19
ICL

19th International
Congress of Linguists
July 21-27 2013
Geneva - Switzerland

Christopher AHERN and Jason QUINLEY

University of Pennsylvania and University of Tübingen
cahern@ling.upenn.edu, jquinley@brookstoneschool.org

*Would you trust me please? A formal model
of polite requests and relationships*

oral presentation in session: 6B Pragmatics, Discourse
and Cognition (Horn & Kecskes)

Would you trust me please? A formal model of polite requests and relationships

Christopher Ahern   Jason Quinley

Abstract

Politeness strategies are employed cross-linguistically as a means of navigating interactions and negotiating relationships. We model the use of such strategies as kind of *trust game* where relationships between individuals are expressed in terms of *other-regarding preferences*. We demonstrate how politeness functions as a kind of currency for efficient social exchange.

1   Introduction

Asking someone for the time or if they could open the door for you when your arms are full are commonplace yet revelatory occurrences. Requests in their many forms are central to social interaction, and indicate an important fact of life: we have neither unlimited information nor unlimited resources to achieve our ends. Many situations require cooperation, but cooperation is a tricky thing. Why should self-interested agents forgo their own gains to benefit others? Simply put, they should not. But, humans are a remarkably cooperative species. Alongside language, it is perhaps the hallmark of our species. We are not the perfectly self-interested *homo economicus* of classical Economic theory, but a *homo empathicus* with concerns extending beyond ourselves. Simply put, we care about others. In this paper we model how these *other-regarding preferences* allow for social exchange through requests. Further, we show that politeness acts as a kind of currency for efficient social exchange under such circumstances.

To begin, we consider our intuitions regarding the requests below as they vary by the magnitude of the request, the degree of politeness used, and the relationship between the interlocutors.

(1)   a.   Could you lend me a dollar?
            b.   I would be forever in your debt if you could lend me a dollar.

(2)   a.   Could you lend me 1,000 dollars?
            b.   I would be forever in your debt if you could lend me 1,000 dollars.

(3)   a.   Excuse me Sir, If it isn't too much trouble, would you be able to tell me the time?
            b.   Could you tell me the time?

Asking a friend to lend you a dollar does not require much ceremony, whereas a larger sum might. The relationship between the requester and the requestee also affects the form. Asking a stranger to give you a dollar might succeed, given the right words, but it is hard to imagine that any amount of locution could coax a larger sum. In contrast, there are certain requests that are assumed to be reasonable, even between strangers. In fact, being too polite in such situations arouses suspicion. For instance, asking someone the time requires a minimum of polite behavior, which may be diminished between interlocutors with closer relationships. Being too polite when asking for the time suggests an ulterior motive, perhaps including a more burdensome request. Politeness can thus be viewed as a rational behavior geared towards garnering the cooperation of another agent.[1]

In addition to our intuitions regarding the form of requests, we also have a sense for the appropriateness of responses. For example, declining to lend a large sum of money, even to a close friend, is not entirely unreasonable. Refusing to lend a single dollar might be seen to indicate a lack of trust, and call into question the foundation of a friendship. Denying either sum to a stranger does not carry with it any such weight. In contrast, one cannot refuse to tell someone the time without appearing extremely, and unnecessarily, aggressive.

(4)   a.   Could you tell me the time?
            b.   F*!k you!

We give a formal account of these intuitions regarding both the appropriate form of requests and responses.

The rest of this paper is organized as follows. We begin in Section 2 by introducing the relevant notions regarding politeness. In Section 3 we

---

[1]Here we take *rational* behavior to be behavior that maximizes success with regard to some goal. This may, or may not, involve conscious deliberation.

introduce trust games as a potential model for requests. In Section 4 we adapt trust games to model requests and demonstrate the effect of politeness: more requests can take place between individuals with a given relationship, and a given request can take place across more types of relationships. That is, friends can ask more of each other, and strangers can guarantee some amount of cooperation. In Section 5 we discuss our results and conclude.

2   Politeness

Taking Goffman's (1967) notion of *face* as a starting point, Brown and Levinson (1978, 1987) articulate a theory of politeness, which has prompted much subsequent theoretical and empirical work.[2] *Face* is the term given to an individual's basic social 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. Requests are a kind of imposition, and thus threaten the negative face of the requestee. Here we outline why and how requests are made and how they relate to the notion of face.

   The abilities or resources of a given agent might be insufficient to bring about a desired outcome. This provides an incentive for an agent in such a situation to make a request of another who is in a better position to bring about the desired outcome. Such requests impinge on the negative face of the addressee. In those cases where a request must be made, the speaker 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. At one extreme, a speaker might avoid making the FTA altogether. This might be the best option if the addressee is very unlikely to grant the request given its magnitude, or for various other reasons. If the speaker decides to go forward with the FTA, then it may be on the record or off the record. That is, the speaker may use indirect speech to implicate the FTA, but retain plausible deniability.[3] If the speaker goes on the record, then the speaker

---

[2]See the introduction of Brown and Levinson (1987) for citations, results, and a discussion.

[3]In what follows we will not deal directly with indirect speech, but see Pinker, Nowak, and Lee (2007) and Mialon and Mialon (2012) for game-theoretic treatments..
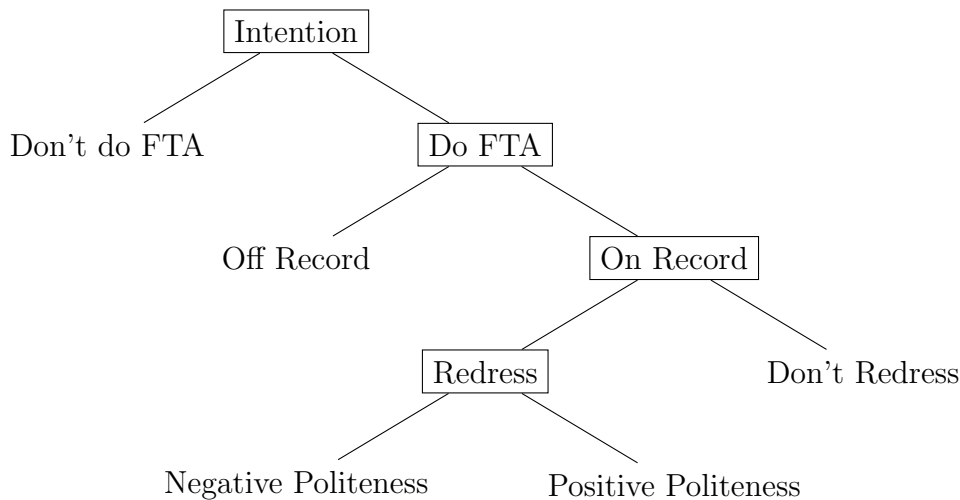
Figure 1: **Brown and Levinson's Politeness Strategies:** As we move down the graph the potential weight of a *face-threatening act* (FTA) increases.

can use flattery to address the hearer's positive face wants, or deference to address the hearer's negative face wants. Finally, the speaker may ignore the adressee's face wants entirely and make the request without any form of politeness, and make a demand.

As a concrete example, consider the situation of having left one's wallet at the office while going out to lunch with a group of friends or colleagues. Here the relevant FTA might be taken as requesting some money. The various strategies of doing so could be implemented as follows.

1. **Don't do FTA:** Don't ask for money.

2. **Off Record:** "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:** "You're such a great friend that I'm sure you wouldn't mind lending me a bit of money for lunch."

5. **Don't Redress:** "Lend me some money for lunch!"

Note that these strategies can be combined. One might begin with an off record statement and only proceed to some form of politeness if the implied request is not understood and responded to favorably. In all, we see that the goal of the speaker is to use a message that conveys the intent of the FTA, but minimizes the weight of the threat to the adressee: the greater the magnitude of the request, the more care that needs to be taken. If lunch is at an inexpensive food stand, it might not require much politeness. If lunch is at the most expensive restaurant in town, then convincing others to lend you the money might require a more indirect approach or more politeness in general.

We should note that while the use of politeness may be highly prevalent, certain situations call for less rather than more. For example, being too polite when asking for a simple favor, such as the time, raises suspicion. Given that increased politeness often indicates a larger imposition. In this case, the form of the message and the request are not congruent. Finally, when exigent circumstances require it, politeness is altogether unnecessary. "Could you move out of the way of that speeding car?" is entirely too deferential for the circumstances. In the face of such physical threats, face wants are secondary.

In these varied uses of politeness we find something eminently strategic. That is, given a certain context, we find FTAs crafted to use not too much nor too little politeness. In light of this, we turn to the use of game theory as a means for describing these strategies and revealing more about the underlying nature of social interaction.

3   Trust Games

Trust Games consist of an agent who has the option to invest some proportion of an endowment with a trustee whom they do not know and will not meet. The amount invested grows by some positive rate and the trustee then decides what proportion, if any, to return to the investor. Suppose we let e be the initial endowment and $r > 1$ be the rate of growth. Further, let us give the investor the option of investing the entire endowment or none of it. Finally, let us give the trustee the option of returning some proportion $p > 0$ of the resulting amount, or none of it. We can represent this game in extensive form as in Figure 2.

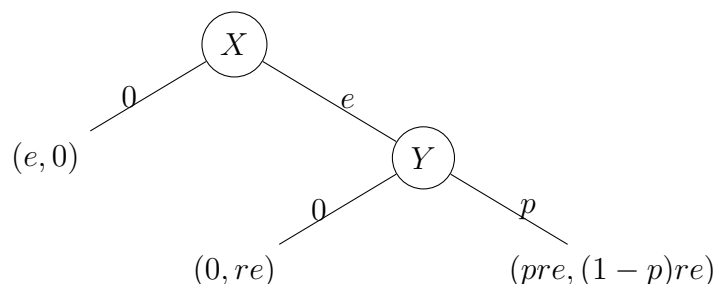It is only worthwhile for the investor to make an investment if the amount

Figure 2: **Trust Game:** The investor, X, can invest all or none of an endowment. The amount invested grows by r. The Trustee, Y, can then return some proportion of the resulting amount, or none of it.

returned is greater than the initial investment. That is, if there is some profit to be made by investing at all. If the Trustee always returned some amount, it would suffice for $pr > 1$ for this to be true. However, it is always best for the trustee to keep any amount that is invested, given that keeping all of the resulting amount is better than keeping only some of it: $re > (1 - p)re$, given that $p > 0$. Knowing that the Trustee has no incentive to return any of the amount invested, the Investor should never invest anything, because it will never yield a profit. Both could do better if some amount were invested and returned, but neither investing nor returning any amount is rational. To achieve this outcome, where both Investor and Trustee are better off, requires that the Investor trust the Trustee to cooperate. It should be noted, however, that this reasoning holds for anonymous one-time interactions. We might expect different results under different circumstances, which is exactly what we consider in the next section.

4   Requests

Trust games are an appropriate model for requests for several reasons. First, the asymmetry that prompts requests is inherent in the structure of the game. Individuals are rarely if ever entirely self-sufficient and often possess different aptitudes and abilities. Second, polite requests entail a loss of face on the part of the requester; so to speak, the requester makes a face "investment" in the requestee in the hopes of a return in the form of the request being granted.

Finally, the requestee is not obligated to grant the request, presenting the agent in need with the risk of losing the face investment. Here we extend trust games to include further actions as they pertain to requests. First, a requester X can ask a request of Y or not. Using a form of politeness entails a certain amount of investment on the part of X in addressing Y's face wants. Y may then choose to grant the request or not. If the request is granted, then X may choose to thank Y, again investing some amount in Y's face. The structure of this interaction can be seen in Figure 3.

The details of the payoffs are as follows. 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 and b_x be the benefit of that outcome. In general, we assume that $b_x < c_x$; otherwise, X could bring about the desired outcome with a net positive gain in utility. The payoff to X for a given outcome is given by the function U_x. For example, $U_x(\neg A) = b_x - c_x$. Similarly, the payoff to Y for a given outcome is given by the function U_y. In this case, $U_y(\neg A) = 0$. 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 asks (A) for help, using a polite request, Y should experience some boost in face based on the attention received, either in terms of positive or negative face. That is, by acting in accordance with Y's face wants, X increases Y's face. Let the amount of face invested by X in Y's face by using a polite 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 invested by X, $m_r f_r$. If Y denies the request ($\neg G$), X has incurred the face cost of asking without receiving any benefit, 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 an increase in face from 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 $m_t > 1$. In total, these details can be seen in Figure 3.

To give a concrete example, consider the case of asking a stranger for the time. The asymmetry in this case might stem from forgetting your watch,
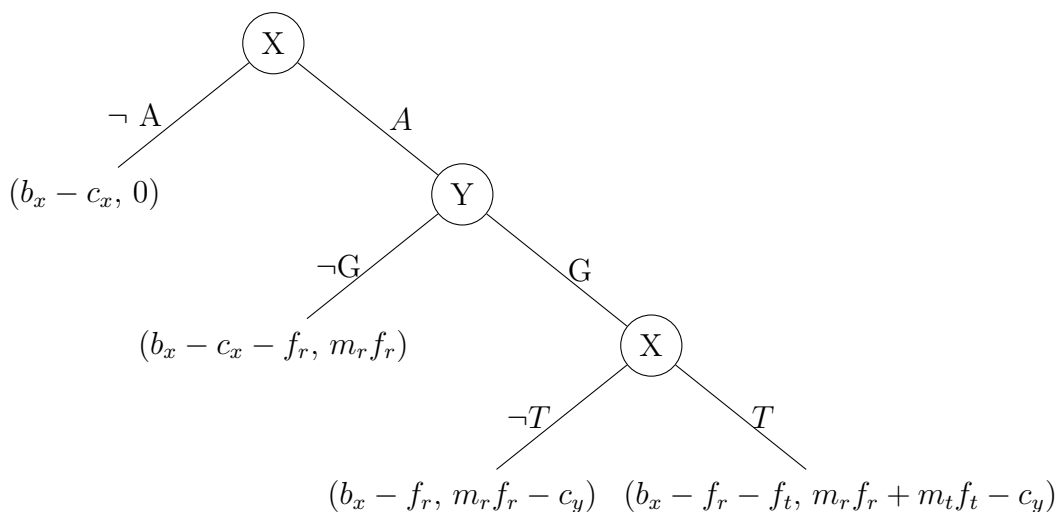
Figure 3: **Request Trust Game:** Player X can choose to Ask (A) something from Player Y, who can then choose to Grant (G) the favor. Player X can choose to Thank (T) player Y.

or having your cell phone run out of battery. In this case the cost to you of finding out the time (c_x) would involve locating a clock. When walking on the street this may prove difficult. You can choose to ask someone walking by. The cost to this person (c_y), if they are wearing a watch or have some other means of telling the time, is very low. Let us assume that you need to know the time to gauge how quickly you should be walking to get to an important meeting on time. Knowing the time will be of great benefit to you; b_x is high. What is the rational course of action in this situation?

We examine the expected result of the game in Figure 3 by determining preferences over outcomes. Faced with the choice between T and $\neg T$, X will choose $\neg T$ given that $f_t > 0$. Knowing this, faced with the choice between $\neg T$ and $\neg G$, Y will choose $\neg G$ given that $c_y > 0$. Knowing this, faced with the choice between $\neg G$ and $\neg A$, X will choose $\neg A$ given that $f_r > 0$. Counter to our intuitions, we would expected to see that instead of asking someone else for the time, one should search for a clock. This certainly does not accord with intuitions or observed behavior, yet we are faced with the same problem as before: cooperation is required for any request to be made or granted.

How are we to make sense of this result? This kind of reasoning in one-shot games is suspect on two accounts. First, even under experimental conditions of complete anonymity, subjects fail to be so calculatingly self-interested (Fehr and Schmidt, 1998; Levine, 1998; Sally, 2000, 2001; Fehr and Schmidt, 2001; Camerer, 2003; Fehr, 2008). Second, in real-life we are willing, perhaps even obliged to grant nominal requests to strangers. How can we explain these facts?

Here we adapt the notion of sympathy suggested by Sally (2000, 2001) to address these facts. 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.

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

To give an example of how this transforms the payoffs of the game, consider the new utility function for the outcome where Y does not grant the request. Remembering that $\delta_y(U_x) + \delta_y(U_y) = 1$ and $\delta_x(U_y) + \delta_x(U_x) = 1$, we can rewrite the utilities of all of the outcomes of the game with other-regarding preferences.

$$V_y(\neg G) = \delta_y(U_x)(b_x - c_x - f_r) + (1 - \delta_y(U_x))(m_r f_r) \tag{2}$$

The impact of other-regarding preferences can be seen in the following reasoning. We can determine what amount of other regard would suffice to make certain actions rational for the agents involved. We begin by considering when it is rational for X to thank Y. To do so, we determine the condition under which the sympathy distribution of X renders thanking (T) preferable to not thanking ($\neg T$). This can be determined by the following inequality.

$$V_x(\neg T) < V_x(T)$$
$$\frac{1}{1 + m_t} < \delta_x(U_y) \tag{3}$$

9

We can interpret this result in two ways. The greater the benefit to Y for thanking, the less X has to care about Y's to do so. Or, given an identical degree of other-regard for two individuals, different m_t's will render thanking rational or not. For example, one might send a thank you card or email to an acquaintance and a friend who have granted a favor, especially if the acquaintance would greatly value the show of gratitude. Similarly, if two friends grant you a favor, but you know that one would appreciate recognition, you are more likely to express thanks to that one in particular. As m_t increases, the threshold approaches 0. The more someone benefits from thanking, the less you actually have to care about them 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 the granting of requests. To do so we determine when Y prefers G to $\neg G$. This can be done based on the assumption that X will thank Y or not. We begin by assuming that X will not, and express the inequality that obtains. This can be expressed as follows.

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

We can also express the degree of other-regard that is required for Y to grant a request if X does thank Y.

$$V_y(\neg G) < V_y(T)$$
$$\frac{(c_y - m_t f_t)}{(c_y - m_t f_t) + c_x - f_t} < \delta_y(U_x) \tag{5}$$

As we would expect, if X thanks Y, then Y need not have such a high degree of regard for X's payoff. This can be seen by showing that the amount of other-regard required when X thanks Y is less than the amount when X does not.

$$\frac{(c_y - m_t f_t)}{(c_y - m_t f_t) + c_x - f_t} < \frac{c_y}{c_y + c_x}$$
$$\frac{c_y}{c_x} < m_t \tag{6}$$

10

Given that $c_y < c_x$ and $1 < m_t$, this is always true. Intuitively speaking, if Y believes that X will thank her, then Y does not have to care as much about X's payoff to grant the request. This makes sense especially if X's thanks serve as a kind of promissory note for a future return of favor.

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. If X cares enough about Y's payoff, then X will thank Y. If Y cares enough about X's payoff, then Y will grant the request. That is, if the specified 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.

Second, we can explain our behavior both with strangers and close friends with the introduction of other-regarding preferences. We suppose a minimal level of other-regard that renders granting certain requests to strangers even when no one else is observing our actions. Moreover, we are often less polite with those we are most intimate with. We can show why this is the case by considering what a request without a face-adressing form would look like. To demonstrate this point, consider a faceless Trust Game, where $f_t = f_r = 0$ for the payoffs in Figure 3. We can think of this as a system where no transfers of face are possible. The structure of the game reduces to that in Figure 4. Asking is as good or better than not asking for X, and it is strictly better than not asking if there is a non-zero probability that Y will grant the request. Given this fact, suppose that X always asks. The game then reduces even further to a choice on the part of Y to grant the request or not. We can then, as before, determine the threshold of other-regarding preferences that suffice to make granting the request rational for Y.

The corresponding threshold of other-regarding preference can be given as the following, where we let $\delta_{y'}(U_{x'})$ be Y's regard for X's payoff in a system without face.

$$V_{y'}(\neg G) < V_{y'}(G)$$
$$\frac{c_y}{c_y + c_x} < \delta_{y'}(U_{x'}) \tag{7}$$

From Eq. (5) and (7) we know that a system with face requires a lower sympathy threshold than one without face just when the following holds.
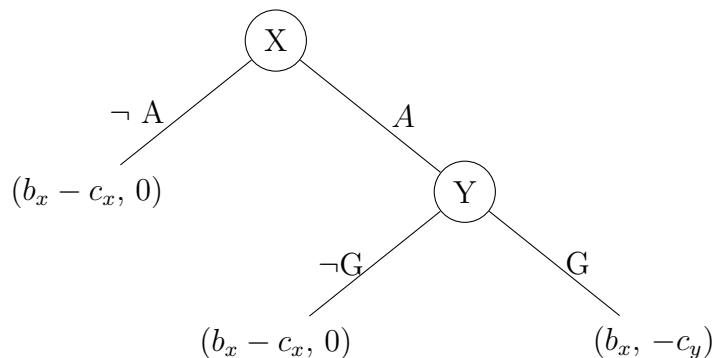
11

Figure 4: **Request Trust Game without face:** Player X can choose to Ask (A) something from Player Y, who can then choose to Grant (G) the favor.

$$\delta_y(U_x) < \delta_{y'}(U_{x'})$$
$$\frac{c_y}{c_x} < m_t \tag{8}$$

Given that $c_y < c_x$ and $1 < m_t$, it is always the case that a closer relationship is required to not use any amount of politeness to address another's face wants. This notion also extends to cases of brutal honesty between friends. For instance, a close enough friend might be willing to truthfully inform you when an outfit is unflattering in order to save you of embarrassment in front of others. An acquaintance might not presume the liberty to do so. The freedom to be honest in this way stems from the assumption of sufficiently aligned interests.

Finally, consider the role that politeness, as a means of addressing the face wants of interlocutors affects cooperation in a group of agents with other-regarding preferences. We can think about the result that a system with politeness requires lower thresholds for cooperation. There are two ways to interpret this result within a group. First, for two individuals with a given relationship, we can specify an amount of regard they have for each other. Being able to address each other's face wants allows them to ask more of each other, and increases the overall level of cooperation within a group. Second, we might consider a particular request. The use of polite forms allows for

12

that request to be made across a wider range of relationships. That is, it allows one to make a request not just of friends but of strangers as well.

5   Conclusion

We have provided a model of requests as trust games and explored the ways in which reputation, in the form of direct reputation and observation, and other-regarding preferences might enable cooperation. To address all observed behavior we used the notion of other-regarding preferences, and showed how politeness forms that address face wants encourage 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. Here we pause to reflect on the role that politeness might play in social evolution and future directions.

Not using politeness forms is better for an agent individually, but it leads to a suboptimal outcome overall given the need for cooperation in requests. However, by observing one norm, an agent might reveal her intentions to follow others. That is, by following the norm of politeness, one indicates that one will also follow the norm of reciprocity. Being aware of and using the politeness forms of a particular group suggests a set of shared obligations (Dunbar, 2004). The ability to reveal these intentions would have been particularly useful in those cases where groups within a larger population had little contact, but also shared interests. That is, as the size of societies grew beyond the size where everyone knew one another, politeness would act as a signal of intentions.

Finally, we note that other-regarding preferences can serve as a useful descriptive framework for exploring different kinds of relationships. For example, we can use the notion of other-regarding preferences in a direct manner to represent particular classes of relationships, such as communality, reciprocity, or dominance (Fiske, 1992). The case of communality holds when X and Y have a high regard for each other and both share in the outcome of the other. This suggests that asking and granting requests is in both of their best interests. For relationships based on reciprocity, we might assume a lower level of other-regard. That is, individuals in this sort of relationship might trade benefits with each other, but not be able to rely entirely upon full cooperation. In a relationship where X has dominance (or authority) over Y, Y is obliged to hold X's preferences in high regard, but not vice versa.

For example, an employer is required to use a minimal amount of politeness, whereas an employee is obliged to use quite a bit more in making requests. The notion of other-regard captures these facts in a straightforward manner.

This work follows in the vein of approaches to pragmatics and politeness from a strategic viewpoint (Benz, Jaeger, and van Rooij, 2006). It defines the conditions under which politeness strategies are rational. A central result is that a system with face allows for a greater level of trust and cooperation between agents with other-regarding preferences. The results presented here demonstrate the growing ability of game-theoretic methods to model pragmatic phenomena, including politeness. While reciprocity and coordination existed outside of and prior to language, language nonetheless serves as an efficient tool for managing them in relationships.

References

Benz, Anton, Gerhard Jaeger, and Robert van Rooij, eds. 2006. *Game Theory and Pragmatics.* Palgrave Macmillan.

Brown, Penelope, and Stephen C. Levinson. 1978. Politeness: Some universals in language use. In *Questions and Politeness*, ed. E.N. Goody, 56–310. Cambridge papers in social anthropology; no. 8, New York: Cambridge University Press.

———. 1987. *Politeness: Some universals in language use.* 2nd ed. Cambridge: Cambridge University Press.

Camerer, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton: Princeton University Press.

Dunbar, Robin. 2004. Gossip in evolutionary perspective. *Review of general psychology* 8(2):100.

Fehr, Ernst. 2008. Social Preferences and the Brain. In *Neuroeconomics: Decision Making and the Brain*, ed. Paul W. Glimcher, Ernst Fehr, Colin Camerer, and Russell Alan Poldrack, chap. 15. Academic Press.

Fehr, Ernst, and Klaus M. Schmidt. 1998. A Theory of Fairness, Competition and Cooperation. CEPR Discussion Papers 1812, C.E.P.R. Discussion Papers.

———. 2001. Theories of Fairness and Reciprocity - Evidence and Economic Applications. CEPR Discussion Papers 2703, C.E.P.R. Discussion Papers.

Fiske, A.P. 1992. The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological review* 99(4):689.

Goffman, Erving. 1967. *Interaction Ritual: Essays on Face-to-Face Behavior*. New York: Anchor Books.

Levine, David K. 1998. Modeling Altruism and Spitefulness in Experiment. *Review of Economic Dynamics* 1(3):593–622.

Mialon, H.M., and S.H. Mialon. 2012. Go Figure: The Strategy of Nonliteral Speech. Unpublished Manuscript, Emory University.

Pinker, Steven, Martin A. Nowak, and James J. Lee. 2007. The logic of indirect speech. *Proceedings of the National Academy of Sciences* 105: 833–838.

Sally, D. 2001. On sympathy and games. *Journal of Economic Behavior & Organization* 44(1):1–30.

Sally, David. 2000. A General Theory of Sympathy, Mind-Reading, and Social Inter- action, with an Application to the Prisoners' Dilemma. *Social Science Information* 39(4):567–634.