Beyond the politeness tightrope: Message design for multiple social goals

David Pautler (pautlerd@ihpc.a-star.edu.sg)
Institute for High Performance Computing, 1 Fusionopolis Way, 16-16 Connexis
138632 Singapore

Abstract

In linguistic pragmatics and social anthropology, several influential researchers believe that politeness is essential for maintaining social order by way of disarming potential aggressiveness [Goffman 1967; Brown & Levinson 1987; Gumpers 1987]. In one of the most detailed of these theories, Brown and Levinson's, speakers pursue a single goal (e.g. getting the hearer to stop doing something) by using a mental model of the hearer to select a position on a one-dimensional spectrum of strategies that identifies the best balance between achieving the speaker's practical goal while avoiding offense to the hearer (as might occur from a purely brusque request). Brown and Levinson's theory can account for many observed conversational sequences, and its use of means-end and abductive reasoning make it a good candidate for a computable cognitive model of socially-appropriate action selection [Green 1994; Bickmore 2003; Gupta 2009]. But are speakers actually limited to this one-dimensional spectrum of strategies? And given that people often pursue more than one goal at once, how might they do so in polite communication? Empirical findings by Kellermann and colleagues argue against a one-dimensional conception [2004]. In this paper, I describe and evaluate a computational model that uses Kellermann's empirical data to generate strategies for multiple simultaneous speaker goals.

Keywords: Politeness theory; Message design; Pragmatics; Computational linguistics.

Politeness Theory

When one person tries to persuade another, what is the general character of the reasoning the speaker uses to craft his message? There are several competing theories across the fields of communication studies, pragmatics, social psychology, and AI (see Pautler [2007: pp. 12-49] for an overview.) The dominant theory in the fields of communication studies and computational linguistics is Brown and Levinson's Politeness Theory [1987] (henceforth, "PT"). Its main thesis is that in conversation, each participant must weigh a tradeoff between speaking efficiently and speaking in a polite manner. Speaking efficiently means adhering to Gricean maxims, which identify four primary dimensions of efficiency: informativeness, truthfulness, relevance, and perspicacity [Grice 1989: pp. 26-27]. Speaking in a polite way means attending to the positive and negative faces of one’s audience and oneself, where "positive face" means the image one wants to project, and "negative face," the set of obligations one wants others to observe. PT claims that speaking politely requires considering each of one’s possible paths forward in the conversation for its face-threatening potential, where the degree of threat is determined by cultural factors, the relative power between the participants, and their social distance. PT claims that speakers who are concerned with being polite judge the degree to which their acts may be face-threatening before carrying them out, and that they use a set of strategies for averting escalating degrees of threat. These strategies are:

1. If the threat is low, simply do the act.(This is the same as following Gricean maxims.)
2. If the threat is minimal, perform the act but indicate to the recipient that no harm is intended, or promote his positive face in another way (e.g., compliment him).
3. If the threat is moderate, perform the act but find a way to promote the recipient's negative face (i.e., his independence) (e.g., offer to help him).
4. At higher levels of threat, the act should be communicated indirectly or abandoned.

This ranking of strategies suggests a one-dimensional spectrum ranging from 'maximal efficiency' (for extant goal(s) and situation) at one end and 'maximal politeness' (for the same conditions) at the other, in which each of these strategies occupies a non-overlapping section of that spectrum. Most computational models of PT are designed to estimate the minimum level of politeness acceptable to the hearer (using input representations of "social distance" and other scalar values) — a position on the spectrum considered optimal for persuasion for the current situation — and then search for a set of utterances that satisfies that balance of politeness and efficiency. For me, this conception invokes an image of a fencer planning the spot he should leap to in order to have maximum advantage in sparring with his partner, but unlike in street-fighting, he is limited by the rules of the sport to moving only within a narrow strip, a tightrope demarcated by convention rather than worldly constraints. How strong is the evidence in support of this conception of interpersonal persuasion?

Conversational Constraint Theory

Kellermann and her students [2006] formulated an alternate account, Conversational Constraint Theory (henceforth, "CCT"). The aim of CCT is to describe a decision process whereby communicative acts such as hinting, changing body pose, and making eye contact are weighed for their suitability for an overarching interpersonal goal and for the situation so that the goal is met in a socially acceptable way. CCT focuses on "compliance-gaining" interactions, where one person tries to persuade another to adopt a particular goal. According to CCT, every potential act can be weighed in terms of its likely efficiency and its

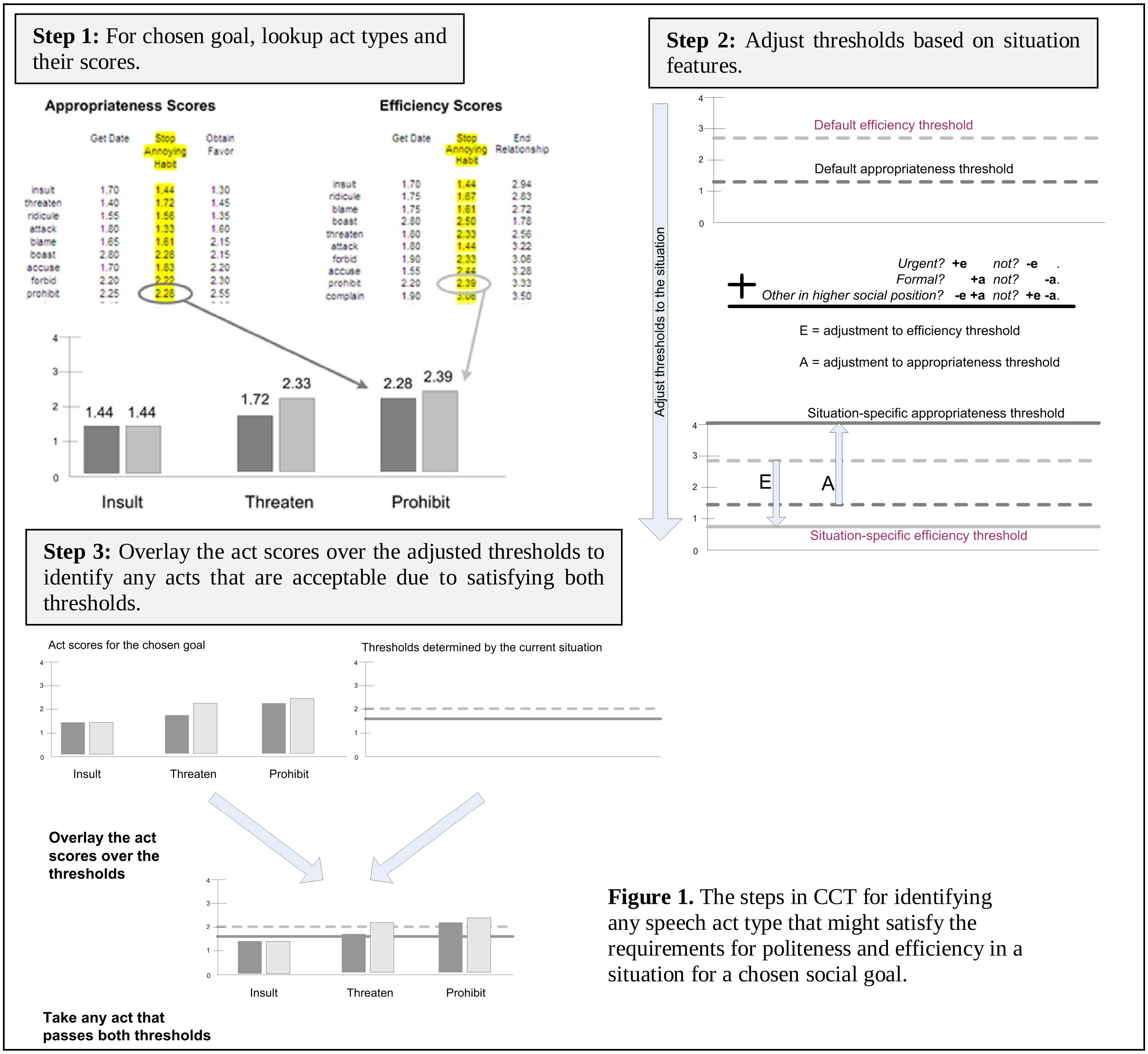

Figure 1. The steps in CCT for identifying any speech act type that might satisfy the requirements for politeness and efficiency in a situation for a chosen social goal.

likely politeness for the goal and situation. To decide whether an act is worth using, one would compare its efficiency score against a minimum preferred level of efficiency, while comparing its politeness score against a minimum preferred level of politeness; only acts that score above both levels are acceptable. That is, CCT shares a minimum preferred level of politeness with PT but adds a minimum preferred level of efficiency. This addition allows for the possibility that these two variables, when envisioned as dimensions of a 2D surface in which the Z-axis indicates overall acceptability, might have local maxima and not be in an exact trade-off; the empirical studies by Kellermann and colleagues found that for many combinations of goals and situation, that that is actually the case [Kellermann 2004].

Consider the interpersonal goal of stopping an annoying habit in the hearer. In PT, such a goal would be considered moderately threatening (or worse) to the hearer's "face", suggesting that the speaker should "promote the hearer's negative face" (i.e., PT strategy 3 above; perhaps by acknowledging his option to decline to comply) or should be indirect (i.e., PT strategy 4 above). Both of those options sacrifice efficiency (i.e., directness) for politeness. Yet, Kellermann found for this same goal of stopping an annoying habit that efficiency does not have to be sacrificed for politeness, even when using speech act types that are quite face-threatening such as insulting, threatening, and prohibiting (see Figure 1, Step 1). In fact, across these three act types, politeness and efficiency scores *increase together* for this goal.

The 2004 CCT study cited above was intended to evaluate a hypothesis that compliance-gaining goals (e.g., obtain a favor) vary in the degree to which they constrain the politeness and efficiency of compliance-gaining behaviors (e.g., requesting versus hinting). The data was created by

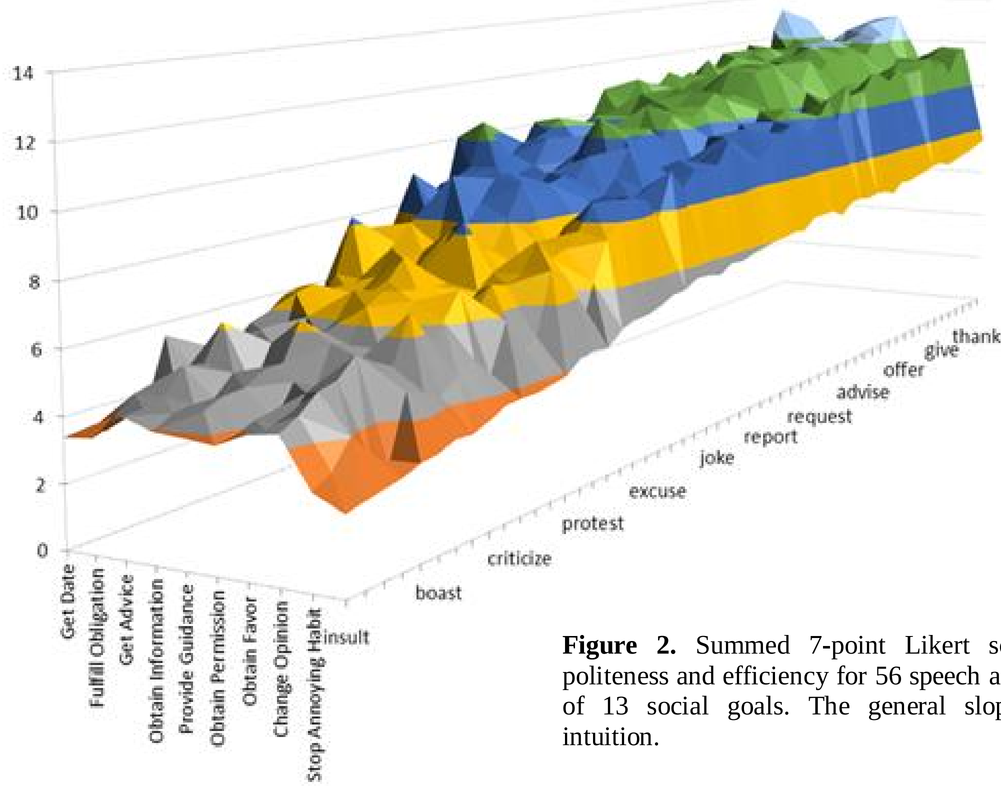

Figure 2. Summed 7-point Likert scale ratings on politeness and efficiency for 56 speech act types for each of 13 social goals. The general slope conforms to intuition.

asking a large pool of undergraduates to rate each act's efficiency and politeness on a 7-point Likert scale. Kellermann explained her normalization method to me by email this way:

> The tactical evaluations were examined for their distribution to make sure they "converged" on the same point. That is, I looked at the distribution of the ratings of each tactic, and checked to see if the scores were normally distributed, or the distribution was "taller" than normal (i.e., the scores converged on the mean with less than expected standard deviation). I found, for the most part, convergence of the ratings onto the mean, or normal distributions. Ratings converged. Z-scores are used to make this assessment. Because all tactics were judged on the same 7 point scale at the same time by participants, standardization is not needed to compare the tactics further, so no additional z-score coding across tactics was undertaken.

To my knowledge, and that of Kellermann [personal communication], no replication of the study has been attempted. Pautler [2007: p. 53] describes an "eyeball comparison" with a similar, earlier study, but the comparison was inconclusive. Yet, if one sums the politeness and efficiency score for each combination of goal and speech act type in the 2004 CCT study, and sorts those sums (as done in Figure 2), the general slope of the sums conforms to intuition. For example, the lowest sum occurs when using an insult to stop an annoying habit in the hearer, while giving thanks to the hearer has high sums across several goals. I take these considerations as sufficient to merit using Kellermann's data in an exploratory computational model that takes as input a set of goals and a representation of the extant situation, and generates a set of acts that satisfy constraints of politeness and efficiency for those inputs.

Computational Model Design

Although the aim of CCT is to describe a decision process for selecting a communication act for a goal, it does not spell out an algorithm for doing so; yet, one can be inferred from its discussion of scores, situation factors, and thresholds. This process is illustrated in Steps 1-3 of Figure 1. As previously discussed, Step 1 shows how the chosen goal results in a set of candidate acts. Step 2 shows how situation factors determine thresholds on the scores of candidate acts. CCT does not suggest whether these two subprocesses are done in parallel nor in a specific order, but neither depends on the other so it does not seem to matter. For the subprocess of determining thresholds, one checks each of the situation factors and adjusts one or both thresholds by a delta value; for example, if the person one is speaking to is in a higher social position, one would decrement the efficiency threshold and increment the appropriateness threshold, while if they were not in a higher position, one would do the opposite and increment efficiency and decrement appropriateness. CCT claims that threshold settings vary over a variety of situational, relational, personal, and interaction-specific factors (see Pautler [2007: pp. 107-109] for the list, plus definitions).

Step 3 shows how the thresholds from Step 2 are overlaid on the candidate acts and their scores from Step 1 to reveal the subset of candidates having scores passing both thresholds; this subset represents all the acts considered acceptable for use in the situation for the chosen goal.

CCT does not describe how multiple acts might be chosen as a coordinated plan for a goal, nor does it describe how multiple goals might be pursued simultaneously; hence, it is implicitly limited to single-goal, single-act scenarios. Yet, CCT's argument against the spectrum of strategies conception is intriguing because it implies that, in some circumstances, a speaker can pursue greater politeness and efficiency *simultaneously* toward a local maximum. Intuitively, this means that attempts at persuasion might succeed not just because they find the right balance in a tradeoff of politeness and efficiency, but also because the attempt found a "sweet spot" that optimized both. Furthermore, if such sweet spots are possible when using a single speech act type for a single goal (as conceived in CCT), it seems conceivable that even greater local maxima might exist if one pursues multiple goals simultaneously through a coordinated set of speech acts. In everyday conversations, it seems commonplace for the interactants to be pursing multiple goals. Therefore, if one extended CCT to work for multiple simultaneous goals so that it could search for local maxima of politeness and efficiency through a bundle of speech acts that might serve as a conversational turn, it might be a good predictor of human judgments of the acceptability of that bundle. I am not aware of any other model in cognitive science that claims to predict the same, and such a model has potential practical benefits as well as theoretic ones [Pautler 2007: pp. 1-10]. The remainder of this paper describes such a computational model, and an evaluation.

Planning Multiple Acts for Multiple Goals

The procedure suggested by CCT for finding a single speech act type to satisfy a single goal in a certain situation can be conceptualized as solving two mathematical inequalities. One of the inequalities is for politeness; it has a minimum preferred level of politeness, Min_{pol}, on one side, and a sum of terms of the form, $Act(i) * Pol_{Act(i)}$, on the other, in which $Act(i)$ is 1 or 0 indicating whether the *i*th act type is chosen, and $Pol_{Act(i)}$ is the politeness score of that act type. The other inequality has the same structure but represents efficiency values instead of politeness ones:

$$Min_{pol} =< Act(1)*Pol_{Act(i)} + \ldots + Act(n)*Pol_{Act(n)}$$

$$Min_{eff} =< Act(1)*Eff_{Act(i)} + \ldots + Act(n)*Eff_{Act(n)}$$

If one can solve this pair of inequalities by giving exactly one of the $Act(i)$ variables a value of 1, then the speech act type identified by *i* will satisfy the input goal for the input situation (which determined the *Min* threshold values). Depending on the threshold values and act scores, there might be several solutions or none.

The requirement that any solution must use at most one act type seems somewhat arbitrary; it might be the case that the scores of two (or more) act types together could pass both thresholds but the scores of no single act type can. The almost trivial example of the paragraph (i.e., a group of coordinated clauses often of different act types) indicates that everyday communication is rife with examples of pursing at least one communicative goal through multiple coordinated speech acts. Dropping the single-act requirement allows finding more solutions, and seems to reflect everyday practice.

Once we drop the requirement, suddenly facing many potential solutions to a single inequality pair, how do we select among them? Grice's maxim of quantity (i.e., brevity) pushes us toward minimizing the number of act types in our preferred solution. With this minimization constraint on the number of act variables that should be set to 1 rather than 0, this formalization of the problem becomes an instance of *integer linear programming* (ILP). For example, to pursue the goal GetDate using just the speech act types {argue, assert, insist, offer, vow}, the corresponding ILP problem can be encoded this way[1]:

```
Minimize
  num_acts: vow + insist + assert + offer + argue

Such that
 Get_Date.effMin: -0.3 vow - 1.15 insist + 0.85 assert +
1.3 offer - 1.65 argue >= 0.7
 Get_Date.appMin: 0.45 vow - 1.2 insist + 1.1 assert +
1.6 offer - 1.55 argue >= 1.2

Binaries
  vow
  insist
  assert
  offer
  argue
```

So now we have a way of solving for a single goal using potentially several acts used in combination. Yet, as discussed earlier, many instances of everyday communication seem to pursue *multiple* goals simultaneously through multiple coordinated acts. Can this scheme be augmented to model such a decision process? Yes, ILP accommodates this quite easily. So far, we have used a pair of inequalities for a single goal; to solve for multiple goals, we merely have to add a pair of inequalities

[1] Note that coefficients here have been normalized to [-3,3] from [1,7] in the original 7-point Likert ratings.

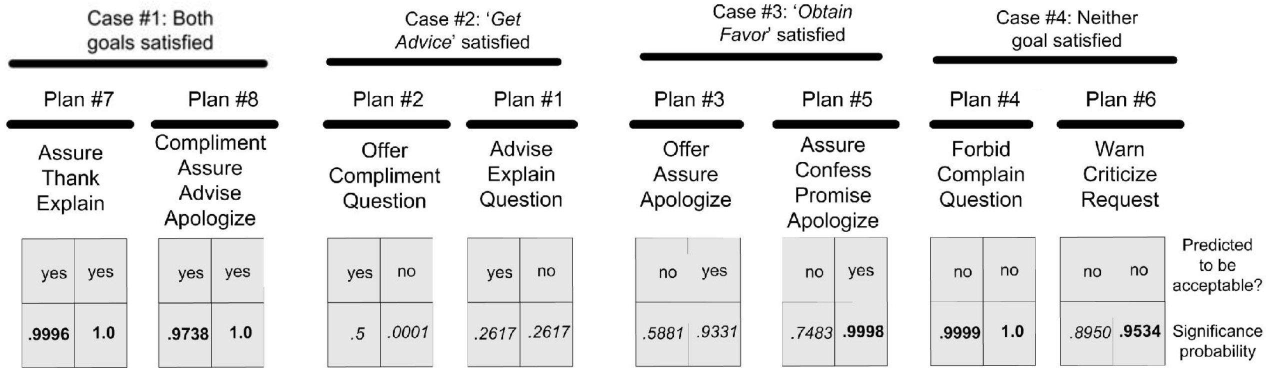

Figure 3. The pilot and study each explored the 4 possible cases of satisfying (or failing to) 2 simultaneous goals (*Get Advice* on the left of each quartile; *Obtain Favor* on the right). These are study results; pilot results were also mixed but only on the same single-goal cases as these.

for each new goal to our system of inequalities. This makes intuitive sense – if two goals are similar, their profile of scores across the act types should also be similar, with the result that solving both pairs of inequalities simultaneously should be only somewhat more constraining on the set of solutions than solving either pair in isolation.

I have named this system design, "Computable Social Communication (CSC)." CSC solves its ILP problems using the GNU Linear Programming Kit, "GLPK" [Free Software Foundation 2006]. Although ILP is NP-hard, and we are solving systems with 56 variables, it typically takes less than 100 milliseconds on a standard consumer computer to find a solution. ILP problems with only a few variables can take hours using the same hardware, but by defining our minimizing function as addition with all coefficients as +1, GLPK is able to use a special "branch-and-bound" algorithm and be extremely efficient.

Algorithm Overview

Searching via ILP for a minimal set of acts that exceeds both thresholds is the core of the CSC algorithm. But as a first step, CSC must set the thresholds, which it does by asking the user to give ratings for a series of questions about the situation, such as its level of formality. The questions are adapted from CCT's set of principles about how thresholds are adjusted based on the situation; the actual threshold default values, and the delta value applied for each matching principle, use assumed values (all are 0.5)[2].

After adjusting the minimal preferred efficiency and politeness thresholds based on the situation, and finding a set of speech act types that satisfies those thresholds and a set of target goals, the acts still may not be good choices. As established by J.L. Austin [1975] and John Searle [1969], each act type has its own preconditions for appropriate use. For example, bona fide *promising* requires that the speaker and hearer mutually believe that the speaker could fulfill the promise, and that that would be desirable to the hearer. To enforce these preconditions, CSC iterates through the candidate solution sets of act type; for each candidate set, it gathers all the associated preconditions (using formalizations based on Wierzbicka's [1987] speech acts dictionary), renders them as queries to the user, and requests a new candidate set if the user indicates that any of the preconditions fail (and blocks use of that act type in any further ILP searches for this scenario). For example, if in the GetDate ILP problem above, one candidate solution is {*insist*, *assert*} but the user indicates one of the preconditions of *insist* fails, then CSC uses ILP to search for another candidate set that does not involve *insist*. The iteration continues until the preconditions of all candidate acts are satisfied, or no more ILP solutions can be found.

[2] No relevant empirical data for setting these values is known to Kellermann [personal communication], and I could find none.

Evaluation

The claimed contribution is: *CSC can generate plans for multiple goals using multiple acts in a way that is sensitive to a wide variety of specific situation conditions.* To evaluate the claim, we designed a survey to gather human judgments on a set of plans generated by CSC. The generated plans all address the same pair of goals and the same situation conditions, but the plans differ by whether CSC expects each plan to be acceptable for both goals, for only one goal, or for neither goal. Generating plans expected to fail is not useful by itself, of course, but is intended to demonstrate that CSC's predictions are reliable across the spectrum of plan quality, not just at the extreme of plans widely agreed to be good ones. Similarly, in the survey, participants also indicate for each goal whether each plan would be acceptable (i.e., for one specific goal, for both goals, or for neither). By comparing simulator and participant judgments at this level, we can not only determine the overall accuracy of the simulator, but we also have enough data to identify possible sources of weakness in the simulator if it is found to be inaccurate in some or all cases.

A pilot and a study were done, and Figure 3 shows the results of the latter. Eight of the sixteen predictions received strong support, and two were weakly supported, similar to the pilot results where nine were strongly supported and one, weakly (see Pautler [2007, pp. 74-90]). And the cases of strong disagreement or weak support all occurred among the predictions having to do with plans expected to satisfy only one of the goals, just as in the pilot. That is, in both the pilot and the study, the model's ability to generate plans satisfying multiple goals through multiple acts received strong support. By tweaking threshold settings, all three of the strong disagreements with predictions (i.e., the acceptability of Plan 1 for *Get Advice*, and the acceptability of Plans 1 and 2 for *Obtain Favor*) can be nudged into weak support or better. The pilot and survey both revealed a supportive trend for the majority of their predictions (for twelve predictions and eleven predictions, respectively, out of sixteen). In particular, plans that were generated to satisfy multiple goals using multiple acts received strong support in all cases but one, where there was moderately strong support. These plans represent the claimed contribution that our tests were meant to verify, so the CSC model has been shown to be well-supported empirically.

Conclusion

Politeness Theory, the dominant theory of interpersonal message design, accounts for many observed conversational exchanges but still seems more limited in scope than everyday conversation. In particular, it claims that efforts at persuasion are limited to a one-dimensional spectrum of 4 standard strategy types. On the other hand, Conversational Constraint Theory claims that strategies can be constructed in an ad hoc manner resulting in greater diversity. This paper described how CCT's single-goal single-act method can be extended to multiple-goal multiple act solutions tractably in the CSC system. A pilot and study provide good empirical support for the predictions that CSC makes.

References

Austin, J. L. (1975). *How to do things with words*. Cambridge, MA: Harvard University Press.

Bickmore, T. (2003). *Relational agents: Effecting change through human-computer relationships*. PhD dissertation, MIT.

Brown, P. and S. C. Levinson. (1987). *Politeness: Some universals in language usage (Studies in interactional sociolinguistics).* Cambridge University Press.

Free Software Foundation. (2006). GNU Linear Programming Kit. http://www.gnu.org/software/glpk/glpk.html

Goffman, E. (1967). *Interaction ritual*. Garden City, NY: Doubleday Anchor.

Green, N. (1994). *A computational model for generating and interpreting indirect answers*. Ph.D. dissertation, Department of Computer and Information Sciences, University of Delaware. Technical Report 95-05.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Gumpers, J.J. (1987). Forward to *Politeness: Some universals in language usage (Studies in interactional sociolinguistics).* Cambridge University Press.

Gupta, S. (2009). *Generating politeness for conversational systems aimed to teach English as a second language*. PhD dissertation, University of Sheffield.

Kellermann, K. 2006. *Theory of selection* (chapter). Unpublished manuscript.

Kellermann, K. 2004. A goal-directed approach to gaining compliance: Relating differences among goals to differences in behavior. *Communication Research* (31) 4. 397-445.

Pautler, D. (2007). *Computable social communication*. PhD dissertation, University of Hawaii-Manoa.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge University Press.

Wierzbicka, A. (1987). *English speech act verbs*. Academic Press.