Questions of Trust

Jason Quinley[1] and Christopher Ahern[2]

[1] University of Tübingen
`jason.quinley@student.uni-tuebingen.de`
[2] University of Pennsylvania
`cahern@ling.upenn.edu`

Abstract. We consider the application of Game Theory in the modeling of different strategies of politeness. In particular, we examine how differences in the linguistic form of requests and proposals map onto the structure of the game being played by interlocutors. We show how considerations of social wants [1, 2] and coordination and cooperation motivate these differences. First, we adapt the notion of other-regarding preferences [3, 4] to show how linguistic systems with the ability to encode politeness strategies allow for requests and cooperation between a wider range of individuals. Finally, we connect the distinction between requests and proposals to the notion of a self-enforcing equilibrium [5].

1 Introduction

Questions in their many forms are central to social interaction. Asking someone for a dollar, or if they would like to see a movie, are commonplace yet revelatory. The indicate how we use language not only to convey information, but also to negotiate relationships. A clear case of these distinctions can be had in the use of the modals *will* and *would* in the following requests:

(1) Will/Would you lend me a dollar?
(2) Will/Would you open the door?
(3) Will/Would you turn that music down?
(4) Will/Would you marry me?

Consider asking these questions of a stranger. It would be impolite to omit the modal. Moreover, between the two modals we also sense a difference in effect. In the first two cases *would* is the more polite form of request. In the third either is acceptable, modulo the degree to which the music affects the speaker. In the last, *will* seems the more appropriate form. Moreover, *would* allows for comedic response: I would if you were rich/handsome/x!

Why are such questions necessary? One reason is that scarcity and ambiguity drive interaction. We have neither unlimited resources nor unlimited information with which to achieve our ends. This leads to the need for cooperation, and with it, strategies to address its fragility. As humans have access to language, they are availed of multiple avenues of cooperation. Studying these allows for a fruitful combination of theories of language and rational interaction.

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. We begin by presenting the relevant notions from politeness- and game theory, then turn to our analysis of requests in these terms and suggest future directions.

2   Politeness Theory and Speech Acts

Beginning from Goffman's [1] notion of *face*, Brown and Levinson [2] articulated an ur-theory of politeness, which has prompted much subsequent theoretical and empirical work. *Face* is the term given to an individual's basic needs, characterized broadly as the need for autonomy (negative face) and acceptance (positive face). Broadly, positive face can be thought of as the wants of the individual, including the desire that those wants be desirable to or approved of by others. Negative face includes both the freedom of action and the freedom from imposition.

Preferences of one agent may conflict with those of others, incentivizing them to make requests, issue threats, or offer proposals. In cases where a request must be made, speakers must commit a *face-threatening act* (*FTA*). In order to mitigate the weight of a FTA, speakers may use several strategies, as laid out in Figure 1.
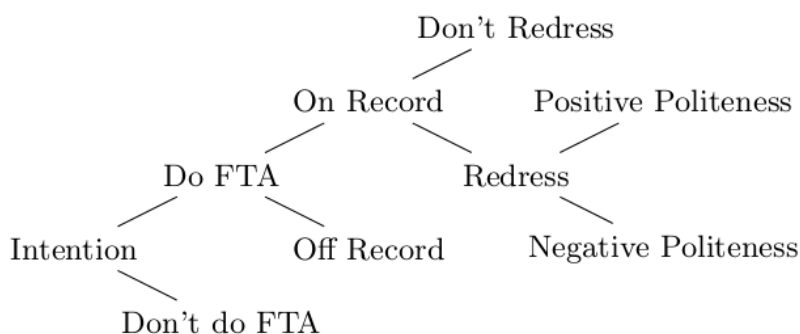

Fig. 1. Brown and Levinson's Politeness Strategies: As we move upwards on the graph, the potential for a *face-threatening act* (FTA) increases.

At the two extremes, a speaker might avoid making the FTA altogether, or state it in a direct manner. In between there are various degrees of deference to the hearer's face wants: indirect speech that is "off the record" and addressing the hearer's positive or negative face. As a concrete example, consider the situation of having left one's wallet at the office while going out to lunch with a group. Here the relevant FTA might

be taken as requesting some money from a friend. The various strategies of doing so could be implemented as:

(1) **Don't do FTA:** Don't ask for money.
(2) **Off Record:**[3] "Oh no! I forgot my wallet in my office!"
(3) **Negative Politeness:** "You don't have to, but would you mind lending me a bit of money?"
(4) **Positive Politeness:** "Congratulations on the raise! Want to lend me some money."

The goal of the speaker is to craft the appropriate message to convey the intent and the weight of the FTA. The greater an imposition a FTA carries, the more care needs to be taken. However, too much politeness is inappropriate given certain FTAs. It would seem odd to be asked, rather circuitously, "Excuse me Sir/Ma'am, but I was hoping that it might be possible if it's not too much trouble that you would be able to tell me the time." Similarly, when expediency is called for, "Please, if you could, move out of the way of that speeding car," would be inappropriate. Thus we might think of different forms of politeness as strategic responses to situations where face may be threatened. Again, the goal of the speaker is to select the appropriate form for the FTA in question; neither too much nor too little deference can be paid. With this notion of strategy in mind we turn to the game-theoretic framework that will figure in our analysis.

3   Basic Game Theory

Game theory gives a mathematical model of strategic interaction between agents. We begin by presenting canonical examples from the field. Crucially, we focus on the difference between cooperation and coordination in sequential play. Sequential play allows for an optimal outcome under rational behavior in cases of coordination, but not for cooperation.

Formally, a sequential game is a tuple $\langle N, O, A_j, U_i \rangle$. N is the set of players in the game. O is a sequence over N that determines the order of play; for $j \in O$, A_j is the set of actions available to the jth player in the order of play. Finally, U_i is a preference for player i over the set of possible paths of play. The payoffs are represented as numeric values, where higher values are taken to be more preferred outcomes. The Prisoner's Dilemma (PD) offers the canonical example of choosing between cooperation and defection. The game reflects a scenario wherein two prisoners must choose between cooperatively staying silent (C), telling the police nothing about their crime, or defecting on each other (D) and confessing the details to the police. Jointly, both prisoners do better if they remain silent, but individually, they do better by ratting out their accomplice. We represent the structure of the Prisoner's Dilemma in *extensive form* in Figure 2. Each node in the game is labeled with the letter of the player whose turn it is to take an action, $O = \langle X, Y \rangle$. The payoffs, as determined by the utility functions, are listed as (U_X, U_Y) at the bottom of the tree.

We use the notion of a *rollback equilibrium* to examine expected behavior in the game. The reasoning proceeds as follows. We begin by considering the lowest nodes in the game and putting in bold the best action available. For any node where Y can make a choice, she should choose D as it is always the better option. Knowing that

---

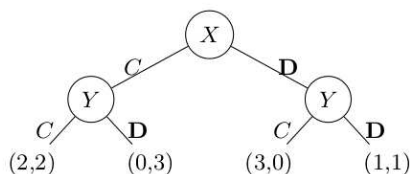[3]See Pinker et al. [6] as well as Mialon [7] for game-theoretic treatments of indirect speech.
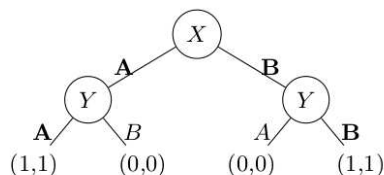
Fig. 2. A Sequential PD   **Fig. 3. A Sequential PCG**

this is the case, X should always choose D, as it always the best option based on what Y will do. That is, cooperation will only ever be met with defection, so it should never be explored. We will refer to those instances where players have diverging interests but come together to yield the optimal outcome for all as instances of cooperation. In instances of cooperation, as in the PD, reaching the best outcome for the players as a group requires some sacrifice in terms of individual payoff. That is, each player must forgo the temptation payoff of defecting in order to maintain cooperation.

In contrast, in cases of coordination, players' incentives do not conflict. Consider the case of a Pure Coordination Game (PCG) in Figure 3. Here players are ambivalent between the actions they take, they only prefer to take the same action. An example might be a scenario where two friends want to meet up for lunch at noon. If one player suggests a restaurant, then the other should indeed go to that restaurant. If X plays A (B), then Y should play A (B). Sequential games allow for the optimal outcome for both players in pure coordination games.

These game structures allow us to distinguish between the notions of cooperation and coordination. We might think of the first as exemplified in the Prisoner's Dilemma, and the latter in the case of Pure Coordination Games. With this background in place, we now turn to the analysis of modals in requests and examine the different rationales for polite behavior.

4   Trust and Modals

In this section we show how face-addressing forms allow for requests between a wider range of individuals. This serves as a broad motivation for using such forms with strangers. We then turn to a distinction between requests in general and marriage proposals in particular, where we argue for a distinction between the different sorts of speech acts involved as they relate to the notion of self-enforcing equilibria.

4.1   Requests as Extended Trust Games

Quinley [8] adapts Trust Games [9] as a model of requests. We borrow techniques and insights from this approach and introduce Extended Trust Games to capture the sequential dynamics of requests. We note the effects of repetition, reputation, and observation on polite forms in requests, but suggest that they are not sufficient to fully explain the use of politeness strategies. Instead, we propose other-regarding preferences

as a means to explain the use of modals and other forms of linguistic politeness in a variety of situations.

Trust games are an appropriate model for requests due to several factors. First, individuals are rarely if ever entirely self-sufficient. Moreover, agents possess different aptitudes and abilities, and this asymmetry prompts requests. Requests entail a loss of face on the part of the requester; so to speak, the requester makes a face "payment" to the requestee. Finally, the requestee is not obligated to grant the request, presenting the agent in need with the risk of both a loss of face and having their request denied.

Trust games depict a scenario where Player X has an initial option to defer to Player Y for a potentially larger payoff for both. We extend this notion further, incorporating a third step in the order of play. Here the play of the game is shown in extensive form in Figure 4 and consists of the first player asking or not asking for some favor, the second player granting or not granting the request, and the requester thanking or not thanking the requestee. We considered a more detailed motivation of the utility structure below.

If X does not ask ($\neg A$), then the status quo remains and X is left to her own devices. Let c_x be the cost to X to achieve the desired outcome. Let c_y be the cost to Y to achieve the same outcome. As noted before, assume an asymmetry in ability or disposition such that $c_y < c_x$; Y is in a better position than X to bring about X's desired state of affairs. If X does ask (A) for help, using a polite request, Y should experience some boost in self-esteem based on the attention received. That is, by acting in accordance with Y's face wants, X increases Y's face. Let the amount of face paid by X to Y in the request be f_r. Let m_r be a multiplicative factor that acts upon f_r to determine the payoff to Y. If talk is cheap, then flattery is certainly sweet; a little bit of face goes a long way, so we assume that $m_r > 1$. Even if Y chooses not to grant the request, Y still comes away with some benefit based on the face paid by X, $m_r f_r$. If Y denies the request ($\neg G$), X has incurred the face cost of asking without receiving any benefits, and must also bear the cost of performing the action, c_x. If Y chooses to grant the request (G), then Y incurs some cost of the action, but still receives the benefit of face from X. Let the benefit to X of Y granting the request be b_x. In general, we assume that $b_x < c_x$. If the request is granted, then X has an opportunity to express to Y some sort of thanks (T) or not ($\neg T$). This expression of thanks, again, comes at some cost f_t, and, again, carries with it some face benefit to Y as determined by a factor $m_t > 1$.

We are faced with the same problem as the prisoner's dilemma; requests are FTAs that require cooperation. X prefers $\neg T$ to T, Y prefers $\neg G$ to $\neg T$, and X prefers $\neg A$ to $\neg G$. Thus, if we are only maximizing individual utility, it never makes sense in a one-shot scenario to ask, grant, or thank, even though both players might prefer the interaction under certain assumptions. We thus consider the effect of repetition, reputation, and observation on the outcome.

4.2   Repetition and Reputation

Under various conditions, repetition engenders cooperation [10]. More specifically, with a given probability of another round of play, group welfare becomes individual welfare; i.e. a PD becomes a Stag Hunt [11]. In a Stag Hunt, players' interest are highly aligned, and the only pitfall is the possibility of mis-coordination. Importantly, in a Stag Hunt players wish to coordinate, but may not necessarily know how when playing simultaneously. In Figure 5, the Stag Hunt structure assumes that players have aligned preferences, and shared preferences over outcomes.
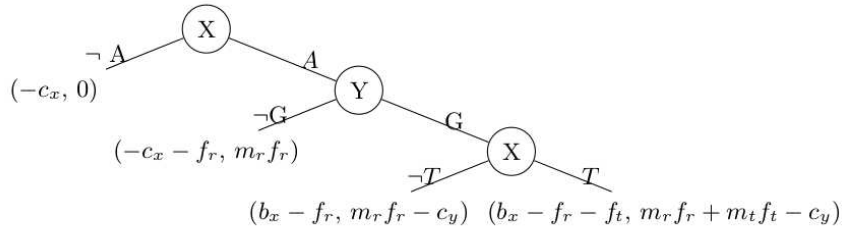
Fig. 4. Request Trust Game: Player X can choose to Ask (A) something from Player Y, who can then choose to Grant (G) the favor. Player X can choose to Thank (T) or not Thank ($\neg T$) player Y.

		Stag	Rabbit
Stag		4,4	0,1
Rabbit		1,0	1,1

Fig. 5. Stag Hunt (SH):
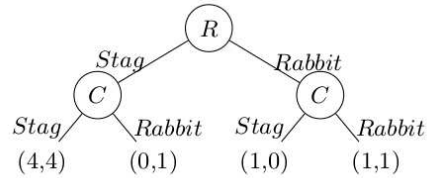in strategic form

Fig. 6. Stag Hunt (SH):
in extensive form

Here, as in the case of coordination games, sequential play allows the players to achieve the optimal outcome. That is, if the first player plays $Stag$, then the second player should as well. Repetition transforms a Prisoner's Dilemma to a Stag Hunt. However, repetition cannot be all there is to the outcome of the interactions we consider here. People are polite to strangers they will never see again.

The effects of reputation and observation on different strategies in trust games are explored in Quinley [8]. Namely, asking requests of other agents is rational when there is a sufficient likelihood that the request will be granted based on the requestee's reputation. Or, in the case here, granting requests is rational when there is sufficient likelihood that X will play T. If Y has sufficient experience or knowledge about the behavior of X with regards to $Pr(T)$, then this suffices to render granting the request rational strategy or not. The novel contribution of Quinley is the inclusion of face effects due to third-party observation. In line with experimental results [12], such observation, framed as a loss (gain) in face for Y when denying (granting) a request, is shown to ensure that requests are asked and granted by and large. This can be extended similarly to X's actions when choosing to thank Y or not.

4.3   Other-Regarding: Reciprocity Without Repetition

While reputation and observation offer rationales for asking and granting requests, they are unlikely to explain all of the behavior we observe. Modals are used to make requests in one-shot interactions where nothing is known about the other individual and there are no third-party observers. In fact, the polite use of modals is even more expected in these sorts of situations. This suggests that reputation and observation are not alone in explaining the behavior observed in requests A rationale for politeness strategies in such situations can be found when we consider other-regarding preferences.

There exists a wealth of theoretical work on [13–15], and behavioral [16, 17] and neurobiological evidence [18] of other-regarding preferences. Here we adapt the notion of sympathy as advanced by Sally [3, 4] to explain the observed behavior. The central notion is that of a sympathy distribution over the payoffs of all the agents involved in the game. For each agent, there is a distribution, $\delta_i \in \Delta(U)$, such that $\sum_j \delta_i(U_j) = 1$, which determines how much that agent cares about her own payoffs and those of others. For example, the perfectly self-interested agent of classical Economic theory is such that $\delta_i(U_j) = 0$ for all $j \neq i$. A selfless agent would be such that $\delta_i(U_i) = 0$.

Here we consider the limiting case of a single interlocutor. Based on the sympathy distribution and the utility function U of the original game, we define a new utility function V.

$$V_i = \delta_i(U_i) \cdot U_i + (1 - \delta_i(U_i)) \cdot U_j \tag{1}$$

The impact of other-regarding preferences can be seen in the following. Consider what values would suffice to make thanking rational for X in the Extended Trust Game of Figure 4. Namely, we wish to determine the condition under which the sympathy distribution of X renders thanking (T) preferable to not thanking ($\neg T$). This holds just when:

$$V_x(\neg T) < V_x(T)$$
$$\frac{1}{1 + m_t} < \delta_x(U_y) \tag{2}$$

Given that $m_t > 1$, the highest threshold will be bounded from above by $\frac{1}{2}$. As m_t increases, the threshold approaches 0. The greater the benefit to Y for thanking, the less X has to care about Y's payoff to do so. This undoes some of the unraveling effect of divergent preferences. We move on to determine the conditions on Y's preferences that suffice to allow for cooperation. That is, we wish to determine when Y prefers T to $\neg G$.

$$V_y(\neg G) < V_y(T)$$
$$\frac{(c_y - m_t f_t)}{(c_y - m_t f_t) + b_x + c_x - f_t} < \delta_y(U_x) \tag{3}$$

There several important points to consider. First, the thresholds we have outlined here are the conditions under which the underlying game of cooperation is transformed into one of coordination. That is, if these thresholds are surpassed, then the game is one of coordination rather than cooperation, and we should expect requests to be made, granted, and thanks expressed. If X's condition on thanking is not met, then the request should be granted if Y prefers $\neg T$ to $\neg G$, which is true just when:

$$V_y(\neg G) < V_y(\neg T)$$
$$\frac{c_y}{c_y + b_x + c_x} < \delta_y(U_x) \tag{4}$$

Otherwise, we should expect requests not to be made.

Second, we find that if $m_t f_t > c_y$, then the request should be granted for anyone, regardless of the sympathy distribution. We might be tempted to think of T in terms of expressing a future commitment to cooperation. While T has this flavor, it does not have this force; thanks, like talk, are cheap. For the expression of thanks to outweigh the cost, c_y, would require either something particularly important to Y, or some strong guarantee on the part of X. Again, future guarantees are not available in the case of single interactions with strangers.

Third, note that $c_x - f_t > c_y - m_t f_t$ given that $c_x > c_y$ and $m > 1$. Moreover, note that $b_x > c_x$, and thus $b_x > c_y - m_t f_t$. As such, as we collapse the non-fixed values towards zero, we see that $\frac{1}{3}$ serves as an upper bound on the threshold. The use of a face addressing form, such as the polite use of modals, allows for a lower threshold of other-regarding preferences for requestees compared to requesters. This makes intuitive sense as requesters are more inherently self-interested.

Finally, the use of politeness strategies that address face allow for a lower threshold than a system without such forms. Consider a faceless Trust Game, where $f_t = f_r = 0$ for the payoffs in Figure 4. We can think of this as a system where no transfers of face are possible. The structure of the game reduces to a choice on the part of Y between granting or not granting the request. The corresponding threshold of other-regarding preference can be given as follows:

$$V_{y'}(\neg G) < V_{y'}(G)$$
$$\frac{c_y}{c_y + b_x + c_x} < \delta_{y'}(U_{x'}) \tag{5}$$

From Eq. (3) and (5) we know that a system with face requires a lower sympathy threshold than one without face just when:

$$\delta_y(U_x) < \delta_{y'}(U_{x'})$$
$$\frac{c_y}{b_x + c_x} < m_t \tag{6}$$

Given $c_y < c_x$, we know that $\frac{c_y}{b_x + c_x} < 1$. Since, $m_t > 1$, it is always the case that a system with face requires a lower threshold than a system without, thus allowing for requests between a wider range of individuals. In this sense, when considered in the context of other-regarding preferences, face allows for cooperation by smoothing out the payoffs of the interlocutors. By "investing" in each other's face, we can guarantee cooperation more easily, even with people we do not know.

4.4   Proposals and Credible Signaling

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. 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!!!*
> **Xavier:** *Woah, I was just asking hypothetically!*
>
> **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.*

We argue that *will* and *would*, for the most part, have the same illocutionary force. However, they differ in that *would* allows for disavowal. To tease out how they do differ, we consider the notion of self-enforcing equilibria.

Aumann considers the game in Figure 7 with pre-play communication. The game has two *Nash Equilibria*: combinations of actions from which neither player can profitably deviate from unilaterally. The equilibria are (C, C) and (D, D). It would seem that both players should settle on playing C, since it is the *payoff dominant* equilibrium. However, this outcome is not guaranteed, even with communication. Suppose both players agree to play C. Suppose X pauses to think about Y. If Y does not trust him, then Y will play D despite the agreement to play C. Y would still want X to play C regardless of what Y does. So, just because both players have agreed to play C, it does not mean that they will; the agreement and the associated equilibrium are not self-enforcing.

	C	D
C	3,3	0,2
D	2,0	1,1

Fig. 7. Aumann's Game

	M_r	M_i
A_r	$v - f_n,\ v + f_n$	$-f_n - f_p, f_n$
A_i	$0, -f_p$	0,0

Fig. 8. Adjusted Aumann's Game

In light of the dialogues above, we might think of the strategies available to X(avier) as either asking for information (A_i) or asking as a request (A_r). Similarly, think of the strategies available to Y(vonne) as interpreting the question as asking for information (M_i) or as a request (M_r). We motivate the utility structure as follows. Suppose that (A_i, M_i) results in some baseline payoff where both players receive 0. Now, suppose that X intends the question as a request, A_r, but Y takes it as a request for information, M_i. X has made some effort to address Y's negative face, and thus is out some effort, f_n, which is transferred to Y. Moreover, X is embarrassed by the miscommunication and loses some amount of positive face because Y does not have the same wants as

him. Similarly, if Y assumes a request, but X does not, then Y loses some amount of positive face. Finally, when X intends a request and Y interprets it as such, then both achieve some payoff, v, modulo a transfer of negative face. These payoffs are given in Figure 8.

On the reasonable assumption that $v > f_n$, there are two pure Nash equilibria: (A_i, M_i), (A_r, M_r). The payoff dominant equilibrium, (A_r, M_r), is not self-enforcing. X prefers for Y to play M_r regardless of what X intends to do; Y prefers for X to play A_r regardless of what Y intends to do. Thus, we can predict the disavowals that occur. However, by and large, we do commit ourselves to making requests, A_r, with *would* and this is because other-regarding preferences transform the payoff structure. The use of *would* is self-enforcing just in case $0 < \delta_x(U_y)$ and $\frac{f_n}{2f_n + f_p} < \delta_y(U_x)$. There are two things to note. First, the comedy of the dialogues above stems from the mismatch between a generally expected amount of sympathy and that displayed. Second, the disavowal on the part of the speaker seems far crueler than what could be an honest mistake on the part of the hearer, as predicted by the fact that $\delta_x(U_y) < \delta_y(U_x)$.

The crucial distinction between *would* and *will*, and why *will* is the appropriate choice for a marriage proposal is evidenced by the effect of not paying negative face to the hearer, as in Figure 10. That is, in a marriage proposal, (A_r, M_r) is a self-enforcing equilibrium much like the classical Stag Hunt, where both players benefit by coordinating on the payoff-dominant choice.

		$Stag$	$Rabbit$
$Stag$		4,4	0,1
$Rabbit$		1,0	1,1

Fig. 9. Stag Hunt (SH): in strategic form

		M_r	M_i
A_r		v, v	$-f_p, 0$
A_i		$0, -f_p$	0,0

Fig. 10. Marriage Game

Thus, using the modal *will* ignores the listener's negative face, but renders the request self-enforcing. This aligns perfectly with our intuition that one cannot back out after asking "Will you marry me?". Moreover, this reasoning about face and other-regarding preferences provides a rationale for why commissive speech acts are possible, and the form they take.

4.5   Summary

We have shown that the transfer of face via politeness strategies with other-regarding preferences allows requests and trust between a wider range of individuals. Specifically, we have shown the necessary amount of sympathy between two individuals that suffices to transform a game of cooperation into one of coordination, and that face lowers this threshold. In addition, we have shown that *would* and *will* differ fundamentally in terms of illocutionary force, and the underlying structure of the interaction. *would* allows for disavowal and is not necessarily self-enforcing, whereas *will* as a commissive speech act commits the speaker to a course of action. In parallel to results from dynamic epistemic logic [19], saying *will* creates common knowledge between the participants of the hearer's commitment to future action, and thus it is only rational in the case that both participants have a benefit towards taking that action and that the action cannot be repeated.

5   Conclusion

This work follows in the vein of approaches to pragmatics and politeness from a strategic viewpoint. It defines the conditions under which politeness strategies are rational in those situations where repetition, reputation, and observation do not hold. A central result is that a system with face allows for a greater level of trust between agents with other-regarding preferences. Also, it outlines how the modals *will* and *would* map onto fundamentally different game structures and predicts both the humorous possibilities of denial and the real power of socially-binding statements. Future directions include extending the current analysis to threats such as *Will you cut that music out!* and requests for information *Will you be here later?*, and providing a broader theoretical framework for the description of speech acts. The results presented here demonstrate the growing ability of game-theoretic methods to model pragmatic phenomena, including politeness. Moreover, though reciprocity and coordination existed outside of and prior to language, language nonetheless serves as an efficient tool for managing them in relationships.

References

1. Goffman, E.: Interaction Ritual: Essays on Face-to-Face Behavior. 1st pantheon books edn. Pantheon, New York (1982)
2. Brown, P., Levinson, S.C.: Politeness: Some universals in language use. Cambridge University Press, Cambridge (1978)
3. Sally, D.: A general theory of sympathy, mind-reading, and social inter- action, with an application to the prisoners' dilemma. Social Science Information **39**(4) (2000) 567–634
4. Sally, D.: On sympathy and games. Journal of Economic Behavior and Organization **44**(1) (2001) 1–30
5. Aumann, R.J.: 34. In: Nash Equilibria Are Not Self-Enforcing. Elsevier (1990) 667–677
6. the National Academy of Sciences of the United States of America: The logic of indirect speech, the National Academy of Sciences of the United States of America (2007)
7. Mialon, H., Mialon, S.: Go figure: The strategy of nonliteral speech. (2012)
8. Quinley, J.: Trust games as a model for requests. ESSLLI 2010 and ESSLLI 2011 Student Sessions. Selected Papers (2012) 221–233
9. Berg, J., Dickhaut, J., McCabe, K.: Trust, reciprocity, and social history. Games and Economic Behavior **10**(1) (July 1995) 122–142
10. Mailath, G.J., Samuelson, L.: Repeated games and reputations: long-run relationships. Oxford University Press, Oxford (2006)
11. Skyrms, B.: The Stag Hunt and the Evolution of Social Structure. Cambridge University Press, Cambridge (2004)
12. Fehr, E., Fischbacher, U.: Third-party punishment and social norms. Experimental 0409002, EconWPA (September 2004)
13. Rabin, M.: Incorporating fairness into game theory and economics. American Economic Review **83**(5) (December 1993) 1281–1302
14. Fehr, E., Schmidt, K.M.: A theory of fairness, competition and cooperation. CEPR Discussion Papers 1812, C.E.P.R. Discussion Papers (March 1998)

15. Levine, D.K.: Modeling altruism and spitefulness in experiment. Review of Economic Dynamics **1**(3) (July 1998) 593–622
16. Fehr, E., Schmidt, K.M.: Theories of fairness and reciprocity - evidence and economic applications. CEPR Discussion Papers 2703, C.E.P.R. Discussion Papers (February 2001)
17. Camerer, C.: Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton (2003)
18. Fehr, E.: 15. In: Social Preferences and the Brain. Academic Press (2008)
19. Baltag, A., Moss, L.S., Solecki, S.: The logic of public announcements, common knowledge and private suspicions. Technical Report TR534, Indiana University, Bloomington (November 1999)