Contents

Chapter 7

Proposals

> *"Would you marry me?" he said. "Well of course I would, but not if you ask like that." "What do you mean?" he replied. "You have to ask for real," I told him.*
> –An unnamed linguist referencing her boyfriend's awkward marriage proposal. Penn Linguistics Colloquium

Why do some calls to action elicit different responses than others? And why do they take different forms? Whether we want someone to open the door for us or go with us to a movie, we are making a a *face-threatening act*(FTA), as we are asking them to change their current course of action, thus lowering their negative face. For instance, we can consider the following pairs of utterances:

Would you open that door?
Would you go to the Justice League movie with me?

Both ask the hearer to accomodate the speaker's desires and thus can be interpreted as requests. Yet we might see a different tack when considering the second item. Consider now this pair:

Do you want to open that door?
Do you want to go to the Justice League movie with me?

In this set of utterances, the first seems odd without context, and in certain instances, it might be cast as a negative face threat. The second, however, might seem like a more natural way to ask a friend about seeing a movie. We claim that this second way puts the utterance in *proposal* territory. The key here is that there are very few instances *prima facie* where the hearer would *want* to perform the action of opening the door, without further incentive. In contrast, the hearer might plausibly *enjoy* seeing the movie. Unraveling the reasons for why requests differ from proposals in grammatical form relies on understanding why the incentives underlying them are different.

Politeness is not a dominant strategy however. I.e. there are situations where direct speech avails, even though the speaker is attempting to incite action from the hearer. In such cases, politeness, especially involving strategies with negative politeness, is not well-suited.

In contrast with requests, proposals encode an interaction potentially to the benefit of both participants. Returning to marriage, we noted the use of modals in certain contexts differs. For purposes of both humor and invoking the undercurrent of common knowledge, we observed that *would* allows for a certain amount of disavowal whereas *will* does not. Consider examples like the marriage proposal in the opening and the possible replies. For example, the following dialogues can be completed for comedic effect:

Xavier: Would you marry me?
Yvonne: *I would...if you were rich.*
Xavier: **Sigh**
(or)
Yvonne: *Yes!!! I thought you'd never ask!*
Xavier: *Woah, I was just asking hypothetically!*

In these examples, we see that there is no guaranteee of commitment to joint action between the two agents. Yvonne's initial agreement still allows for counterfactual statements, references to possible situations in which she would marry Xavier, or avoiding obligation. On Xavier's part, even if Yvonne agrees, he can plausibly deny that he was interested in the act of marriage [**?**]. Even smaller stakes like going to a movie can elicit confusion when proposing a joint action.

Xavier: Would you like to see a movie?
Yvonne: *Yeah, there are a few I'd like to see.*
Xavier: *Great! When can I pick you up?*
Yvonne: *Oh! I didn't realize you meant with you.*
(or)
Yvonne: *Yeah! When do you want to go?*
Xavier: *Oh! I didn't mean with me, just in general.*

Here is where we differ in some ways from **?**, as we find a reference to what they call a positively polite request:

> Hey Joe, let's go take in a flick tonight, OK? - I mean if you're not too busy

Without the hedge of *I mean if you're not too busy*, we would classify this as a proposal. The underlying assumption here is that seeing a movie would be enjoyable for both parties. We can also imagine that, given variations on preferences, we could have a multitude of mutually satisfying outcomes for the two parties.

For the conditions of a request to be satisfied, we claim that it must be the case that the one being asked, in the above case *Joe*, would somehow be better off not going to the movies. This could very well be the case, for instance if Joe had a date with his girlfriend or an important meeting the next day.

We can even see this kind of preference elicitation [**????**] in advertising, economic policy proposals, law enforcement, service situations, and even the humorous case from *Lady Marmalade*:

How would you like a nice Hawaiian Punch?
Would you like to shrink the welfare state?
Voulez-vous coucher avec moi?

The latter even has an element of eliciting preferences and using negative politeness with the *Vous*-form of address. In this case, the proposal is to sleep with a madame, with the madame soliciting a paying customer.

In these cases, the utterance asks for the preference of the agents, although it does not ask about a course of action. Other cases of *will* and *would*, however, seem to ask about a commitment to future action. We would like to uncover the reasons for this.

Modeling the Conventionalization of English Marriage Proposals

Although the utterance *Will you marry me?* may seem natural enough to native English speakers, a closer examination into the surrounding context reveals that the operative modal verb, in this case *will*, may be a fossilized form from an earlier time that does not indicate a commitment to a future action but rather a desire to take that action. To connect this semantic shift to the other chapters of the book, I would like to make two claims:

1. *Will you marry me?* evolved from an earlier expression connoting *Do you want to marry me?*

2. Game-theoretic models of preference revelation, commitment, and coordination model the dynamics behind the semantics

Evaluating these claims on a semantic level means going to evidence from a sample of Indo-European languages, the modern context of wedding vows, and corpus data from older forms of English. Analyzing the second, game-theoretic claim means we need to understand games like the *Stag Hunt* and how they connect to signaling.

Illocutionary Force, Intention, and Incentives

We argue that *will* and *would*, for the most part, have the same illocutionary force. They both ask our hearer to potentially commit to an action. However, they differ in that *would* allows for disavowal. In a request, this makes sense, as we cannot obligate someone to do something potentially against their self-interest. What if we do know something about their preferences? To tease out how they do differ, we consider the notion of self-enforcing equilibria. The point here is that even though we may know something about the preferences of another agent, that may not be enough to guarantee our knowledge of their actions, particularly when more than one equilibrium exists.

7.1 Separating Requests from Proposals

Requests and proposals differ in that requests serve primarily to benefit the speaker. The previous chapters have shown that combinations of repetition, sympathy, and face-respect (e.g. avoiding FTAs) can affect the success of a request, as it may difficult to compel someone to do something that is not in their immediate interest. In effect, what certain types of politeness do is transform a win-lose scenario to a win-win.

In contrast, the end result of proposals ideally benefit both parties. To know whether a proposal will truly benefit both parties, a speaker may ask about the preferences, abilities, or intentions of the hearer. This also reveals why certain formalities seem unnecessary when asking someone to commit to something in their own interest. This scale is potentially continuous and accounts for the diversity of proposal types.

? gives a first step at the game-theoretic background behind how modal verbs feature in questions like:

- Will/Would you lend me a dollar?
- Will/Would you open the door?
- Will/Would you turn that music down?
- Will/Would you marry me?

Their claim is that the relationship between the speaker and hearer can dictate how likely one is to use one utterance over the other, as cases like asking *Will you lend me a dollar?* or *Will you marry me?* to a stranger might seem odd. In contrast, *Would you marry me?*, when asked to one's long-time girlfriend, seems to be infelicitous as well. A parent might be able to ask *Will you turn that music down?* whereas a sibling may not. **?** goes on to elaborate possible replies to *Would you marry me?* like *I would if you were rich or handsome!*

We want to explore the following regarding requests and proposals:

- Morphosyntactic Form: What are the similarities and differences in how we make requests and proposals? Why do certain modal verbs lend themselves to one and not the other?
- Semantic/ Pragmatic Content: Why do requests and proposals seem to have similar forms but differ in the types of speech acts they perform?
- Behavioral Incentives: How do game-theoretic models of cooperation underlie repeated requests and proposals?

In what follows, we examine the use of modals in requests and proposals. We show how modals, and other politeness strategies, when thought of in terms of other-regarding preferences, allow for expanded interaction between individuals. We also show how and why the use of the modal *will* in requests is binding, whereas *would* is not necessarily so.

Modality in Requests

Modal verbs like *could, would, may* etc. often feature in requests to accommodate negative face needs, as they reflect situations of possibility. The simple reason is that there is uncertainty when asking someone to do something for us when they may not benefit from it. For example, requests taken from the *Stanford Politeness Corpus* include:

- *Yeah, it's just search, sadly. Any chance I could get an accept for that? :)*
- *Could you suggest me a different way to reach the result? Another rom, or a market app?*
- *I would be interested in whether there is any impact to the life of the battery based on charging using one method over the other. Any suggestions or links?*

This is certainly not limited to English, for in German and French we have requests like:

- *Könnten Sie mir helfen? (Could you help me?)*
- *Ich möchte eine Cola, bitte. (I would like a Cola, please.)*

and

- *Puis-je m'assoie ici? (May I seat myself here?)*
- *Savez-vous conduire a la gare? (Can you drive to the station?)*

Notice that these requests need not be questions, as *Ich möchte eine Cola* or the third sentence above (*I would be interested . . .*) demonstrate. Even indirect requests like, *If you could pass the guacamole, that would be awesome!* [**?**] can appear without the syntax of a question.

Based on **?**

, there is a rich, varied, and active literature on modality. We will limit our attention to English for this brief synopsis. We can articulate different categories of modality and their exemplars within requests.

Epistemic modality expresses a speaker's knowledge or belief about the topic at hand, as *Would you have time tomorrow?* expresses the speaker's uncertainty regarding the schedule of the hearer.

Root modality covers subcategories of *deontic* and **dynamic** modality.

- **Deontic** modality expresses an agent's obligations, as in *Do you have to play the trumpet so loud?*
- **Dynamic** modality expresses an agent's ability to do something, as in *Could you get the phone for me?*

Although there are other categorizations of modality, we will settle on these.

Modality in Proposals

Modal verbs also appear in proposals, albeit in a different context. This might include cases like *Shall we dance?* or *How would you like to catch a ballgame at the bar?* Notice that the first invokes an obligation and the second asks for the preferences of the hearer.

We can see examples of *epistemic* modal verbs used in proposals like:

- *This could be the start of something big.(Steve Allen)*
- *I may/might/should be free tomorrow if you want to meet then.*

We can also see deontic modals in:

- *Shouldn't we get the car started sooner?*
- *Wouldn't it be nice if we were older . . . and we could be married! (Beach Boys)*
- *We have got to (must) find a better way to do this!*
- *How would you like to stop fighting and start working together?*

Requests and Proposals as Illocutionary Acts

Drawing from **?**, [1] , with editions for clarity. we have the following classification of illocutionary speech acts:

> **Directives** are *attempts (of varying degrees) by the speaker to get the hearer to do something*, e.g. requests, commands and advice.
>
> **Commissives** commit a speaker to *some future course of action*, e.g. verbal contracts, promises, and oaths.
>
> **Expressives** express *the psychological state specified in the sincerity condition about a state of affairs*, e.g. congratulations, excuses and thanks.
>
> **Declarations** *bring about the correspondence between the propositional content and reality*, e.g. christenings, legal verdicts, or pronouncing someone husband and wife.
>
> **Assertives** (or **Representatives**) *commit the speaker (in varying degrees) to something's being the case, to the truth of the expressed proposition.*

As we will emphasize, requests and proposals differ in that they signal to whose benefit the action will be. Requests typically signal an action only benefiting the speaker, and proposals typically benefit both parties. Thus the type of speech act used reveals the underlying incentives. [2]

In terms of **?**, a marriage proposal seems like a *commissive* speech act: it asks the hearer to commit to a course of action. Historically, it really asks about

[1] Italicized text is direct from **?**

[2] Marketing vs. Evangelism = Only benefiting hearer but with sympathy

attitudes and preferences. [3] Hence it can be thought of as *expressive* speech act. This is also the case with the German and French analogues:

Willst du mich heiraten?
Veux-tu m' épouser?

Both utterances encode *wanting* to marry the speaker. They do not, however, obligate the hearer to perform the action of marrying. Thus we have a curiosity in the English form of *Will you marry me?* as an utterance appearing to obligate future action. Further, we have the case that although both parties may be sure of the others' range of preferences, there is no guarantee of coordination. [4]

As we will soon see, mutual knowledge of preferences is not enough to establish coordination in many circumstances, and to explore this we will discuss findings in **?**, **??**, and **?**. To understand the incentives behind proposals as they contrast to requests, we will return to the theme of *Cooperation vs. Coordination* over multiple equilibria. This will include the Stag Hunt and Prisoner's Dilemma as contrasting examples of cooperation and coordination. We will also examine what internally motivated mechanisms alter agent decision procedures and how these mechanisms have linguistic analogues. In tandem, we will explore mechanisms altering game play like repetition and evolution, giving us a way to not only view an utterance within the course of a relationship between speakers but also to view an utterance as a strategy itself that can be replicated over time.

7.2 Stag Hunts and Social Contracts

Starting a new business arrangement or proposing marriage is risky, especially when we have another fallback option. Activities like this require someone else to sustain the potential benefits, and thus they require our opponent to trust that we will commit to a course of action that is mutually beneficial. As we have seen before, games like the one-shot Prisoner's Dilemma give us no credibility when signaling our action[**?**] , but games like the Stag Hunt can give us the emergence of social structure via agreements to adopt more profitable actions [**?**]. In previous chapters we saw how cooperative situations like trust games can evolve into coordination problems of following a norm. This occured via strategies involving reciprocity and punishment for deviating from the norm [**?**]. Despite the higher payoff of following a norm, in a one-shot game with multiple equilibria, an agent unsure of his partner's actions might require further reason to believe that his partner will commit to a risky equilibrium strategy[**?**].

The Stag Hunt is a model of social contracts, dating back to its coinage by Rousseau [**?**]. The reasoning was that the Stag Hunt possesses two equilibria: one low-risk with low reward; the other high-risk for the individual with high reward for the group. The terms of the contract bring about convergence to the socially optimal state (S, S). They also recognize the stability of the risk-dominant state

[3] Citation? History English Modals

[4] The anti-proposal of *You Won't Fall For Me If You Know What's Good For You* begs more analysis.

(H, H). For our purposes, we may consider examples like arranging a business deal, making plans to attend a movie, or getting married as social contracts. These outcomes (implicitly) benefit both parties more than the original state of nature, but all require a degree of extra effort to reach.

A	S	R
S	4,4	0,3
R	3,0	2,2

B	S	R
S	5,5	0,4
R	4,0	2,2

C	S	R
S	9,9	0,8
R	8,0	7,7

Table 7.1: A trio of variants for the Stag Hunt as discussed in **?**. The table in (C) is discussed in Aumann's work on self-enforcing of equilibria. All three have a socially optimal equilibrium of (S,S) and a risk-dominant equilibrium of (H,H).

We go on to assess utterances encoding situations of mutual benefit with multiple stable outcomes. We then look deeper into games like the Stag Hunt that provide a model of these situations, seeing how mutual knowledge of preferences can give us reasons to trust our partners.

Proposals and Stag Hunts

Given a relationship, proposals can have the effect of initiating a new one, transforming a previously stable arrangement into one potentially more rewarding. Examples we are interested in include business contracts and grant applications, for instance when we see companies merge or a research initiative approved. We also turn to individual arrangements, whether friends are seeing a movie or a couple becomes engaged to be married. A primary question is exactly what is signaled when a proposal is made. Is it a commitment to action or a revelation of preferences? We claim that the second can, in some cases, bring about the first.

I	C	D
C	a,a	0,4
D	4,0	2,2

II	C	D
C	5,5	0,4
D	4,0	2,2

III	C	D
C	3,3	0,4
D	4,0	2,2

Table 7.2: A trio of public goods games Stag Hunt and Prisoner's Dilemma depending on a parameter a. We have a Stag Hunt if $a > 4$(II), and we have a Prisoner's Dilemma if $a < 4$(III). A proposal with common knowledge of preferences revealed can reveal which game the agents are playing.

In Table 7.2, we have a base game in (I), but if agents reveal their preferences of the parameter a, the game can turn into one like the Stag Hunt in (II) or the Prisoner's Dilemma in (III). Here agents could discuss whether they prefer the (C,C) outcome to the other outcomes if they could engage in pre-play communication or have a strong commitment from the other partner to adopt that strategy. Otherwise, deviating from the (D, D) outcome could result in an expected loss.

Skyrms and the Social Contract

Two works by Skyrms trace the evolution of social contracts, using the Stag Hunt as the primary model for the emergence of social structure[**??**]. In these volumes, he describes a few ways in which a socially optimal outcome can both arise and stabilize. Although these volumes both discuss bargaining as a variant of arriving at social contracts, we omit those results here. In summarizing the issues with the social contract,[**?**] Skyrms remarks:

> *For a social contract theory to make sense, the state of nature must be an equilibrium. Otherwise, there would not be the problem of transcending it. And the state where the social contract has been adopted must also be an equilibrium. Otherwise, the social contract would not be viable. . . . The problem of instituting, or improving, the social contract can be thought of as the problem of moving from riskless hunt hare equilibrium to the risky but rewarding stag hunt equilibrium.*

We would like to examine variants for *solving* the Stag Hunt through both group dynamics and individual preferences. By *solving* we mean finding mechanisms that can make selecting the socially optimal outcome easier. This may include increasing its basin of attraction or giving agents other reasons to choose the *Stag* action. This may also include conditions which stabilize the *Stag* outcome under perturbations of the game.

7.3 Solving the Stag Hunt in a Pair

One set of options for *solving* the Stag Hunt involves pairwise mechanisms affecting how agents choose actions when in a two-player game. These include thresholds for certain outcomes, sympathetic preferences, and discounting future rewards. We begin with the general conditions for risk-dominance in the Stag Hunt and move to sympathetic types and relational thresholds in one-shot games. We then discuss repeated Stag Hunt games and the effects of sympathy on the discount values.

Generalized Payoffs and Risk-Dominance

In general, we can return to our original conditions on the Stag Hunt, where under the inequality $P > R > S > Q$, as seen in Table 7.3, we have a game with two strict equilibria and one of them having a higher payoff.

To compute when *Stag* will have a larger basin of attraction, we can find when $(P-R)^2 > (S-Q)^2$. We will assume $Q = 0$ and $P > R$, and thus

$$(P-R)^2 > (S-Q)^2 \Rightarrow P - R > S \Rightarrow P > R + S$$

In our case, a small value for S or a great difference in P and R will suffice to make *Stag* the risk-dominant equilibrium, although *Hare* remains a Nash Equilibrium as well. In the metaphor from Rousseau, we can think of half of a stag P

	S	R
S	5,5	0,4
R	4,0	2,2

	C	D
C	P,P	Q,R
D	R,Q	S,S

Table 7.3: Stag Hunt with specific and general payoffs. The table with general payoffs uses C for *Stag* and D for *Hare* to disambiguate from the general values. Assume $Q = 0$ for simplicity.

weighing more than three rabbits ($R + S$). If we look back to Table 7.3, we see that this is not the case, as $5 < 4 + 2$. Contrast this to Table 7.4 and we see *Stag* as a less risky strategy, where $5 > 2 + 1$. This could also be thought of as a share from a bargain or contract outweighing what the two agents might jointly receive if only one of them aimed for the bargain and the other chose the status quo.

In terms of the basin of attraction in general, we will soon see that the generalized symmetric form Stag Hunt has a stable point in the replicator dynamics when the frequency of *Stag* is $x = \frac{S-Q}{S-Q+P-R}$. This means that *Stag* becomes a more likely endpoint when strategies are replicated when S and Q are close or when P is much greater than R. This amounts to lower rewards for mutual pursuit of *Hare* or a larger reward for choosing *Stag*. Going back to $Q = 0$, we have $x = \frac{S}{S+P-R}$. As an example, these two tables below have larger basins for the *Stag* action, as mentioned earlier.

	S	R
S	7,7	0,4
R	4,0	2,2

	S	R
S	5,5	0,2
R	2,0	1,1

Table 7.4: Modified Stag Hunt with larger basin of attraction for *Stag* hunters. Notice it is more costly to deviate from the *Stag* action.

Social Preferences: Sympathy Solves the Stag Hunt

The Stag Hunt can also be solved by other-regarding preferences. In particular, sympathy can remove the risk-dominance of the *Hare* equilibrium. This does not remove choosing *Hare* as an equilibrium altogether, but it does alter the effect that uncertainty can have on the expected payoffs. In contrast, spite will not solve the Stag Hunt as it can remove the *Stag* equilibrium and deliver the Prisoner's Dilemma instead.

Using the payoffs from the Stag Hunt, let us now compute the conditions under which (S,S) becomes risk-dominant. We first begin with a payoff matrix with utility U altered by a sympathetic parameter s and given by $V_X(s) = (1 - s)U_X + sU_Y$:

$$U_X = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 5 & 0 \\ 4 & 2 \end{bmatrix} \end{matrix} \Rightarrow V_X = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 5 & 4s \\ 4-4s & 2 \end{bmatrix} \end{matrix}$$

	S	R
S	5,5	0,4
R	4,0	2,2

	S	R
S	5,5	4s,4-4s
R	4-4s,4s	2,2

Table 7.5: Stag Hunt with standard and sympathetic payoffs.

Let us consider the Stag Hunt in Table 7.5 with standard and sympathetic payoffs and the subsequent conditions for risk-dominance.

$$(5-(4-4s))^2 > (2-4s)^2 \Rightarrow s > \frac{1}{8}$$

Thus under a sympathy condition of $\frac{1}{8} < s < \frac{1}{2}$, we can transform (S, S) into a risk-dominant equilibrium while also maintaining the case that HH is still an equilibrium on its own, as $2 > 4s$ for $s < \frac{1}{2}$. On a more general level, we have

Solving the Stag Hunt with General Values

$$U_X = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} P & Q \\ R & S \end{bmatrix} \end{matrix} \Rightarrow M_X(s) = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & \quad D \end{matrix} \\ \begin{bmatrix} P & Q(1-s)+Rs \\ R(1-s)+Qs & S \end{bmatrix} \end{matrix}$$

Recall that in this case we know that $P > R > S > Q$. In particular, note that this should give us that $S > Q(1-s) + Rs$ and $P > R(1-s) + Qs$. We use this to compute the conditions under which we have a risk-dominant and payoff-dominant equilibrium.

$$(P-(R(1-s)+Qs))^2 > (S-(Q(1-s)+Rs))^2 \Rightarrow s > \frac{R+S-(P+Q)}{2(R-Q)}$$

As before, we have that $\frac{R+S-(P+Q)}{2(R-Q)} < s < \frac{1}{2}$ will give us *Stag* as a risk-dominant and payoff dominant equilibrium. Returning to our assumption that $Q = 0$, we have

$$(P-R(1-s))^2 > (S-Rs)^2 \Rightarrow s > \frac{R+S-P}{2R}$$

Defaults and Relational Thresholds: Amicability

Owing to **???**, we can first consider that default behaviors will solve the Stag Hunt without appeals to unbounded rationality. Among those with a default of rejecting outcomes that perform poorly for the groups, we can achieve this. As before, we invoke the notion of *k-amicable* outcomes, states of the game that outperform a threshold. Consider the Stag Hunt and the Prisoner's Dilemma. If the players have a default to

1. Consider equilibria vs. actions AND

2. Select the equlibrium above a certain threshold

then we can arrive at the socially optimal state in the Stag Hunt. In contrast, (C, C) is not a Nash Equilibrium of the Prisoner's Dilemma and as such will not be eligible for selection. In the case of the Stag Hunt in Table 7.6 below, we would need $k > 4$.

PD	C	D
C	4,4	0,5
D	5,0	$\boxed{2,2}$

SH	C	D
C	$\boxed{5,5}$	0,4
D	4,0	$\boxed{2,2}$

Table 7.6: Although the Prisoner's Dilemma and Stag Hunt contain outcomes better for the group, only the Stag Hunt has (C, C) as an equilibrium.

Looking back at Table 7.6, we can also see that while the Prisoner's Dilemma does contain two asymmetric outcomes that outperform $k > 4$, neither are equilibria. Note also that sympathy only affects asymmetric outcomes, and thus sympathy would not change the decision process here, as our Pareto-optimal outcome is also symmetric.

Repetition with Discounting and Sympathy

We can also go back to the repeated Stag Hunt and examine the feasible regions under various conditions to see how sympathy interacts with the discount factor or the fairness of the outcomes. When considering the figure in Figure 7.1, we see that what sympathy in fact does is

- increase the perception of fairness in the off-diagonal outcomes and

- reduce the attractiveness of exploiting one's partner in the off-diagonal outcomes.

Although this could also mean that some outcomes where one player exploits the other are more stable equilibria in the repeated game than they would be otherwise, these cannot be sustained without strategies more complicated than trigger strategies.

$$U_X = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 5 & 0 \\ 4 & 2 \end{bmatrix} \end{matrix} \Rightarrow V_X = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 5 & 2\delta \\ 4-2\delta & 2 \end{bmatrix} \end{matrix}$$

Stag is still the payoff-dominant outcome, but we are interested in where *Stag* becomes risk-dominant. This occurs when

$$5-(4-2\delta) > 2-2\delta \Rightarrow \delta > \frac{1}{4}$$

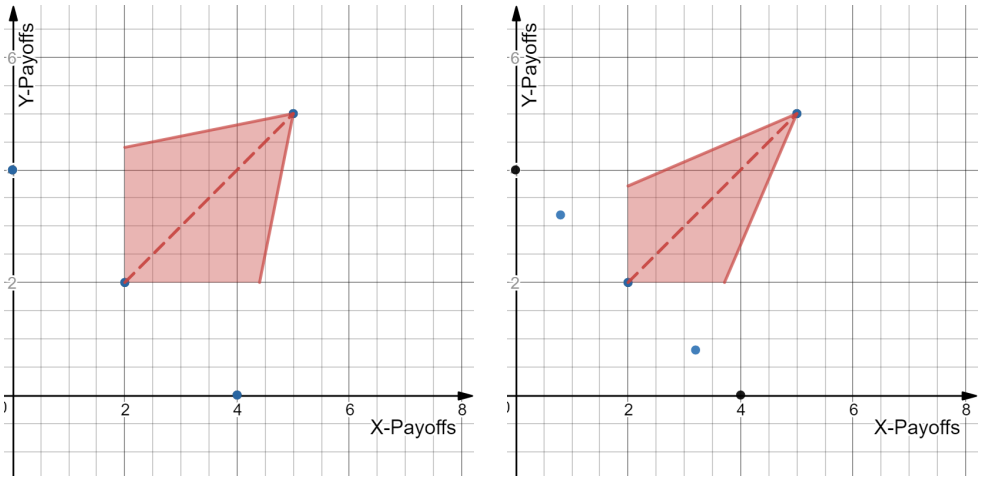

Figure 7.1: Stag Hunt from Table 7.5 with sympathy of $s = 0$ and $s = .2$ and . The black dots represent the original off-diagonal payoffs. Notice how this region of feasible payoffs is slightly smaller as we increase sympathy. This means that agents do not need to be as patient, since they will be less likely to choose outcomes unfavorable to their partner.

Adding Sympathy to the Repeated Stag Hunt

How does sympathy affect the strategies in the game if repeated? Figure 7.1 contrasts the feasible region of equlibria in the repeated Stag Hunt with the region modified by sympathetic payoffs. Here we have the repeated game table with sympathetic payoffs

$$U_X = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 5 & 0 \\ 4 & 2 \end{bmatrix} \end{matrix} \Rightarrow V_X = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & \quad\quad D \end{matrix} \\ \begin{bmatrix} 5 & (1-s)(2\delta)+s(4-2\delta) \\ (1-s)(4-2\delta)+4s\delta & 2 \end{bmatrix} \end{matrix}$$

$$5-((1-s)(4-2\delta)+4s\delta) > 2-((1-s)(2\delta)+s(4-2\delta)) \Rightarrow s > \frac{1}{8}(\frac{1-4\delta}{1-\delta})$$

Looking back at Figure 7.1, we can see the region with $s = .2$, one example of the result where $s > \frac{1}{8}$. If we solve for δ,

$$\delta > \frac{1}{4}(\frac{1-8s}{1-2s})$$

Both of these results for making the *Stag* outcome risk-dominant are consistent with our earlier cases of $s = 0$ in the classical utility models or $\delta = 0$ in the one-shot games. The limiting value for δ in this case is $\delta = .25$, and this boundary function decreases with s until $s = \frac{1}{8}$. What this means is that we require less patient players as sympathy increases. With sympathy greater than this limit, we get nonsensical values for δ. We can see this relationship in the Figure 7.2.

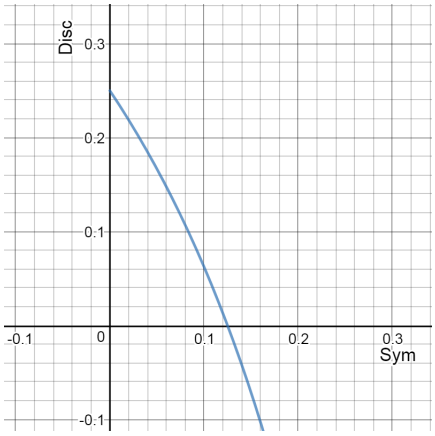

Figure 7.2: Decreasing relationship between sympathy s and the patience requires, or discount, δ.

7.4 Solving the Stag Hunt in a Population

Thinking of actions as strategies, we now turn to some of the Stag Hunt and evolutionary dynamics. Instead of two players choosing between actions, we can think of a population of stag-hunters and hare-hunters. As the game is repeated, we can think of agents learning from their neighbors or strategies being replicated according to their previous success. We discuss the first type of update, learning dynamics.

? discusses two update mechanisms by which a population can learn: *Imitate the Best* and *Best Response*. In one example of *Imitate the Best*, agents play an initial round of the Stag Hunt game, and then they follow their play by observing the utilities of their neighbors on a square lattice. Depending on the initial play, the population can end up in a stable configuration of all playing *Hare* or *Stag*. However, given a relative population of 30 percent playing *Stag*, the population can also develop islands of Stag-hunters in a sea of Hare-hunters, so to speak. If there are islands of Hare-hunters, they can be overwhelmed by a group of Stag-hunters, as when there are two Stag-hunters nearby, they will have the highest

score. Skyrms also goes on to mention that when comparing "imitate the best individual in the neighborhood" vs. realistic "imitate the strategy that performs best on average in the neighborhood", the second strategy can also lead to the dissolution of Hare-hunters in the population.

Considering the dynamics of "best response to your neighborhood", agents play a round of the Stag Hunt, and then they observe the strategy that their neighbors last played. Depending on the topology of the network connections in the neighborhood, this can drive Stag-hunters to extinction quickly. The reason is that Hare-hunters require more than one Stag-hunter in their neighborhood to switch out of their current state. In contrast, if there are not enough Stag-hunters in the population, a player who plays Stag in one round will quickly change to Hare in the next.

The need for a sufficient number of Stag-hunters in the population for the behavior to stabilize is related to our earlier mention of the *replicator dynamics* and the *basin of attraction*, which we now discuss in short.

Evolutionary Game Theory and Replicator Dynamics

As discussed before in Ch 4**??**, games need not be one-shot, rather they can be repeated. Evolutionary Game Theory is similar in that the game is repeated, but in this case strategies replicate proportionally to their utility in the stage game. Agents here are no longer rational, as their actions in the game table represent a behavioral profile and the utilities represent the success of replication for that profile.

Once a game features a potential social contract, we may be interested in the conditions under which that outcome will be replicated in a larger population of agents. Consider the Stag Hunt from **?** in the first entry from Table 7.7. Under what initial conditions would we expect a population to contain only Stag-hunters? Here we turn to the *replicator dynamics* in Equation 7.1.

Formally, for a frequency x of a strategy S in a population, we can write the replicator dynamics as a function $D(x)$ of that frequency. $D(x) = \frac{dx}{dt}$, or the instantaneous rate of change of the population behaving according to S. If $EU_S(x)$ is the expected payoff to the fraction of the population playing S, and $\overline{EU}$ represents the weighted average of the population's performance as a whole, we have the *replicator equation:*

$$D(x) = x(EU_S(x) - \overline{EU}) \tag{7.1}$$

To find the range of initial frequencies that will allow S to take over (the basin of attraction), we compute the points where $D(x) = 0$. I.e. there is no change from one generation of the population to the next. This tells us that

i All agents have converged to a single strategy OR

ii The population is at a resting point, where the strategies will stay in their current proportions.

A	S	H
S	9,9	0,8
H	8,0	7,7

B	S	H
S	5,5	0,4
H	4,0	2,2

I	S	H
S	p^2	$p(1-p)$
H	$p(1-p)$	$(1-p)^2$

Table 7.7: Stag Hunt from **?**vs. a variant from **?**.

If $p = Pr(S)$, will first find $EU_S(p)$ and then $\overline{EU}$. Let A be the set of actions in the game and let $U(S, a)$ denote the utility of S against a. Then let p_a, p_b represent the proportion of agents playing the actions a, b. $EU_S(p)$ will be the utility of only playing S against the population with these proportions. $\overline{EU}$ will be the expected utility summed across each potential strategy pairing (b, a).

$$EU_S(p) = \sum_{a \in A} p_a U(S, a)$$
$$\overline{EU} = \sum_{b \in A} p_b [\sum_{a \in A} p_a U(b, a)]$$

Not surprisingly, if we have $Pr(S) = 0$ or $Pr(S) = 1$, we will see no further change in the proportions, as either the equation does not take mutation into account. Seen simply from the original equation, we would have that $x = 0$ or $x = 1 \Rightarrow EU(S) = \overline{EU}$.

The Stag Hunt has the feature of a *risk-dominant* equilibrium in (H, H). Simply put, if strategies replicate proportionally to their ability to outperform the average strategy, a risk-dominant strategy can take over the population under a wider range of initial frequencies. That range of initial frequencies is the *basin of attraction* under the *replicator equation*. Under games with two strategies, this can be seen with algebra, as we show with Aumann's Stag Hunt vs. a variant from **?** in Table 7.7.

I	S	H
S	$9p$	$0(1-p)$
H	$8 \times 0p$	$7 \times 0(1-p)$

II	S	H
S	$9p^2$	$0p(1-p)$
S	$8p(1-p)$	$7(1-p)^2$

Table 7.8: Stag Hunt from **?**vs. a variant from **?**. If $Pr(S) = p$ and $Pr(H) = 1-p$, each cell has the payoffs shown in II. For $Pr(S) = 1$ and $Pr(H) = 0$, we have table I, whose cells give us $EU_S(p)$.

The expected probabilities and payoffs for each strategy combination in Aumann's Stag Hunt (A) from Table 7.7 can be seen in Table 7.8. In game A, for $p = Pr(S)$ we have

$$EU(S) = 9p; \overline{EU} = 9p^2 + 0p(1-p) + 8p(1-p) + 7(1-p)^2$$
$$D(p) = p(EU(S) - \overline{EU}) = 0 \Rightarrow 0 = -p(8p^2 - 15p + 7))$$
$$\Rightarrow p = 0; p = \frac{7}{8}; p = 1$$

Thus we need at least $Pr(S) = \frac{7}{8}$ for the *Stag* strategy to take over in Aumann's game. There is a different case in game B, as for $p = Pr(S)$ we have

$$EU(S) = 5p, \overline{EU} = 5p^2 + 0p(1-p) + 4p(1-p) + 2(1-p)^2$$
$$D(p) = p(EU(S) - \overline{EU}) = 0 \Rightarrow 0 = -p(3p^2 - 5p + 2))$$
$$\Rightarrow p = 0; p = \frac{2}{3}; p = 1$$

If we look back at the tables from Table 7.7 and the related diagrams from the replicator dynamics, we see that choosing the Stag option is riskier in Aumann's version, as we lose a larger amount if our partner fails to coordinate.

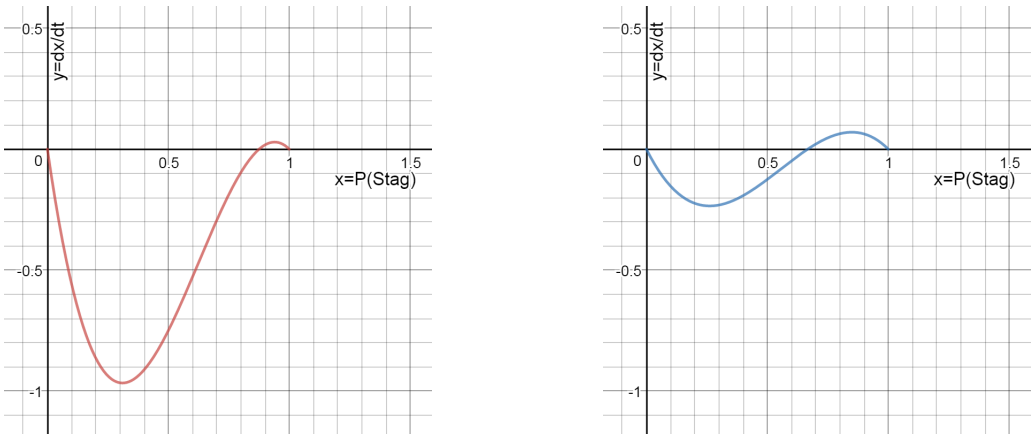

Figure 7.3: Replicator Dynamics Stable States from Aumann's and Skyrms's Stag Hunt in Table 7.7.

General Payoff Symmetric Form Games

	C	D
C	P,P	Q,R
D	R,Q	S,S

Figure 7.4: Replicator Dynamics and Stable States in the Stag Hunt, where $C = Stag$ and $D = Hare$.

In general for 2×2 symmetric form games, we can find the basin of attraction as such, where $x = Pr(C)$.

$$D(x) = x(EU(x) - \overline{EU}) = -x[(P + S - Q - R)x - (S - Q)](x - 1)$$

This implies we have stationary points for the replicator dynamics at $x = 0, x = 1$, and $x = \frac{S-Q}{S-Q+P-R}$. We can think of the proportion s playing *Stag* and the proportion $1 - x$ playing *Hare* as another way of understanding the idea of risk dominance. Recall earlier that *Hare* being risk-dominant would mean $S - Q > P - R$. So comparing the risk of deviating from these strategies would give us $\frac{S-Q}{P-R} > 1$.

$$x = \frac{S - Q}{S - Q + P - R} \Rightarrow \frac{x}{1 - x} = \frac{S - Q}{P - R} \Rightarrow \frac{x}{1 - x} > 1$$

This result of $\frac{x}{1-x} > 1$ is necessary for the replicator dynamics to reach a steady distribution, but it is not sufficient. Consider the example from **?** earlier. Although a proportion of $\frac{Pr(Stag)}{Pr(Hare)} = \frac{6}{4} > 1$, we need that $\frac{Pr(Stag)}{Pr(Hare)} > \frac{7}{1} = \frac{7-0}{9-8}$ for *Stag* to take over.

In other words, for a risky strategy to overtake the population, its initial proportions must be enough to outpace the comparison of the losses faced by deviating from either strategy in the initial table. The only way for the proportions in the population to stabilize with such a mix of strategies is to have exactly the same proportion of agents playing the risky strategy to the safe strategy as the proportion of those two losses. In Aumann's game this means comparing $7-0 = 7$ loss from switching out of *Hare* to the $9 - 8 = 1$ loss from switching out of *Stag*. When the population playing the other strategy is in these proportions ($7 : 1$ for *Stag: Hare*) , the proportions will stabilize. Any more, and the population's proportion of the risky strategy will increase, as we can see $\frac{dx}{dt} > 0$ for those values and $\frac{dx}{dt} < 0$ for values of x less than the stationary point, as we can see from Figure 7.3.

7.5 Marriage Markets as Social Contracts

A forthcoming paper entitled *Losing Face* by Reinstein and Gall[5] depicts a marriage proposal as as a matching problem based on inferences derived from the signals that males and females send each other. Just as we have done, they motivate their analysis on other types of mutually beneficial arrangements like school choice and business partnerships.

Abridged Accept-Reject Model

The basic model without *loss of face* pairs a *Male* and a *Female* against each other, each with the option of playing *Accept A* or *Reject R*. We will refer to this

[5]Based on the earlier **?**

game as the *Accept-Reject Model* from here on.

Matching players (A, A) in Table 7.9 receive utility equivalent to their partner's type, [6] whereas if there is no match, players receive a discounted utility proportional to their own type.

HH	A	R
A	h,h	δh,δh
R	δh,δh	δh,δh

HL	A	R
A	l,h	δh,δl
R	δh,δl	δh,δl

LH	A	R
A	h,l	δl,δh
R	δl,δh	δl,δh

LL	A	R
A	l,l	δl,δl
R	δl,δl	δl,δl

Table 7.9: Reinstein's basic model of *Accept*(A) or *Reject*R proposals from combinations of *High(h)* and *Low(l)* types *without loss of face*. If both accept, utility depends on the other player. If one rejects, players receive a discounted portion of their own type.

Players can observe signals from the other, and they know their own type $x \in \{l, h\}$. The signals s are drawn independently from the continuum $s \in [0, 1]$ according to a distribution based on their type f_x. The type and signal space obeys the following properties:

1. **High types are better off single:** $\delta h > l$.
2. **Monotone likelihood ratio:** $s > s' \Rightarrow \frac{f_h(s)}{f_l(s)} > \frac{f_h(s')}{f_l(s')}$.
3. **Floor and ceiling:** $f_h(\underline{s}) = 0; f_l(\overline{s}) = 0; f_l(\underline{s}) > 0; f_h(\overline{s}) > 0$.

The first condition governs the discount δ, so that high types would be irrational to knowingly marry a low type. The second conditions means that high types are more likely to send higher signals. Last, the third condition means that, given the range of signals possible $[\underline{s}, \overline{s}]$, high types will never send the lowest signal and low types will never send the highest signal. To see a more concrete case (Table 7.10) of Reinstein and Gall's model not in their original paper[7], take the example that

$$U(h) = 10 \qquad U(l) = 5 \qquad \delta = .6$$

Results: High types may not always accept, but low types always do in any trembling-hand equilibrium, as a low type would not be worse off if a high type makes a mistake. Hence high types have to be reasonably confident they are

[6] Their utility model also includes a *pizazz* factor $\tilde{x}$ for a type x, which we omit here for simplicity.

[7] These tables were approved of by the authors.

HH	A	R
A	10,10	6,6
R	6,6	6,6

HL	A	R
A	5,10	6,3
R	6,3	6,3

LH	A	R
A	10,5	3,6
R	3,6	3,6

LL	A	R
A	5,5	3,3
R	3,3	3,3

Table 7.10: Concrete payoffs for Reinstein's basic model of *Accept*(A) or *Reject*(R) proposals from combinations of*High(h)* and *Low(l)* types *without loss of face*. If both accept, utility depends on the other player. If one rejects, players receive a discounted portion of their own type.

not being fooled into marrying a low type, despite this occurring with positive probability. These types make inferences from their partner's signal, choosing to reject unless the signal is above a certain threshold.

As for the resulting matchings, low types only marry if they fool a high type or meet another low type. High types only marry if they are fooled by a low type or meet another high type with a sufficiently strong signal. Thresholds for signals are symmetric for both males and females.

Loss of Face Model

The model then adapts a utility cost entitled *Loss of Face* for a row player that plays accept but is rejected. This factor becomes important in what the authors call asymmetric revelation, where the male (first mover, M) does not know what action the female (F) takes. The stipulation is that:

- M plays Accept while F plays reject,
- F knows M played Accept, and
- M knows all of the above.

The utility tables are very similar except in the outcomes (A, R), where M receives a penalty via *Loss of Face*, $\delta h - L$. We give below the general payoffs (Table 7.11)and then a concrete exampleTable 7.12.

Just as before, agents who marry receive benefit based on their partner's type, while single agents receive benefit based on a discount of their own type. In the concrete example (Table 7.12) of Reinstein and Gall's model, we let

$$U(h) = 10 \qquad U(l) = 5 \qquad \delta = .6 \qquad L = 2$$

Results: The asymmetry in payoffs induces asymmetry in strategies. High type females may not always accept, but low type females always do in any trembling-hand equilibrium, as a low type female would not be worse off if a high type male makes a mistake. Hence high type females have to be confident they are not being fooled into marrying a low type. High type males accept if the female's signal is sufficiently high, but low types will not accept if a female's signal is too high. This is because the potential loss of face means that a low type has to be

HH	A	R
A	h,h	$\delta h - L$,δh
R	δh,δh	δh,δh

HL	A	R
A	l,h	$\delta h - L$,δl
R	δh,δl	δh,δl

LH	A	R
A	h,l	$\delta l - L$,δh
R	δl,δh	δl,δh

LL	A	R
A	l,l	$\delta l - L$,δl
R	δl,δl	δl,δl

Table 7.11: *Loss of Face: Accept*(A) or *Reject*(R) as before from combinations of *High(h)* and *Low(l)* types. If both accept, utility depends on the other player. If one or both rejects, players receive a discounted portion of their own type, with the row player M losing face if he plays accept. Utilities are (M, F).

HH	A	R
A	10,10	4,6
R	6,6	6,6

HL	A	R
A	5,10	4,3
R	6,3	6,3

LH	A	R
A	5,10	1,6
R	3,6	3,6

LL	A	R
A	5,5	1,3
R	3,3	3,3

Table 7.12: Concrete payoffs (M, F) for Reinstein's basic model of *Accept*(A) or *Reject*(R) proposals from combinations of *High(h)* and *Low(l)* types *without loss of face*. If both accept, utility depends on the other player. If one rejects, players receive a discounted portion of their own type.

sufficiently confident that their partner will also accept. The authors point to the creation of "reverse snobs" in this result.

In contrast to the basic model, high type males require a stronger signal from females with loss of face. Further, although the non-matching equilibrium was unstable in the basic model (being rejected and rejecting had the same payoffs), the outcome (R, R) is now stable due to the loss of face for males. This makes high type males more selective than high type females, although the end result of a smaller number of males accepting means that females who do accept are more likely to have a low type male than in the basic model. Last, the model with loss of face induces a lower overall marriage rate and lower overall utility for the population involved.

7.6 Weak Commitment Models

We wish to modify the Stag Hunt and the Accept-Reject Model from Reinstein and Gall to introduce a third option, *Weak Commitment*. In natural language, we indicate this through modal verbs like *would*. The model we give allows a player playing a weak commitment to *free-ride* on the commitment of the other while maintaining some of the security of the risk-dominant strategy. Consider the utterances:

- *I will/want to marry you.*

- *I would marry you (if it's advantageous).*

- *I will not marry you.*

These utterances give us a model for augmenting the game tables from Table 7.10. We would like to add the assumption that agents would still prefer a weak commitment to being single, so we add a value ω to lower the payoff given to the partner but still maintain the benefit of receiving utility from a partner's full acceptance. We will proceed to give the tables with the abstract and concrete values side-by-side. In this case, we now have $h > \omega h > \delta h > l > \omega l > \delta l$.

$$U(h) = 10 \qquad U(l) = 5 \qquad \omega = .8 \qquad \delta = .6$$

HH	A	W	R
A	h,h	ωh,h	δh,δh
W	h,ωh	ωh,ωh	δh,δh
R	δh,δh	δh,δh	δh,δh

HH	A	W	R
A	10,10	8,10	6,6
W	10,8	8,8	6,6
R	6,6	6,6	6,6

Table 7.13: High vs. High: Adding a weak commitment like *I would marry you* gives the game a new set of equilibria.

LH	A	W	R
A	h,l	ωh,l	δl,δh
W	h,ωl	ωh,ωl	δl,δh
R	δl,δh	δl,δh	δl,δh

LH	A	W	R
A	10,5	8,5	3,6
W	10,4	8,8	3,6
R	3,6	3,6	3,6

Table 7.14: Low vs. High: Adding a weak commitment like *I would marry you* gives the game a new set of equilibria.

HL	A	W	R
A	l,h	ωl,h	δh,δl
W	l,ωh	ωl,ωh	δh,δl
R	δh,δl	δh,δl	δh,δl

HL	A	W	R
A	5,10	4,10	6,3
W	5,8	4,8	6,3
R	6,3	6,3	6,3

Table 7.15: High vs. Low: Adding a weak commitment like *I would marry you* gives the game a new set of equilibria.

LL	A	W	R
A	l,l	ωl,l	δl,δl
W	l,ωl	ωl,ωl	δl,δl
R	δl,δl	δl,δl	δl,δl

LL	A	W	R
A	5,5	4,5	3,3
W	5,4	3,4	3,3
R	3,3	3,3	3,3

Table 7.16: Low vs. Low: Adding a weak commitment like *I would marry you* gives the game a new set of equilibria.

Sympathy and Amicability Remove Weak Commitment

If we add sympathetic payoffs to the game with weak commitment, we can remove Nash equilibria that are not an ESS. We return to the high type vs. high type table above, but with a sympathy value of $s = .5$ to simplify calculations. Note that any sympathy value $s > 0$ would have the same effect.

HH	A	W	R
A	10,10	8,10	6,6
W	10,8	8,8	6,6
R	6,6	6,6	6,6

HH	A	W	R
A	10,10	9,9	6,6
W	9,9	8,8	6,6
R	6,6	6,6	6,6

Table 7.17: Adding sympathy $s > 0$ to the game with weak commitment removes non-strict Nash equilibria (boxed). Here $s = .5$ for readability.

The case with the low types would play out in an equivalent fashion. One question might be whether perception of the other player's type could influence the inclusion of a weak commitment strategy or the emergence of sympathy. For instance, if players perceived their opponent to be like them, might we have an increase in sympathy? We can also imagine a dialog like:

Harry: *Would you marry me?*

Sally: *Would I? If you loved me you wouldn't ask like that!*

Relational thresholds like Amicability also play into the Accept-Reject Model. For players trying to maximize the group outcome, any threshold higher than $k > 18$ will only select the mutual acceptance outcome (A, A) with the payoffs of $(10, 10)$.

Weak Commitment with Loss of Face

We continue adapting the *Accept-Reject Model* to weak commitment, this time building in loss of face. To follow our examples from the beginning of the chapter as motivated by **?** and **?**, we will build in the constraint that being rejected under a weak proposal loses less face than being rejected while making a clear proposal with common knowledge. We will call this face cost L_W, and thus have $L_A > L_W$. As before, males have asymmetric exposure to face loss with outcomes

listed (M, F); e.g. the game labeled HL pits a high type male against a low type female.

$$U(h) = 10 \qquad U(l) = 5 \qquad \omega = .8 \qquad \delta = .6 \qquad L_W = 1 \qquad L_A = 2$$

HH	A	W	R
A	h,h	ωh,h	$\delta h - L_A$,δh
W	h,ωh	ωh,ωh	$\delta h - L_W$,δh
R	δh,δh	δh,δh	δh,δh

HH	A	W	R
A	10,10	8,10	4,6
W	10,8	8,8	5,6
R	6,6	6,6	6,6

Table 7.18: High vs. High: Weak commitment with *Loss of Face.*

LH	A	W	R
A	h,l	ωh,l	$\delta l - L_A$,δh
W	h,ωl	ωh,ωl	$\delta l - L_W$,δh
R	δl,δh	δl,δh	δl,δh

LH	A	W	R
A	10,5	8,5	1,6
W	10,4	8,4	2,6
R	3,6	3,6	3,6

Table 7.19: Low vs. High: Weak commitment with *Loss of Face.*

HL	A	W	R
A	l,h	ωl,h	$\delta h - L_A$,δl
W	l,ωh	ωl,ωh	$\delta h - L_W$,δl
R	δh,δl	δh,δl	δh,δl

HL	A	W	R
A	5,10	4,10	4,3
W	5,8	4,8	5,3
R	6,3	6,3	6,3

Table 7.20: High vs. Low: Weak commitment with *Loss of Face.*

Results in the Weak Accept-Reject Model(WAR)

Note that the corners of every 3×3 game table give us the original, allowing us to see what changes in behavior would occur if we restricted the W strategy. If we contrast the equilibrium behavior of the four types of players to their behavior in the original *Accept-Reject Model*, we see that

- Low Type Males are more likely to play W in WAR than A in the AR model, as $U(W) \geq U(A)$ for any type of female. Moreover, W performs equally as well as A when matched against a low type female and significantly better than R would against a high type female, in the event that the high type female is fooled. As W reduces the face loss for the low type males, this should lower the "reverse-snob" effect, whereby low type males might reject a perceived high type female.

LL	A	W	R
A	l,l	ωl,l	$\delta l - L_A$,δl
W	l,ωl	ωl,ωl	$\delta l - L_W$,δl
R	δl,δl	δl,δl	δl,δl

LL	A	W	R
A	5,5	4,5	1,3
W	5,4	3,4	2,3
R	3,3	3,3	3,3

Table 7.21: Low vs. Low: Weak commitment with *Loss of Face.*

- Low Type Females should behave similarly to low type males, as no matter who their partner is, they will be indifferent to playing A or W.

- High Type Males are more likely to play W in WAR than A or R in the AR model when sufficiently unsure of their partner's type. This because W weakly dominates A against a high type female, but W does not perform significantly worse against a low type female. This should create an additional threshold in the best response curve for these types. Thus this raises the threshold for playing A and lowers the threshold for playing R, creating a situation where a high type male might weakly commit to low type females sending higher signals and high type females sending lower signals.

- High Type Females will be indifferent to playing A or W in the WAR model, but against a low type male, they will always prefer R. That W weakly dominates the other actions for all types of males also means that females should have a higher threshold for choosing when to play A than in the AR model.

Weak Commitment in the Stag Hunt

Similar to the modifications of Reinstein and Gall's Accept-Reject model, the Stag Hunt gives us another chance to adapt the weak commitment payoffs to a similar case of social contracts.

Stag	S	W	H
S	5,5	4,5	0,4
W	5,4	4,4	2,4
H	4,0	4,2	2,2

Stag(s)	S	W	H
S	5,5	4.5, 4.5	2,2
W	4.5, 4.5	4,4	3,3
H	2,2	3,3	2,2

Table 7.22: Stag Hunt with Weak Commitment: Nash Equilibria (boxed). Sympathy removes many equilibria. $s = .5$ for readability in the second game.

Weak commitment adds a strategy that takes advantage of the other player's cooperation but minimizes risk. Adding sympathy $s > 0$ to the Stag Hunt with weak commitment removes non-strict Nash equilibria, as seen before.

7.7 Conventionalization and Signaling

As mentioned earlier, we will take the archaic meaning of *will* to include the modern connotation of *want*, as in examples from the King James Bible:

> For ye have the poor with you always, and whensoever ye **will** ye may do them good: but me ye have not always. ... And he said, Abba, Father, all things are possible unto thee; take away this cup from me: nevertheless not what I **will**, but what thou **wilt**. –Mark 22

Given that the semantics of *Would you marry me?* could include the meaning of *Will you marry me?*, why do we see the distinction in the pragmatics? We claim that this follows the same pattern as scalar (*some-all*)implicatures, and thus we can resolve the conventionalization process with *iterated best response (IBR)*, as seen in **???**.

Historical Background on English Marriage Proposals

Writings like the English *Book of Common Prayer*(BCP), as inspired by the *Sarum Rite*, and the works on William Shakespeare give more instances of *will* indicating preference over possibilites, as seen in the ritualized taking of marriage vows from the BCP from 1549:

> Wilte thou have this woman to thy wedded wife, to live together after Goddes ordeinaunce in the holy estate of matrimonie? Wilt thou love her, coumforte her, honor, and kepe her in sickenesse and in health? And forsaking all other kepe thee only to her, so long as you both shall live?

The text from the Sarum Rite includes the address to the man as:

> **Priest:** *Vis* habere hanc mulierem in sponsam?
> Do you wish to have this woman in marriage?
>
> **Man:** *Volo.*
> I so wish.

The works of Shakespeare contain several instances of the *thou wilt* construction revealing the preferences of the speaker or inquiring as to the preferences of the hearer. We give some examples below, with the name of the speaker in boldface and play in italics.

> **Countess:** Tell me thy reason why thou wilt marry. (*All's Well That Ends Well*)
> **Menas:** Wilt thou be lord of all the world? (*Antony and Cleopatra*)
> **Oliver:** Wilt thou lay hands on me, villain? ... And wilt thou have me? ... You say you'll marry me, if I be willing? (*As You Like It*)
> **Iachimo:** Wilt thou hear more, my lord? (Cymbeline)
> **Hamlet:** Or if thou wilt needs marry, marry a fool. (Hamlet)

Henry : By mine honour, in true English, I love thee, Kate. ... Wilt thou have me? (*Henry V*)
Juliet: I will not marry yet; and, when I do, I swear, It shall be Romeo! (*Romeo and Juliet*)
Petruchio: And will you, nill you, I will marry you. (*Taming of the Shrew*) [8]
Miranda: I am your wife, if you will marry me. (*Tempest*)

These examples, in particular those relating to marriage, give us more evidence of the preference-based reading. That said, what does revealing one's preferences indicate as far as strategy? And what might uncertainty of a partner's preferences do to change the outcome? These questions we explore as we discuss the conventionalization of *Will you marry me?*

Iterated Best Response on Possibility

Work from **?** and **?** can connect the semantics of *Will* to *Would* in two of the following ways. Take the semantics of *I would marry you* to mean *There exist possible worlds in which I would want to marry you.* Then take the semantics of *I will marry you* to mean *For all possible worlds I want to marry you.*

Scalar Implicatures from ?, ?

Consider the Some-All singaling game from **???**, interpreted first as a declaration of the male's preferences. We take here the messages to be costless and the utility of both agents to be totally aligned with the receiver guessing the sender's type correctly. I.e. for either agent i

$$U_i = \begin{matrix} \\ \exists\neg\forall \\ \forall \end{matrix} \begin{matrix} \begin{matrix} \exists\neg\forall & \forall \end{matrix} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{matrix}$$

Now proceeding with the assumption of successively higher levels of rationality as **?**, we give the best responses to a naive sender S_0 by a level-1 receiver R_1, and so on.

$$S_0 = \begin{Bmatrix} \exists\neg\forall \Rightarrow Would \\ \forall \Rightarrow Would, Will \end{Bmatrix} \qquad P_0(m|t) = \begin{matrix} \\ \exists\neg\forall \\ \forall \end{matrix} \begin{matrix} \begin{matrix} Would & Will \end{matrix} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix} \end{matrix}$$

$$R_1 = \begin{Bmatrix} Would \Rightarrow \exists\neg\forall \\ Will \Rightarrow \forall \end{Bmatrix} \qquad P_1(t|m) = \begin{matrix} \\ Would \\ Will \end{matrix} \begin{matrix} \begin{matrix} \exists\neg\forall & \forall \end{matrix} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{matrix}$$

[8] This expression is a corruption of *Volle-Nolle*, spoken in modern English as *Willy-Nilly*, meaning whether the hearer is willing or not.

$$S_2 = \begin{Bmatrix} \exists\neg\forall \Rightarrow Would \\ \forall \Rightarrow Will \end{Bmatrix} \qquad P_0(m|t) = \begin{matrix} \\ \exists\neg\forall \\ \forall \end{matrix} \begin{matrix} \begin{matrix} Would & Will \end{matrix} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{matrix}$$

$$R_{k+1} = R_{k-1} \qquad S_{k+1} = S_{k-1}$$

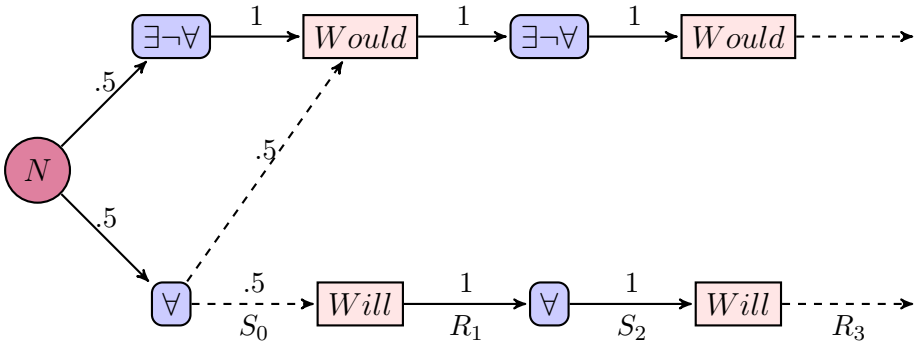

Figure 7.5: Iterated Best Response on *Will vs. Would*.

Lifting to Epistemic Models and Inquisitive Semantics

Questions specify partitions of possible worlds that are acceptable answers. In our case, asking *Will you marry me?* can only have a Yes/ No answer, but asking *Would you marry me?* could have a multitude of answers and additional specifications. The messages in fact partition the available action space when it comes to the receiver's reply, so now we augment the model with inquisitive semantics and the notes from **?** on resolving decision problems.

Our questions are:

- How do we depict this with an extensive form game?
- How does being unsure of one's partner's preference indicate belief of one's dispreference?
- How do face (positive/ negative) costs play in here? Rejection carries loss of positive face but acceptance means gain of positive face.
- Do Horn's rules on costly expressions vs. frequency come in here? (We don't marry every day)

Self-Enforcing Equilibria and Self-Signaling Reputation

Following the background from **?**, we discuss how the conventionalization predicted by the signaling game literature stabilizes the convention as *self-signaling.*

Self-Signaling Set: A set of types C with the property that precisely types in the set C gain from inducing the best response to C relative to a fixed equilibrium.

Although we have looked at games like Aumann's Stag Hunt as a variant of the social contract, and his claim is that these games do not admit self-signaling sets, in our case, talk is anything but cheap, as face concerns reflect poorly on a partner who backs out of the Pareto-optimal state, whether it be in business or in marriage.

7.8 Questions for Editors/ Notes

- I have some musings on a Bayesian-style game with four possible information states similar to the Accept-Reject model. This model would be like four 2×2 games, where the male and/or female want/ don't want to get married, as opposed to high/ low types. The male would move first, but what would "will"/ "would" indicate? Should there be different distributions of these states? One state would be the Stag Hunt, as both would want to accept but might lose face if the other rejected.
- Much like the implicatures seen in **?** or **?**, one question I have is whether the statement "I am not sure whether you want to marry" might mean "I am sure you don't want to marry" or "I am sure I don't want to marry", or the disjunction.