On signalling games in improvised theatre

Lochlan Morrissey

1 Overview

This thesis offers a descriptive game theoretic model of a particular style of improvised theatrical performance. Its purpose is to provide a formal framework for the analysis of individual scenes of so-called **longform** improvisation in which actors are seen as rational agents who select utterance and interpretation strategies that afford them the greatest utility. An actor's repertoire of linguistic strategies spans the entire linguistic faculty, for it is almost solely through the performance and interpretation of utterances that actors are able to execute performances. Executing an improvised performance requires planning of the plot's trajectory, which necessarily take place during the performance. So, actors must produce utterances that signal this planning information while concealing it from the audience, for such information damages the performance's verisimilitude.

To build the game model of improvised theatre, I use a dataset constituted of four performances of longform improvisation. The model aims to account for the strategies that the actors use in the course of these performances. As such, it is *descriptive*, rather than predictive, insofar as it provides a formal means of scaffolding instances of a particular genre of speech, instead of specifying a game model that aims to predict agent behaviour in conversations, and that can be applied to conversations generally. This methodology *inverts* the standard approach to devising game models, in which introspective reflection informs the specification of basic parameters, and these parameters are tested against data [see Fra11]. My approach in this thesis is to first examine a dataset of improvised theatre, and base the parameters on these observations.

I make the following three major claims in the course of the thesis: **A**. That common ground in improvised theatre is built through iterative processes of negotiation. **B**. That for every possible combination of the parties to an improvised performance, multiple, divergent *partial grounds* are generated. **C**. That actors consciously exploit differences between the partial grounds to transmit information which is inaccessible to the audience.

The game model draws on existing work in linguistic game theory and extends existing game theoretic models in four significant ways: (a) it introduces a formal, structured model of discourse, which is built iteratively by a conversation's interlocutors; (b) it provides a means of modelling repeated interactions between the same agents, and the emergent effects that arise in such interactions; (c) it specifies conditions of epistemic access to background information, and in particular describes how divergences in background information lead to differences in interpretation; and (iv) it proposes a model of **natural language (NL) steganography**, which allows for an information theoretic model of hiding information in natural language utterances. In formulating these extensions, I examine three primary communicative situations of improvised theatre: the establishment and continued construction of a plot-world (in chapter 5); negotiation of actors' preferred plot-trajectories (in chapter 6); and concealment of planning information (in chapter 7). I identify heuristics that describe optimality with respect to the purpose of a particular strategy, which capture the linguistic norms of improvised performance. I end with an application of the model to a scene in chapter 8.

2 Improvised theatre

Improvised theatre is understood as *a form of theatre in which composition of the performance's dialogue converges with its execution.* A particularly popular form of improvised theatre is the **Harold** [see HCJ94], which is characterised chiefly in two ways. First, a Harold follows a codified structure of nine scenes, partitioned into three groups, with short intermezzi between each of these groups. Second, each Harold begins with an audience suggestion, which forms the basis of the generation of a number of thematic motifs that recur throughout the performance. Performances of Harold longform are interesting objects of study for linguists primarily because (a) there are no artificial restrictions on the actors' speech; and (b) the reference to previously-established (i.e., contextual) information is a particular feature of Harold performances. The dataset that is analysed in this dissertation is made up of four performances of Harolds, which each lasts approximately thirty minutes. These performances all occurred in a single evening in a pub in Melbourne, Australia in late 2014.

A viable model of speech in improvised theatre requires a model of improvised theatre itself. In chapter 3, I propose the following model.

All performances of improvised theatre take place in a **plot-world**, which is a Stalnkarian possible world—a "way things might have been", or a fully specified "state of affairs" [Sta76]. A **plot** is simply a series of events that occur within a plot-world; it follows that these events are possible within the particular plot-world. Events are visible actions—those actions that are shown to the audience—, and do not include, for instance, the thoughts of the characters, or long-term events at a cosmic scale. In linguistic terms, this includes the performance of utterances, and not interpretations. The plot is partitioned into **scenes**, each of which is generally confined to a single location, and a finite set of characters. Each scene is further composed of a number of **states** of the scene [BM09; BM10]. States are identified with certain states of the world in which those states are true; that is, a state is identified by its extension. The information about the plot-world that is established in the course of a performance thus contributes directly to the identity of the state. Furthermore, certain states are rendered impossible by the introduction of information: this expresses a requirement that plots be self-consistent, and that plot-worlds have their own set of rules.

The most basic unit of analysis of an improvised performance is not the individual utterance, but is instead the **offer–response pair** (o–r pair). Any utterance produced in an improvised performance is either an offer, a response, or else both simultaneously. O–r pairs are categorised by how they affect the state of the scene. When an actor produces an offer, she may wish to signal either that the state be shifted, or else that the state be maintained. Since states are identified with extensions, state maintaining moves do not introduce or revise information that would alter the state's extension, while state shifting moves do. Responses are categorised according to two axes: whether they accept the offer's assertion—called the response's **polarity**, which can either be *affirmative* or *negative*—, and whether the response accepts the state operation proposed by the offer—called the response's **conformity**, which can either be *augmentative* or *redirective*. The following table summarises the types of response:

	accepts assertion	does not accept assertion
conforms	affirmative augmentation	negative augmentation
does not conform	affirmative redirection	negative redirection

I find that there are some restrictions on the kinds of responses that can follow certain sorts of offers: affirmative redirections do not occur after state shifting offers, and negative augmentations do not occur after state maintaining offers.

One of the primary difficulties in formulating a formal model of improvised theatre is that the character of agency is particular to the genre. Actors make decisions that are attributed to characters in the plot-world, and the verisimilitude of the plot—the preservation of which, recall, appears to be the actors' primary motivator—relies on actors behaving believably as characters. The difficulty lies in the fact that actors often know more than characters do. Characters do not, for instance, realise that they are being observed by an audience, and neither are they privy to conversations that occur when they are offstage, as actors are. For the most part, I ignore such complications, since the model pertains to language use in improvised theatre, rather than the metaphysics thereof. And so, I propose that a character is identified by a set of **traits**; the formal specification of a character is an attribute–value matrix. A character's traits are are ascribed to her through the *deliberate action of the actors*. That is, intrinsic traits of the actor playing the character—such as the actor's accent—are not automatically inherited by the character. The ultimate goal in behaving as a character is that the character be **consistent**. I define a character's consistency by whether there is an entity in the plot-world that felicitously matches that character for the entirety of a scene. Inconsistencies most often occur when a trait has been assigned, and then another trait is assigned that is inconsistent with the first. A simple but common example involves applying two different names to a single character.

3 Improvised performance games

3.1 Semantic model

To model utterance semantics, as well as interpretations and the common ground, I introduce an intensional semantic model that integrates with the possible worlds-based model of improvised theatre that is outlined in the previous section. The model that I devise is a subset of **Discourse Representation Theory** (DRT) [KGR11], which was selected because (a) versions thereof are formulated with an intensional model [see KGR11]; (b) **Discourse Representation Structures** (DRSs)—the basic unit analysis in DRT, analogous to sentences or formulae in other formal languages—can fulfil a range of semantic functions, by representing utterance semantics, interpretations, and the common ground; and (c) it provides a mechanism of update, so that it can represent background information that is established cumulatively.

The DRS language $\mathcal{L}$ that I use in this dissertation is similar to the model defined by Kamp, Genabith, and Reyle [KGR11], in that it includes attitude complexes that describe propositional attitudes; $\mathcal{L}$ includes modal indicators in its vocabulary. The syntax of $\mathcal{L}$ is that of a standard DRS language with attitude complexes. A DRS is defined as a pair, containing a set of discourse referents and a set of conditions over those referents. Valid conditions are specified according to those that are given by Kamp, Genabith, and Reyle [KGR11], and these are used to define **attribute description sets**, which are the primary means of formally representing the propositional attitudes held by agents. The cumulative update of information is represented in DRS by **merging** two DRSs K and K', which results in a DRS K'', whose set of referents and set of conditions are the unions of the equivalent sets of K and K'.

The intensional semantics of the DRS language model relies on a model constituted of three entities: a set of possible worlds $\mathcal{W}$, a non-empty domain of individuals (i.e., entities) $\mathcal{D}$, and an interpretation function $\mathcal{I}$ that, depending on the input, assigns names to individuals in $\mathcal{D}$, or assigns predicates to individual entities, such that this assignment is valid *only within a subset of the possible worlds in* $\mathcal{W}$. In DRT, DRSs are treated as partial models of a given possible world. That is, for each possible world, there is assumed to be a DRS that lists every entity in that possible world and the conditions that are applied to that entity; and any given DRSs is a subset of this larger DRS. Therefore, the truth of a DRS relative to a possible world is determined by whether the DRS can be **embedded** within the world's fully-specified DRS. From this definition of truth, a **proposition** is defined as a set of worlds in which a given DRS is true.

3.2 Game model

Improvised performances are modelled by **improvised performance games** (IPGs), which are, in essence, interpreted signalling games. Interpreted signalling games are defined by specifying five parameters: (a) A set of players N; (b) a **strategy space**, which is the Cartesian product of the players' **strategy sets**—the latter contains the actions that a player can perform, which in this case are either utterances or interpretations; (c) a set of worlds or world-states $\mathcal{W}$; (d) a semantic denotation function that associates utterances with worlds in which they are true; (e) a utility function that associates outcomes of the game with real numbers.

I make two major changes to signalling games. First, the number of players may exceed two, with the roles of Sender and Receiver being recast as *categories of player*. Each IPG has one Sender, but it may have potentially many Receivers. This is important in modelling the concealment of information from the audience that pervades communication in improvised theatre, which I discuss below. Second, I include in the Sender's strategy set an empty utterance ϵ. The inclusion of this strategy means that players can opt to remain silent, and it is introduced to capture situations in improvised theatre where offers are made but are not responded to directly. The absence of a response itself is an intentional act, and is, according to the data, generally interpreted as acceptance.

A performance is modelled as a *series* of IPGs, each play of which is called a **period**, and has the following form:

1. Nature selects a state of the scene, which is observed by that period's Sender.
2. The Sender produces an utterance.
3. Each Receiver selects an interpretation. The order of interpretations does not matter, since each player may only observe her interpretation, and thus the choice of an interpretation is not contingent on other players' choices.

Additionally, at the beginning of the performance, there is a period zero, in which Nature selects a plot-world—this restricts which states of the scene are possible. The Sender of period one views this selection. In game theory, private information of this sort may represent a state of the world, or a communicative intention, or the hidden price of certain commodities. Since actors appear to have preferences over the trajectory of the scene's plot, then, Nature's selection of a state of the scene is tantamount to stating that an actor selects an intention for the plot-trajectory that is known only to her. And so, the addition of a zeroth period models the fact that the first actor to speak in a scene generally determines certain grounding facts about that scene—for instance, where the scene is set, the relations of the characters to one another, the problem at the core of the scene, and so on. Nature's selection at each period represents a similar process of motive-setting, except that it allows for a continuous updating of actors' preferences throughout the course of a performance.

At the conclusion of each period, actors are informed of their **private histories**. The constitution of private histories differs for each player, since each player has access to different information: only the Sender can observe the state that Nature selects, and only the Receiver can observe which interpretation she has selected. However, players can make estimates of the other players' hidden actions, based on their beliefs. So, the Sender's private history for a given turn includes the state of the scene, the utterance, and the interpretation she considers most likely; and the Receiver's is the state of the scene she considers most probable, the utterance, and her interpretation.

The players' utilities are awarded at the *conclusion* of the IPG, which means that during the game itself they are unaware of their utilities. This models an actor's uncertainty of the success of a scene until that scene's conclusion. However, actors do have access to their private histories, and so at the end of each period, they use this to calculate their **provisional expected utility**, which accumulates throughout a scene. Estimating the provisional expected utility depends primarily on the player's private history, as well as her beliefs about the plot-world.

4 Linguistic strategy in improvised theatre

After I have devised a provisional form of the formal model, I proceed to analyse the linguistic strategies used by actors. These strategies are grouped together according to the function of the strategies—that is, the purported intention in selecting a particular intention at a particular turn.

4.1 Plot-world building and resumption

The first functions that I describe are those that are used to build, or specify, a plot-world; chapter 5 is devoted to this topic. I identify two situations that are salient to this function: the creation of a plot-world without any prior information, and setting a scene in a plot-world that has been established in a previous scene—called the **resumption** of that plot-world.

In the former case, I examine the use of presuppositions and attribute ascriptions through sociolinguistic variation as techniques to either propose new facts about the plot-world if the strategy is an offer, or accept an offer if the strategy is a response.

(a) Presupposition strategies operate by causing the audience and other actors to accommodate the presupposed facts. Presuppositions are used widely because they give a sense that facts have already been established in a plot-world *prior to the scene's commencement.* That is, if something is presupposed by an utterance, and it is accepted by the characters, then the audience also accepts it as another fact about the plot-world of which they are unaware.
(b) Attribute ascriptions through sociolinguistic variation use accent, word choice, and grammatical variation to convey attributes about their characters. This sort of strategy can also be used more broadly to assert facts about the plot-world. For instance, if an actor uses a particular accent when portraying her character, then that character is ascribed with certain, generally stereotypical characteristics of the group associated with that accent.

These sorts of strategy are used to provide a plot-world with the sense that it exists prior to the execution of the scene, and do so by positing new information that is accepted by the actors participating in the scene and the audience. Strategies that resume plot-worlds instead rely on information that has already been established in the course of the performance. Actors that resume a plot-world tend to unambiguously signal the previous scene using some piece of information that was established in that scene, and they do so immediately as the scene begins. But they also signal what has changed between the scenes. Other than what is explicitly stated to have changed, everything else is presumed to remain consistent between the two scenes. Actors commonly use this to subvert expectations in later scenes, generally for comedic effect.

Errors in consistency sometimes occur during the process of plot-world building and resumption, and in attempting to repair these errors, actors reveal their preferred actions. Analysing errors, then, can be useful in ascertaining what actors' preferences are. I note, for instance, that a kind of common error of attributing two different names to a single character occurs a small number of times in the data. Different strategies are used to remedy this error, which share a common goal of incorporating the error into the logic of the plot-world. This reinforces the analysis of actors' motives that they prefer the plot-world to remain verisimilar throughout the scene.

At the end of this discussion, I propose five heuristics:

(1) An action is called optimally verisimilar if it appears to believably belong to the plot-world; that is, if it does not appear to be produced by an actor pretending to be a character.
(2) A strategy is optimally compatible if the facts expressed by that strategy do not contradict what is established about the plot-world prior to that strategy's deployment.
(3) If a Sender intends that a scene be resumed, a strategy is optimally salient with respect to that scene (called the target scene) if there is no other scene that is more probably the target scene given the strategy.
(4) If a Sender intends that a scene be resumed, a strategy is optimally distinct if only the intended differences between the target scene's context and the present scene's context are communicated.
(5) A strategy is optimally plausible in the plot-world if, according to the information that is already known about the plot-world, the plot-world has a high probability of being an element of the extension of the strategy's DRS.

I provide formal definitions of each of these. Additionally, I identify the following assumption:

- If a scene is resumed, all facts are assumed to be identical to those in the target scene, unless this assumption is explicitly overturned.

4.2 Establishing a common ground

The process of establishing a plot-world amounts to the iterative construction of a structure of background knowledge that the actors use as a basis for the design and interpretation of utterances. This background information is cast as the **common ground** that is shared by the actors and the audience, after the definition given by Stalnaker [Sta02, p. 716]: "it is common ground that ϕ in a group if all members *accept* (for the purpose of the conversation) that ϕ, and all *believe* that all accept that ϕ, and all *believe* that all *believe* that all accept that ϕ, etc.". The common ground is thus used "as a resource for the communication of further information, and against which he [the speaker] will expect his speech acts to be understood" [Sta75, p. 273]. Since agents who share a common ground are not required to believe what is posited for it to be included in the common ground, it provides a better basis for the epistemological condition of agents participating in or being spectators of an improvised performance. I emphasise that the process of building a common ground is a thoroughly *negotiated* process. Negotiation, in the sense that it is used in this dissertation, relies on concepts developed by Arundale [Aru10]: **provisional** and **operative** meanings of utterances. According to Arundale, the process of arriving at an operative interpretation broadly follows a three-stage schema:

(a) The first speaker produces an utterance that *projects*[1] a number of interpretations, with a particular interpretation intended to be conveyed;
(b) the second speaker produces an utterance that projects the same interpretation that the first speaker intended, called the provisional interpretation; and
(c) the first speaker produces an utterance that confirms the provisional interpretation as the operative interpretation that will be used in the remainder of the conversation.

And so, an operative interpretation is one that has been assented to by all parties to a conversation. This assent does not need to be explicit, but may simply arise by not disputing the meaning of the utterance. Furthermore, it may take a number of turns to arrive at the operative interpretation of a given utterance. Crucially, at any turn the hearer of an utterance may select an interpretation that *may not be foreseen by the speaker as a projected interpretation of that utterance.*

The common ground is negotiated in this sense: interlocutors accept particular interpretations of information introduced during the conversation, and this information is used in the negotiation of other interpretations later in the conversation. The common ground arises only through the speech acts of a number of agents, and each of these agents may propose (a) that certain facts be included in the common ground; and (b) that proposed facts be accepted or rejected. I propose a formal definition of common ground simply as a DRS that has a nonempty extension; that is, the common ground should be *minimally plausible*—possible in at least one possible world $w \in \mathcal{W}$. The greatest effect of this definition is the redefinition of players' beliefs in the game model. The Sender's belief becomes a conditional probability distribution of interpretations given the common ground; the Receiver's prior belief over the set of possible worlds is conditionalised on the common ground—since the common ground must be true in at least one $w \in \mathcal{W}$, it narrows the possible range of possible worlds; and in the revised Receiver's belief, the probability of a given $w \in \mathcal{W}$ is conditionalised on the conditional probably of an utterance given the common ground. These three redefinitions of belief simply include the common ground as a conditional term, and yet the effects prove to be extremely robust and subtle.

Similar to chapter 5, I analyse strategies that actors use when negotiating during improvised performance. In particular, I focus on strategies that actors use (a) to negotiate *operative* interpretations of utterances; and (b) to signal their preferences over the direction of the performance's plot. In the former case, I examine the use of repetition as a strategy to provide a body of evidence for a particular interpretation of an utterance, particularly when there is a dispute in choosing between multiple viable candidates for the operative interpretation of that utterance. In the latter case, I examine the actors' use of prosody and repetition to emphasise their intended functions on the state of the scene. The findings of this analysis are summarised in heuristics proposed at the end of the chapter:

[1] In Arundale's usage, *projects* appears to mean *evidences*, such that an utterance provides evidence for particular interpretations.

(6) Given a particular utterance, and a series of responses such that one response is made by each of the parties to a conversation, an operative interpretation of the utterance is optimally marked if it is most probable given *all* of the responses.
(7) Given an intended interpretation and a series of utterances meant to convey this interpretation, a strategy of repetition achieves optimal convergence if and only if the intersection of the extensions of each of the utterances contains *only* the intended interpretation.

Furthermore, I propose two heuristics that govern how actors interpret response strategies:

(8) A response to an offer is *optimally polar* if it is unambiguously the case that what is intended to be conveyed by the offer utterance is taken up and subsequently accepted or rejected.
(9) A response strategy is by default positive.

5 Natural language steganography

The formulation of the model culminates in chapter 7 with the introduction of a model of natural language steganography. Steganography is the art and science of transmitting usually private information *in the presence of adversaries* by concealing it within an apparently innocent 'cover' medium [PH03]. Steganography is closely related to the better-known **cryptography**, but there is an important difference between the two. While the result of encrypting a file is a garbled sequence of seemingly random, unusable data, a so-called **stegosystem** outputs data that are structured, so that the very presence of hidden information may be plausibly denied. The **security** of a stegosystem is a metric that measures the degree to which an object with hidden information notably differs from an object with no hidden information.

5.1 Cachin's information theoretic model of steganography

Information theoretic studies of steganography serve as a good foundation for a formal model of communication in improvised theatre, since, as mentioned above, actors' speech in improvised theatre operates on two levels simultaneously: the level of the plot—at which utterances must appear verisimilar—and the level of planning coordinated action—which must remain hidden, if verisimilitude is to be maintained. The utterances that are optimal in improvised theatre are generally those that are most believable in the scene's plot-world, since they most convincingly appear free of hidden planning information.

The NL stegosystem that I propose is based on a subset of the information theoretic model formulated by Cachin [Cac04]. In the subset of Cachin's model that I use, there are three agents: a Sender, a Receiver, and an Adversary. The basic sequence of an act of communication is tripartite: the Sender emits a signal through a public channel; the signal is observed by both the Receiver and the Adversary; and the Receiver extracts evidence of the hidden meaning. There are two states that the Sender is possibly in. In the case that she is *inactive* (state 0), she sends an innocent signal that doesn't contain any concealed information—this utterance is called a **coverform** c. If she is *active* (state 1), she sends a **stegoform** s. Observe that both of these may be the same utterance, but they differ in the degree of complexity of their respective formation. The coverform is simply a signal drawn from a set of possible signals, while the stegoform is generated by an **embedding function**. The embedding function has access to a **random source**—which is known only to the Sender—; a **key**—which is selected randomly and shared across a private channel by the Sender and the Receiver, so that it is known mutually by these agents—; and the coverform as an argument, since this is the signal that sent across the public channel. The key represents shared knowledge that is known only to the Sender and the Receiver. This aspect in particular becomes important for the model of NL. Note that the **hidden message** $z \in Z$ itself is not included as an argument of the embedding function in Cachin's model. Following more general information theoretic models of communication [see CT12], Cachin claims the stegoform ought to provide the Receiver with evidence of the hidden message, rather than delivering the hidden message itself. Therefore, Receiver performs an extraction function that takes s and k, and returns an *estimate* of z, $\hat{z}$.

In Cachin's model, the Receiver always has perfect knowledge of the Sender's state, and so her primary task is to determine the value of z given $\hat{z}$, by solving $\arg\max_{z \in Z} \Pr(z|\hat{z})$. The Adversary's decision problem involves determining whether the signal contains some hidden information, and therefore she must apprehend whether the signal was selected from a coverform probability distribution $\delta(C)$ or a stegoform probability distribution $\delta(S)$. A stegosystem's efficacy—its **security**—is defined by the **relative entropy** of $\delta(C)$ and $\delta(S)$, which is a metric of the relative uniformity of the probability distributions. If a particular utterance is highly probably in, say, $\delta(C)$, then it is more obvious that it is has been selected from this distribution. If the distributions are less dissimilar, then it becomes difficult to make inferences of this kind.

5.2 A game model of NL steganography

In Cachin's model, the Sender and Receiver share a key across a private channel, which is used both in the embedding and extraction functions. My model of natural langauge steganography likewise relies on information that interlocutors share, and particularly disjunctions in this information. To provide a formal model of this, I propose to split the common ground into a number of **partial grounds**, so that the former is the union of the latter. For any given subset of the set of interlocutors with a size greater than two, there is a partial ground that describes what the agents in that set accept for the purpose of the conversation. The formal implementation of this relies on an addition of a system of **permissions** that apply to DRSs, which specify which agents are able to 'read' that DRS. I define a **permission implication function** that relates DRSs' permission sets to others.

The order of events in my model is as follows:

1. A state is set by Nature, which determines whether the utterance produced by the Sender contains hidden information. This specifies how the Sender treats the partial grounds between her and the Receiver, and her and the Adversary. In state 0, she uses the *intersection* of the partial grounds, meaning that she designs the utterance so that it is able to be interpreted identically by all agents, since their interpretations are based on the same information. In state 1, the Sender uses instead the *symmetric difference* between the partial grounds, so that her utterance can be understood differently by each of her interlocutors. So, in states 1 and 0, a stegoform and a coverform is sent across the channel, respectively.
2. The Adversary and Receiver both observe the utterance, and interpret it. Both of the agents select an interpretation, represented as an estimate of the state set by Nature.

The formal implementation of **coverform** and **stegoform utterances** is very simply, and relies on a definition of **projected interpretations**, such that for each of the Receiver and Adversary, there is a DRS that represents the Speaker's preferred interpretation. For a coverform, these projected interpretations are equivalent—since there is no information that is hidden from the Adversary—and for a stegoform, they are not.

There are three major ways that this model diverges from Cachin's:

(a) In Cachin's model, the selection of a coverform precedes the determination of the Sender's state. This is possible because the coverform can exist prior to the message being sent—for instance, if the coverform is an image used to hide the information. This is not the case in my model, since the selection of a stego- or coverform occurs simultaneously with its production as an NL utterance.

(b) The Sender and Receiver in Cachin's model have a single key that they share unilaterally. I instead define the **keys** as *any DRS condition that is present in the Sender and Receiver's partial ground, but not in the Sender and Adversary's.* This means that there are likely multiple keys, and it is based on these keys that any information might be hidden.

(c) Security in my model is a measure of the relative uniformity of the cover- and stegoform distributions, but these distributions are conditionalised on the set of possible worlds. The NL stegosystem's security is therefore more complex than Cachin's general model, for successful concealment of NL information requires consideration of the state of the world.

To incorporate the NL stegosystem into the game model, the DRT model is incorporated into the formal model of the stegosystem, so that the stegosystem is defined as a DRS model with cover- and stegoform distributions; and the game model—which is recast as a **stegogame**. Stegogames thus follow the normal order of an interpretation game, with each player except the Sender selecting an interpretation, but the formal linguistic model is subsumed in the model of the NL stegosystem.

5.3 Strategies to conceal information

I examine the use of three kinds of utterance, and the information that they conceal: declarative utterances; questions; and implicates. These three kinds of utterance have a finite amount of information that can be concealed: they may **propose** that certain information be accepted into the common ground of the performance, and subsequently **accept** or **reject** this information. So, for instance, the utterance "Well, they fired me" conceals the communication "*I propose that* they fired me". These three functions are the only possible messages that may be concealed in improvised theatre, and while the different sorts of utterance that I examine approach their concealment by different means, they do not extend what may be concealed.

As in previous chapters, I propose heuristics in this chapter that pertain to the concealment of information in improvised theatre.

(10) For every subset of the set of actors involved in an improvised performance, there is always a partial ground that they share, such that the partial ground contains the information that (a) the other agent is an actor; (b) the activity they are performing involves constructing a plot; (c) the structure of communication in the genre is that of iterated offer–response pairs; and (d) there is an audience present.

This heuristic provides actors with a basis for *every* utterance to conceal information, without the need for any common ground information to be established about the plot-world. In other words, actors always expect every utterance in improvised theatre to conceal information, and they mutually share that expectation.

(11) An embedded meaning is optimally functional if the probability of one of the predicates proposes, accepts, or rejects being the embedded meanings, conditional on the actors' shared partial ground, is greater than the respective probabilities of the other two.

At the end of the chapter, I propose a definition of utility in improvised theatre. Based on the data, it appears that the audience isn't interested in comprehending the concealed meanings, and that they in fact prefer not to understand them, for this compromises the enjoyment of the performance. The utility function I define yields the following ordering of outcomes, where the order of utilities is (S, R, A), and where A stands for the audience:

	A understands	A misunderstands
R understands	$(0, 0, -1)$	$(1, 1, 1)$
R misunderstands	$(-1, -1, -1)$	$(-1, -1, 1)$

To summarise, the actors are interested in being understood by their fellows, and by concealing information from the audience, and the audience is interested in not observing the concealed information.

6 Conclusion

I devote chapter 8 to a thorough application of the model to a scene from the dataset. I provide a list of the states of the scene, and the strategies that are used to shift or maintain them, before analysing the first four o–r pairs of the scene.

This analysis serves to demonstrate the efficacy and completeness of the model, insofar as the diverse utterance strategies that the actors use in the scene are both described with respect to their concealed meanings, and accounted for using the game model and its definition of utility. It illustrates how the three claims I make at the beginning of the thesis are justified. Claim **A** is substantiated by demonstrating that actors form narrative common ground through an iterative process of proposal, acceptance, and rejection of information, especially in the case of concealed communication. The resulting picture is one of conscious, concerted effort to establish a narrative common ground through a process of negotiation. Claim **B** is evidenced similarly, through the introduction of partial grounds, and their necessity for a description of covert communication, which appears to be a basic feature of language use in improvised theatre. Without partial grounds, there is no way of accounting for covert communication, for there is no means of expressing the differences between the information that is available to different agents. Claim **C** is demonstrated by the data that suggest the operation of a natural language stegosystem that exploits differences in interlocutors' partial grounds. I show, too, that these differences are used by actors to advance their preferred plot-trajectories, and that they ensure that they exist between the information that is available to the audience, and that availalbe to the actors.

The presence and operation of natural language stegosystems is clear in improvised theatre. However, I propose that there are everyday situations in which interlocutors wish to convey different information to different interlocutors. I claim, then, that the study of natural language steganography in other genres of naturally-occurring speech is likely to be a fruitful endeavour, and that my model of NL steganography—while it is limited to a given genre—provides the mathematical basis for such a study.

References

[Aru10] Robert Arundale. "Constituting face in conversation: Face, facework, and interactional achievement". In: *Journal of Pragmatics* 42.8 (2010), pp. 2078–2105.

[BM09] Allan Baumer and Brian Magerko. "Narrative Development in Improvisational Theatre". In: *Interactive Storytelling*. Ed. by Ido A Iurgel, Nelson Zagalo, and Paolo Petta. Berlin: Springer, 2009, pp. 140–151.

[BM10] Allan Baumer and Brian Magerko. "Proceedings of the Third Joint Conference on Interactive Storytelling". In: ed. by Ruth Aylett et al. Vol. 6432. Lecture Notes in Computer Science. Berlin: Springer, 2010, pp. 165–175.

[Cac04] Christian Cachin. "An information-theoretic model for steganography". In: *Information and Computation* 192.1 (2004), pp. 41–56.

[CT12] Thomas M Cover and Joy A Thomas. *Elements of information theory*. New York: John Wiley and Sons, 2012.

[Fra11] Michael Franke. "Quantity implicatures, exhaustive interpretation, and rational conversation". In: *Semantics and Pragmatics* 4.1 (2011), pp. 1–82.

[HCJ94] Charna Halpern, Del Close, and Kim Johnson. *Truth in Comedy*. 1st ed. Collorado Springs: Meriweather Publishing, 1994.

[KGR11] Hans Kamp, Josef van Genabith, and Uwe Reyle. "Discourse Representation Theory". In: *Handbook of Philosophical Logic*. Ed. by Dov M. Gabbay and Franz Guenthner. Vol. 15. Dordrecht: Springer Netherlands, 2011, pp. 125–394.

[PH03] Niels Provos and Peter Honeyman. "Hide and seek: An introduction to steganography". In: *Security and Privacy, IEEE* 1.3 (2003), pp. 32–44.

[Sta02] Robert C Stalnaker. "Common ground". In: *Linguistics and Philosophy* 25.5 (2002), pp. 701–721.

[Sta75] Robert C Stalnaker. "Indicative conditionals". In: *Philosophia* 5.3 (1975), pp. 269–286.

[Sta76] Robert C Stalnaker. "Propositions". In: *Issues in the Philosophy of Language*. Ed. by David D Merrill and Alfred F MacKay. New Haven: Yale University Press, 1976, pp. 79–91.