositions are represented in terms of individuals.

For example,

(36) Two A are B.

How can we represent (36) in a mental model?

Since predicate logic and mental-model theory are both individual-based systems, they get into the same trouble with non-standard quantifiers. First, *All A are B* is not synonymous with *Two A are B*. Second, if it takes two individuals to represent *two*, then it takes sixty individuals to represent *sixty*, which gets us back to the same problem discussed in connection with predicate-logical representations of cardinalities. Guerts believes that despite going through many revisions, the mental model theory is still not expressive in terms of reasoning with quantified sentences.

A different way of dealing with quantification is Charter and Oaksford's [24] probabilistic semantics which underlies their probability heuristics model of syllogistic reasoning. According to Charter and Oaksford, humans are geared towards reasoning with uncertainty. They are designed by evolution to reason not logically but probabilistically. This account calls for a probabilistic interpretation of quantified expressions.

For example,

(37) *All A are B.*

This probabilistically speaking means, that P(B|A) = 1 i.e., the conditional probability of B given A equals 1.

Similarly,

(38) *No A are B.*

Conveys that P(B|A) = 0, and

(39) *Some A are B.*

Conveys that P(B|A) > 0.

If the conditional probability of the conclusion is 1, a proposition with *all* can be inferred. The probabilistic approach can afford a representation of proportional quantifiers, such as *most*. According to Charter and Oaksford's denition,

(40) *Most A are B.*

means that P(B|A) is high but less than 1.

In this respect, a probabilistic semantics is more expressive than other approaches but still not expressive enough. In general, propositions involving cardinal quantifiers cannot be translated into a probabilistic format.

For example, if it is given that

(41) *Two A are B.*

We do not know what P(B|A) is unless it is also known how many A's there are. One proposal is that (41) should mean that P (B|A)= 2/|A| (where |A| stands for the cardinality of the set of As). Thus, if there are five vegetarians altogether,

(42) Two vegetarians are liberals.

means that there is a 0.4 probability that a given vegetarian is a liberal. This proposal runs into problems, the most obvious one being that it suffices for (42) to be true that there are two liberal vegetarians; the total number of

vegetarians is irrelevant.

In short, the probabilistic account leads to the claim that all quantifiers are proportional, which is unintuitive for quantifiers like *some*, and false for others like the cardinals. It is not just logic-based approaches that suffer from these problems but all theories of reasoning run into the same issue.

Geurt believes logic-based approaches to deduction are more powerful than others; limitations being quantifiers, such as *most* and *at least half of* are not expressible in standard predicate logic. He feels the right way to deal with representational shortcomings in logic-based models is to consider an approach based on sets rather than individuals. He presents a logic-based model of syllogistic reasoning motivated by semantical considerations and dropping the assumption that syllogistic reasoning is always in terms of individuals.

References

[1] Definition of trust. http://oxforddictionaries.com/definition/trust. Oxford Dictionaries Online. Retrieved March 12, 2012.

[2] Linkedin Corp. http://www.google.com/finance?q=NYSE:LNKD&fstype=ii. Google Finance. Retrieved Febuary 1, 2012.

[3] Facebook's Filing: The Highlights. http://bits.blogs.nytimes.com/2012/02/01/facebooks-filing-the-highlights, Febuary 1 2012. The New York Times. Retrieved Febuary 20, 2012.

[4] Form S-1 Registration Statement Facebook, Inc. http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm, Febuary 1 2012. U.S. Securities and Exchange Commission. Retrieved February 21, 2012.

[5] Online Data Helping Campaigns Customize Ads. http://www.nytimes.com/2012/02/21/us/politics/campaigns-use-microtargeting-to-attract-supporters.html?_r=1&ref=todayspaper, Febuary 21 2012. The New York Times. Retrieved Febuary 21, 2012.

[6] Jonathan E. Adler and Lance J. Rips. *Reasoning: Studies of Human Inference and its Foundations*. Cambridge University Press, 2008.

[7] Rachel F. Adler and Farishta Satari. The Application of Virtual Reality Simulations to the Treatment of Anxiety Disorders. *Decision Sciences Institute. San Antonio, TX*, 2006.

[8] Keiko Aoki, Kenju Akai, and Kenta Onoshiro. Deception and confession: Experimental evidence from a deception game in japan. Technical report, Institute of Social and Economic Research, Osaka University, 2010.

[9] Michael Arrington. Twitter's Financial Forecast Shows First Revenue in Q3, 1 billion users in 2013. `http://techcrunch.com/2009/07/15/twitters-financial-forecast-shows-first-revenue-in-q3-1-billion-users-in-2013/`, July 15 2009. TechCrunch. Retrieved Febuary 20, 2012.

[10] James S Atherton. Piaget's Theory of Cognitive Denvelopment. `http://www.learningandteaching.info/learning/piaget.htm`, Dec 2011.

[11] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Computational Representation of Practical Argument. *Synthese*, 152(2):157–206, 2006.

[12] Robert J. Aumann. Agreeing to Disagree. *Institute of Mathematical Statistics (Institute of Mathematical Statistics)*, 4(6):1236–1239, 1976.

[13] John Langshaw Austin. *How to do things with Words: The William James Lectures delivered at Harvard University in 1955.* Ed. J. O. Urmson. Oxford: Clarendon, 1955.

[14] Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki. The Logic of Public Announcements, Common Knowledge, and Private Suspicions. In *TARK 1998: Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge*, pages 43–56, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[15] Johan Van Benthem, Jelle Gerbrandy, and Barteld Kooi. Dynamic update with Probabilities. In *ILLC Prepublication*, 2006.

[16] Anton Benz. Questions, plans, and the utility of Answers. *Syddansk Universitet, Kolding*, 2006.

[17] Anton Benz, Gerhard Jaeger, and Robert Van Rooij. *An Introduction to Game Theory for Linguists*. Palgrave MacMillan, New York, 2005.

[18] Michael Blome-Tillmann. Conversational Implicatures (and How to Spot Them). *Philosophy Compass*, 8(2):170–185, 2013.

[19] Nancy Bonvillain. *Language, Culture, and Communication The Meaning of Messages*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 2000.

[20] Juergen Bracht and Nick Feltovich. Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics*, 93:1036–1044, 2009.

[21] Steven J. Brams. *The Presidential Election Game*. Yale University Press, New Haven and London, 1978.

[22] Ruth M. J. Byrne, Orlando Espino, and Carlos Santamaria. Counterexamples and the suppression of Inferences. *Journal of Memory Language*, 40:347–373, 1999.

[23] Sugato Chakravarty, Yongjin Ma, and Sandra Maximiano. Lying and Friendship. Technical Report 1007, Purdue University, Department of Consumer Sciences, 2011.

[24] N Chater and M Oaksford. The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, 38:191–258, 1999.

[25] Ying Chen, Joel Sobel, Ying Chen, Navin Kartik, and Joel Sobel. Selecting Cheap Talk Equilibria. *Review of Economic Studies*, 2008.

[26] Herbert H. Clark and Thomas B. Carlson. Hearers and Speech Acts. *JSTOR*, 58:332–373, 1982.

[27] Herbert H. Clark and Edward F. Schaefer. Concealing One's Meaning from Overhearers. *Journal of Memory and Language*, 26:209–225, 1987.

[28] Herbert H. Clark and Edward F. Schaefer. *Arenas of Language Use: Chapter 8 Dealing with Overhearers*. University of Chicago Press, 1992.

[29] Vincent P. Crawford, Uri Gneezy, and Yuval Rottenstreich. The Power of Focal Points is Limited: Even Minutes of Payoff Asymmetry May Yield Large Coordination Failures. *American Economic Review*, 2008.

[30] Vincent P. Crawford and Joel Sobel. Strategic Information Transmission. *Econometrica*, 50(6):1431–1451, 1982.

[31] Robin P. Cubitt and Robert Sugden. Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory. *Economics and Philosophy*, 19:175–210, 2003.

[32] Robert Dale. Cooking Up Referring Expressions. *In Proceedings of the 27th Annual Meeting of the Association of Computational Linguistics, University of British Columbia, Vancouver*, 1989.

[33] Robert Dale and Ehud Reiter. Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 19, 1995.

[34] Donald Davidson. Truth and Meaning. *Synthese*, 17, 1967.

[35] Donald Davidson. On Saying That. *Synthese*, 19, 1968.

[36] Donald Davidson. Belief and the Basis of Meaning. *Synthese*, 27(3-4):309–323, 1974.

[37] Donald Davidson. *Moods and Performances*. Springer Netherlands, 1979.

[38] Munindar Singh Department, Munindar P. Singh, Ashok Mallya, Mike Maximilien, and Raghu Sreenath. The Pragmatic Web: Preliminary Thoughts. In *In Proc. of the NSF-EU Workshop on Database and Information Systems Research for Semantic Web and Enterprises, April 3-5, Amicalolo Falls and State*, 2004.

[39] Hans Van. Ditmarsch and Barteld Kooi. The Secret of My Success. *Synthese*, pages 201–232, 2006.

[40] Keith Donnellan. Reference and Definite Descriptions. *in Philosophical Review*, pages 281–304, 1966.

[41] Claire Doutrelant, Peter McGregor, and Rui Oliveirab. The effect of an audience on intrasexual communication in male Siamese fighting fish, Betta splendens. *Behavioral Ecology*, 12:283–286, 2001.

[42] Iddo Samet Dov Samet and David Schmeidler. One Observation behind Two-Envelope Puzzles. *The American Mathematical Monthly*, 111(4):347–351, 2004.

[43] F.I. Dretske. *Knowledge and the Flow of Information*. MIT Press, 1983.

[44] Michael Dummett. What is a theory of meaning? (ii). In *The Seas of Language*. Oxford University Press, 1993.

[45] Peter Eavis and Evelyn M. Rusli. Investors Get the Chance to Assess Facebook's Potential. `http://dealbook.nytimes.com/2012/02/01/investors-get-the-chance-to-assess-facebooks-potential`, Febuary 1 2012. The New York Times. Retrieved Febuary 21, 2012.

[46] Zachary Ernst. What is Common Knowledge? *Episteme*, 8(3):209–226, 2011.

[47] Gareth Evans. The Causal Theory of Names. *in Martinich, A. P. ed. The Philosophy of Language. Oxford University Press.*, 1985.

[48] Joseph Farrell and Robert Gibbons. Cheap Talk with Two Audiences. *The American Economic Review*, 79(5):1214–1223, 1989.

[49] Joseph Farrell and Matthew Rabin. Cheap Talk. *Journal of Economic Perspectives*, 10(3):103–18, 1996.

[50] Keith Ferrazzi. *Who's got your back*. Random House Digital, Inc., 2009.

[51] Stanley Fish. Talking to No Purpose. `http://opinionator.blogs.nytimes.com/2011/04/04/talking-to-no-purpose`, April 4 2011. The New York Times. Retrieved April 10, 2011.

[52] Melvin Fitting. Reasoning About Games. *Studia Logica*, 82:1–25, 2006.

[53] Luciano Floridi. *Information: a very short introduction*. Oxford University Press, 2010.

[54] Michael Franke and Robert Van Rooij. Strategies of Persuation, Manipulation, and Propaganda: psychological and social aspects. 2013.

[55] Gottlob Frege. On Sense and Reference. *in Translations from the Philosophical Writings of Gottlob Frege, edited by Peter Geach and Max Black*, pages 58–70, 1960.

[56] Francis Fukuyama. *Trust*. Free Press Paperbacks EditionSimon, 1996.

[57] Bob Garfield. *The Chaos Scenario*. Stielstra Publishing, 2009.

[58] John Geanakopolos. Common Knowledge. *Journal of Economic Perspectives*, 6(4):53–82, 1992.

[59] Barton Gellman, Aaron Blake, and Greg Miller. Edward Snowden comes forward as source of NSA leaks. `http://www.washingtonpost.com/`

politics/intelligence-leaders-push-back-on-leakers-media/2013/06/09/fff80160-d122-11e2-a73e-826d299ff459_story.html. The Washington Post. Retrieved June 10, 2013.

[60] Jelle Gerbrandy. Communication Strategies in Games. *Journal of Applied Non-Classical Logics*, 17, 2006.

[61] Bart Geurts. Reasoning with quantifiers. *Cognition*, 86(3):223–51, 2003.

[62] Uri Gneezy. Deception: The Role of Consequences. *American Economic Review*, 95(1):384–394, 2005.

[63] Paul H. Grice. *Studies in the Way of Words*. Harvard University Press, Cambridge, Massachusetts, 1989.

[64] Barbara Grosz and Candace Sidner. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, Volume 12, number 3, 1986.

[65] Hans Peter Gruner and Alexandra Kiel. Collective decisions with interdependent valuations. *European Economic Review*, 48(5):1147–1168, 2004.

[66] Ulrike Hahn and Mike Oaksford. The Rationality of Informal Argumentation: A Bayesian Approach to Reasoning Fallacies. *Synthese*, 152(2):207–236, 2006.

[67] Thomas Hobbes. *Leviathan (Oxford World Classics)*. Edited by J.C.A Gaskin. Oxford University Press, 1996.

[68] Sjaak Hurkens and Navin Kartik. (When) Would I Lie To You? Comment on "Deception: The Role of Consequences". 2006.

[69] Sjaak Hurkens and Navin Kartik. Would I lie to you? on social preferences and lying aversion. *Experimental Economics*, 2009.

[70] Jameel Jaffer. Secrecy and Freedom. `http://www.nytimes.com/roomfordebate/2013/06/09/is-the-nsa-surveillance-threat-real-or-imagined?partner=rss&emc=rss`. The New York Times. Retrieved June 10, 2013.

[71] Gerhard Jager. Game dynamics connects semantics and pragmatics. *University of Bielefeld*, 2006.

[72] Gerhard Jager. Game theory in semantics and pragmatics. *University of Bielefeld*, 2008.

[73] Adrianne Jeffries. As Banks Start Nosing Around Facebook and Twitter, the Wrong Friends Might Just Sink Your Credit. `http://betabeat.com/2011/12/as-banks-start-nosing-around-facebook-and-twitter-the-wrong-friends-might-just-sink-your-credit/`, December 13 2011. BetaBeat. Retrieved Febuary 20, 2012.

[74] Philip Johnson-Laird. Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness. *Cambridge, MA: Harvard University Press*, 1983.

[75] Philip N. Johnson-Laird and Ruth M. J. Byrne. Deduction. *Hillsdale, NJ: Erlbaum*, 1991.

[76] David Kaplan. Demonstratives. *in I. Almog et al. (eds.), Themes from Kaplan, Oxford University Press.*, pages 481–563, 1985.

[77] David Kaplan. Dthat. *Syntax and Semantics*, 9, 1989.

[78] Edi Karni. Subjective expected utility theory with costly actions. *Games and Economic Behavior*, 50(1):28–41, 2005.

[79] Navin Kartik. Strategic Communication with Lying Costs. *Review of Economic Studies*, 2009.

[80] Yeon koo Che and Navin Kartik. Opinions as Incentives. *Review of Economic Studies*, 2008.

[81] Saul A. Kripke. *Naming and necessity.* Blackwell Publishing, 1981.

[82] Saul A. Kripke. *'Identity and necessity' In Metaphysics: An Anthology. Edited by Jaegwon Kim, and Ernest Sosa. Malden, MA.* Blackwell Publishing, 1999.

[83] David Lewis. *Convention: A Philosophical Study.* Harvard University Press, Cambridge, Mass., 1969.

[84] Arthur Merin. Information, relevance, and social decision making. *In L. Moss, J. Ginzburg, and M. de Rijke, editors, Logic, Language, and Computation*, 2, 1999.

[85] Claire Cain Miller and Brad Stone. Hacker Exposes Private Twitter Documents. http://bits.blogs.nytimes.com/2009/07/15/hacker-exposes-private-twitter-documents/?hpw., July 15 2009. The New York Times. Retrieved Febuary 20, 2012.

[86] Philip N. and Johnson-Laird. Mental Models and Deduction. *Trends in Cognitive Sciences*, 5(10):434–442, 2001.

[87] Stephen Neale. Paul Grice and the Philosophy of Language. *Review of Paul Grice, Studies in the Ways of Words Cambridge, Mass.: Harvard University Press,*, 1989.

[88] Steve Newstead. Interpretational errors in syllogistic reasoning. *Journal of Memory and Language*, 28:78–91, 1989.

[89] Steve Newstead. Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language*, 34:644–664, 1995.

[90] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[91] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT Press, 1999.

[92] Eric Pacuit, Rohit Parikh, and Eva Cogan. The Logic of Knowledge Based Obligation. *Synthese*, 149, 2006.

[93] Prashant Parikh. *Pragmatics and Games of Partial Information.* In A. Benz, G. Jager, & R. van Rooij (eds.), Game Theory and Pragmatics, pp. 83-100. Palgrave Macmillan, Basingstoke, 2006.

[94] Prashant Parikh. *Language and Equilibrium.* The MIT Press, 2010.

[95] Rohit Parikh. Finite and Infinite Dialogues. *in the Proceedings of a Workshop on Logic from Computer Science, MSRI publications, Springer*, pages 481–498, 1991.

[96] Rohit Parikh. Social Software. *Synthese*, 132:187–211, 2002.

[97] Rohit Parikh. Sentences, Propositions and Logical Omniscience, or What does Deduction tell us? *City University Of New York*, 2007.

[98] Rohit Parikh. Some Puzzles About Probability and Probabilistic Conditionals. *Symposium on Logical Foudnations of Computer Science*, 4514/2007:449–456, 2007.

[99] Rohit Parikh and Paul Krasucki. Communication, Consensus, and Knowledge. *Journal of Economic Theory*, 52(1):178–189, 1990.

[100] Rohit Parikh and Ramaswamy Ramanujam. A Knowledge Based Semantics of Messages. *J. of Logic, Lang. and Inf.*, 12(4):453–467, 2003.

[101] John Perry. Frege on Demonstratives. *Philosophical Review*, 86:474–497, 1977.

[102] John Perry. The Problem of the Essential Indexical. *Nous*, 13(1):3–21, 1979.

[103] John Perry. The Prince and the Phone Booth: Reporting Puzzling Beliefs. *Journal of Philosophy*, 86:685–711, 1986.

[104] Ahti Veikko Pietarinen, editor. *Game Theory and Linguistic Meaning*. Current Research in the Semantics / Pragmatics Interface. Elsevier Ltd, Oxford, 2007.

[105] Steven Pinker. *The Stuff of Thought: Language as a Window into Human Nature*. Penguin, 2007.

[106] Steven Pinker, Martin Nowak, and James Lee. The logic of indirect speech. *Proceedings of the National Academy of Sciences of the USA*, 105(3):833–838, 2008.

[107] Plato. Cratylus. Technical report, Trans. C. D. C. Reeve. In Complete Works. Ed. John Cooper., 1997.

[108] Jan A. Plaza. Logics of Public Communications. *in: M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, Z. W. Ras (Eds.), Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems: Poster Session Program*, pages 201–216, 1989.

[109] Massimo Poesio, Rosemary Stevenson, Barbara Di Eugenio, and Janet Hitzeman. Centering: A Parametric Theory and Its Instantiations. *Computational Linguistics*, Volumne 30, Number 3, 2004.

[110] Willard Quine. Two Dogmas of Empiricism. *Philosophical Review*, 60(1):20–43, 1951.

[111] Willard Quine. Meaning and Translation. *Brower, R. (ed.), On Translation, Cambridge Mass.*, pages 148–172, 1959.

[112] Howard Rachlin. Notes on Discounting. *Journal of the Experimental Analysis of Behavior*, 85(3):425–435, 2006.

[113] Howard Rachlin and Bryan Jones. Social Discounting. *Psychological Science*, 17(4):283–286(4), 2006.

[114] Howard Rachlin and Bryan Jones. Altruism among relatives and non-relatives. *Behavioural Processes*, 79(1):120–123, 2008.

[115] Howard Rachlin and Matthew Locey. A behavioral analysis of altruism. *Behavioural Processes*, 87(1):25–33, 2011.

[116] Eric Rasmusen. *Game and Information: An Introduction to Game Theory.* Blackwell, Cambridge, MA, USA & Oxford, UK, 1st edition, 1990.

[117] Eric Rasmusen. *Games and Information An Introduction to Game Theory.* Wiley-Blackwell, 4th edition, 2006.

[118] Ehud Reiter and Robert Dale. A Fast Algorithm for the Generation of Referring Expressions. *Proceedings of COLING-92, Nantes*, 1992.

[119] Philip J. Reny. Arrow's theorem and the Gibbard-Satterthwaite theorem: a unified approach. *Economics Letters*, 70(1):99–105, 2001.

[120] Alexander Repenning and James Sullivan. The Pragmatic Web: Agent-Based Multimodal Web Interaction with no Browser in Sight. *Human-Computer Interaction – INTERACT 2003*, pages 212–219, 1973.

[121] Lance J. Rips. *The psychology of proof: Deduction in human thinking.* Cambridge, MA: MIT Press, 1994.

[122] Alvin E. Roth, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *The American Economic Review*, 81(5):1068–1095, 1991.

[123] Bertrand Russell. On Denoting. *Mind*, 14:479–493, 1905.

[124] Bertrand Russell. Descriptions. *in Russell's Introduction to Mathematical Philosophy*, 1919.

[125] Samer Salame, Eric Pacuit, and Rohit Parikh. Some Results on Adjusted Winner. *Synthese*, 2005.

[126] David Sally. Can I say "bobobo" and mean "There's no such thing as cheap talk"? *Journal of Economic Behavior & Organization*, 57:245–266, 2005.

[127] Leonard J Savage. *The Foundations of Statistics.* John Wiley and Sons, New York, 1954.

[128] T C Schelling. *Micromotives and Macrobehavior.* Norton, 1978.

[129] Thomas Crombie Schelling. *The Strategy of Conflict.* Harvard University Press, Cambridge, Massachusetts, 1960.

[130] Michael F. Schober and Herbert H. Clark. Understanding by Addresses and Overhearers. *Cognitive Psychology*, 21:211–232, 1989.

[131] John Searle. *A Taxonomy of Illocutionary Acts*, pages 334–369. University of Minnesota Press, Minneapolis, 1975.

[132] John R. Searle. Proper Names. *Mind*, 67(266):166–173, 1958.

[133] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1964.

[134] Brian Skyrms. *Signals Evolution, Learning, Information*. Oxford University Press, 2010.

[135] Raymond M. Smullyan. *First-Order Logic*. Dover Publications, New York, 1995.

[136] Scott Soames. Truth, Meaning, and Understanding. *Philosophical Studies*, 65:17–35, 1992.

[137] Edit Staff. 10 ways big data changes everything. `http://gigaom.com/2012/03/11/10-ways-big-data-is-changing-everything/6/`. GigaOM. Retrieved March 14, 2012.

[138] Robert Stalnaker. *Saying and Meaning, Cheap Talk and Credibility*. In A. Benz, G. Jager, & R. van Rooij (eds.), Game Theory and Pragmatics, pp. 83-100. Palgrave Macmillan, Basingstoke, 2006.

[139] Matthew Stone. Specifying Generation of Referring Expressions by Example, 2003.

[140] Peter Frederick Strawson. On Referring. *Mind*, 1950.

[141] Chisato Takahashi, Toshio Yamagishi, James Liu, Feixue Wang, Yicheng Lin, and Szihsien Yu. The intercultural trust paradigm: Studying joint cultural interaction and social exchange in real time over the Internet. *International Journal of Intercultural Relations*, 32:215–228, 2008.

[142] Deborah Tannen. *You Just Don't Understand: Women and Men in Conversation*. Harper Collins Publishers, NY, New York, 1991.

[143] Alfred Tarski. The Semantic Conception of Truth: And the Foundations of Semantics. *Philosophy and Phenomenological Research*, 1944.

[144] Alfred Tarski. *Logic, Semantics, Metamathematics*. Oxford at the Clarendon Press; 1st edition, 1956.

[145] Chris Taylor. Social networking 'utopia' isn't coming. `http://articles.cnn.com/2011-06-27/tech/limits.social.networking.taylor_1_twitter-users-facebook-friends-connections?_s=PM:TECH`, June 27 2011. CNN. Retrieved Febuary 21, 2011.

[146] Gordon P. Thomas. Mutual Knowledge: A Theoretical Basis for Analyzing Audience. *College English*, 48(6):580–594, 1986.

[147] By The New York Times. Daily Report: Dismay in Silicon Valley at N.S.A.'s Prism Project. `http://bits.blogs.nytimes.com/2013/06/10/daily-report-dismay-in-silicon-valley-at-n-s-a-s-prism-project/`. The New York Times. Retrieved June 10, 2013.

[148] Michael Tomasello. How Are Humans Unique? http://www.nytimes.com/2008/05/25/magazine/25wwln-essay-t.html?_r=1&scp=2&sq=michael+tomasello&st=nyt. The New York Times. Retrieved March 15, 2012.

[149] Michael Tomasello. *Why We Cooperate*. MIT Press, 2009.

[150] Manuel Valdes and Shannon McFarland. Job seekers getting asked for Facebook passwords. http://news.yahoo.com/job-seekers-getting-asked-facebook-passwords-071251682.html. http://news.yahoo.com. Retrieved March 23, 2012.

[151] John Kerry Video. Kerry accuses Romney of flip-flopping. http://www.cnn.com/video/#/video/politics/2012/09/07/dnc-bts-kerry-bin-laden-better-off.cnn. www.cnn.com. Retrieved September 6, 2012.

[152] Douglas Walton and David M. Godden. The Impact of Argumentation on Artificial Intelligence. *in Considering Pragma-Dialectics, ed. Peter Houtlosser and Agnes van Rees*, pages 287–299, 2006.

[153] Peter Cathcart Wason. *Reasoning*. In Foss, B. M.. New horizons in psychology. Harmondsworth, Middx: Penguin, 1966.

[154] Paul Weirch. Interactive Epistemology. *Episteme*, 8(3):201–208, 2011.

[155] Rebecca S. Wheeler. *The Working of Language From Prescriptions to Perspectives*. Praeger Publishers, Westport, Connecticut, 1999.

[156] Ludwig Wittgenstein. *Philosophical Investigations*. Blackwell Publishing, 1953.

[157] Andrew F. Wood and Matthew J. Smith. *Online Communication: Linking Technology, Identity, and Culture (2 Edition)*. Routledge, 2004.

[158] Michael Wooldridge. *An Introduction to Multiagent Systems*. John Wiley & Sons, Chichester, England, 2002.

[159] M Wout and A.G. Sanfey. Friend or foe: The effect of implicit trustworthiness judgements in social decision-making. *Cognition*, 108:796–803, 2008.