Pragmatics and Social Interaction of Conditional Influencing Promises and Threats

Cristiano Castelfranchi[1], Marco Guerini[2]

1- National Research Council - ISTC- Institute of Cognitive Sciences and Technologies
via San Martino della Battaglia 44, 00185 – Roma, ITALY
{c.castelfranchi@istc.cnr.it}
2- ITC- irst, Istituto per la Ricerca Scientifica e Tecnologica
Via Sommarie 18, 38050 - Trento, ITALY
{guerini@itc.it}

Abstract

In this paper we investigate some aspects of the process that underlie the generation of Promises and Threats (P/T) utterances. Starting from the distinction between the deep and surface form of Conditional Influencing P/T (CIP/T) - those P/T used to persuade the addressee - we focus on some possible factors that influence the choice between different deep and surface realizations. We individuate four main classes of these factors: objective, personality/emotional, relational/social and pragmatic/linguistic.

1 Introduction

This paper is based on the theory about P/T as exposed in [Guerini & Castelfranchi, 06]. Here, we expand this theory by exploring some of the factors that influence the final realization of the P/T utterance. In particular we focus on CIP/T, those P/T used to persuade the addressee. In the context of negotiating agents some simplified formalizations of CIP/T has been put forward, see for example [Kraus *et al.,* 98; Amgoud & Prade, 04; Sierra *et al.*, 98]. Still, here we will focus on the process of CIP/T utterance selection and not on P/T use in negotiation.

In the following variables *x*, *y*, *ax*, *ay* indicate respectively: the sender of the message, the receiver, an action that *x* can perform, and an action that *y* can perform. The reference model is Cohen-Levesque's one [Cohen & Levesque, 87; Cohen & Levesque, 90].

We propose two different formalizations of "goal of persuading" (formulae 1 and 2). Formula 1 implies formula 2 when *y* is an autonomous agent (i.e. every action performed by an agent follows from an intention).

1. PERSUADE(x y ay) = GOAL(x DOES(y ay))

2. PERSUADE(x y ay) = GOAL(x INTEND(y ay))

Considering formula 2, in persuasion the speaker presupposes that the receiver is not already performing or planning the required action *ay*. In a more strict definition it can also be presupposed that the receiver has some barriers against *ay*: *y* wouldn't spontaneously intend to do so. Persuasion is then concerned with finding means to overcome these barriers by conveying the appropriate beliefs to *y*.

The relation between persuasion and dissuasion is non-trivial, though, here we will consider the goal of dissuasion as the goal of persuasion to not perform a given action.

3. DISSUADE(x y ay) = PERSUADE(x y ¬ay)

The utterance structure for CIP/T is:
"*If ay then ax*"
In CIP/T structure, the condition of the utterance ("if *ay*") is equal to the achievement or avoidance goal of the persuasive act and can be based on the elicitation or activation of a non-active goal of *y*.

- In P the condition expresses what *y* has to 'adopt'. *x* is proposing an 'exchange' of reciprocal 'adoption': "if you adopt my goal (*ay*) I will adopt your goal (*ax*)".
- In T the condition is what *x* wants to avoid and he is prospecting a 'reciprocation' of damages: "if you do what I dislike (*ay*), I will harm you (*ax*)".

In CIP/T class the concepts of promise and threat are - logically - two faces of the same coin, but this is not the case for the general class of P/T. In fact the intended meaning of these social acts is an IFF: the promise act is always and necessarily accompanied by a threat act (although hiddenly): "*if you do your homework I will bring you to the cinema*" entails "*if you do not do your homework I will not bring you to the cinema*".

However we are not claiming that the two CIP/T are the same act, but that they are two necessary and complementary part of the same communicative plan; otherwise the act would be ineffective. The problem is to understand which are the factors that bring persuader to choose one of the two forms over the other (that is: the promise side or the threat side of the CIP/T). In fact the impact on the receiver can vary a lot depending on the chosen surface realization [Verbrugge *et al.*, 04; Verbrugge *et al.*, 05].

2 Deep and surface CIP/T & their grammar

Only a pragmatic difference seems to distinguish between P and T as two faces of the same act. However, common sense and language have the intuition of something deeper. What is missed is an additional dimension, where promises refer to real gains, while threats refer to losses and aggression. We need to divide CIP/T along two orthogonal dimensions: the deep and surface one.

1. The deep (substantial) dimension regards the "gain" and "losses" for the receiver related to speaker's action. Gain: the fact that one realizes a goal that he does not already have, passing from the state of *Goal p & not p*, to the state that *Goal p & p* (the realization of an 'achievement' goal in Cohen-Levesque's terminology); in this case the welfare of the agent is increased. Losses: the fact that one already has *p* and has the goal to continue to have *p* ('maintenance' goals in Cohen-Levesque terminology); in case of losses one passes from having *p* - as desired - to no longer having *p*; in this case the welfare of the agent is decreased.
2. The surface dimension regards the linguistic form of the CIP/T: the use of the P or T face.

In Table 1 on the columns we have losses and gains (with regard to *ax* in *y*'s perspective). These two columns represent:

- deep threatening (loss): a choice between two losses ("harm or costs?" no gain), let us call it *Deep-T*
- deep promises (gain): a choice between a gain (greater then the cost) or a missed gain, *Deep-P*

On the rows we have the surface form of the corresponding communicative acts: in the case of surface promise (let us call it *P-form*) what is promised is a missing loss or a gain, while in the case of surface threat (let us call it *T-form*) what is promised is a loss or a missing gain. The distinction (for a same deep structure) is granted by the IFF form of CIP/T.

	Deep-T: Loss	**Deep-P: Gain**
Surface Promise (P-form)	If *ay* then *not-loss* "If you do the homework I will not spank you"	If *ay* then *gain* "If you do the homework I will bring you to the cinema"
Surface Threat (T-form)	If *not-ay* then *loss* "If you do not do the homework I will spank you"	If *not-ay* then *not-gain* "If you do not do the homework I will not bring you to the cinema"

Table 1 – Deep and surface form of CIP/T

If we consider CIP/T as social acts based on the prospect of incentives (the promised or threatened action *ax*) then gains are due to *prizes* and losses to *punishments*. In particular:

a) If *ax* is something given because it is wanted by *y*, then it is a prize:

$$4. \text{ GOAL(y ax)} \rightarrow \text{PRIZE(ax)}$$

b) If *ax* is something given because it is not wanted by *y*, then it is a punishment:

$$5. \text{ GOAL(y ¬ax)} \rightarrow \text{PUNISHMENT(ax)}$$

What is explained in Table 1 is the general framework, but, for example we must distinguish "defensive" promises/threats (defensive from *x*'s perspective: *x* does not want *ay* and uses *ax* to stop *y*) from "requesting" ones (in which *ay* is something wanted by *x*)[1].

In particular in *requesting* P/T *ay* is something wanted by *x* (presupposing a persuasive aim on *ay*, formula 2) in *defensive P/T ay* is something not wanted by *x* (presupposing a dissuasive aim on *ay*, formula 3):

$$6. \text{ GOAL(x ay)} \rightarrow \text{REQUESTING-P/T}$$

$$7. \text{ GOAL(x ¬ay)} \rightarrow \text{DEFENSIVE-P/T}$$

3 Promising or Threatening? Pragmatics and Social Interaction

The question we address is: on which basis *x* decides between a Deep-P and a Deep-T and between the corresponding surface realizations (P-form vs. T-form)? Why and when to make explicit and apparent one facet is better than the other way around? Which are the pragmatic, social and cognitive differences?

There is for sure a criterion of non redundancy in avoiding the explicit complete expression of a CIP/T. In fact saying "If you do *ay* I will do *ax*, but if you do not do *ay* I will not do *ax*" is redundant. The second portion is left implicit but part of the intended meaning. So the criterion of redundancy says nothing about explicitly expressing the promise side or the threat side. Other criteria are needed; several factors play a role at the personality, relational, social, and pragmatic-linguistic layers. These factors are mixed together when deciding for a deep P/T and the corresponding *P-form* or *T-form*. We individuate four classes of them:

1. *Objective* factors (influencing only the deep dimension): e.g. *x* cannot compromise any of *y*'s goals
2. *Personality and Emotional* factors (influencing both the deep and surface dimensions): the personality and emotional traits both of *x* and *y* can change the effect of a CIP/T
3. *Relational and social* factors (influencing both the deep and surface dimensions): e.g. if *x* does not threat *y* he will loose his social image. If *x* promises something to *y*, *x* can gain his favors
4. *Pragmatical-linguistic* factors (influencing both the deep and surface dimensions): e.g. the premise and presuppositions of the uttered CIP/T.

---

[1] In our previous work [Guerini & Castelfranchi, 06] we called this class "aggressive", but we found it misleading because it could evoke a threatening nature.

The **objective factors** depend quite a lot on the kind of *incentive power* of *x* [Castelfranchi, 03]. There are two cases:

- *Positive incentive active power (PosPower): x* is able and in condition, through her action *ax*, to produce additional positive gains, incentives for *y*. Thus the P-face is to do *ax* and produce the prize; the T-face is just abstaining from *ax* and not producing the positive incentive. **T-form = not-prize** (no promised prize).
- *Negative incentives active power (NegPower): x* is able and in condition through her action *ax* to produce harms, damages for *y*. Thus the P-face is just abstaining from *ax* and not producing the 'penalty', while the T-face is doing *ax* and producing the harm. **P-form = not-punishment** (no giving the threatened penalty).

If *x* has negative power the T-form should be more direct, clear and natural; while having the positive power the P-form should prevail. In the following formulae the sign ">" represent an ordered relation of preference for *x* between possible alternatives:

8. NegPower: T-form > P-form (= not-punishment)

9. PosPower: P-form > T-form (= not-prize) [2]

Some factors at the **personality and emotional** level are:

y's perceived personality: is *y* more sensible and prone to influence by the prospect of losses and harms or more sensible to possible gains?

x's perceived personality: is *x* for *y* more credible as a good willing guy, as providing prizes (choosing *PosPower* option), or as rigorous or hostile guy providing punishments (choosing *NegPower* option)?

Framing effect: A negative/harm framing of the situation induces different reactions and even different reasoning and decision than a positive framing [Kahneman *et al.*, 82]. *x* can plan to exploit frame-effect to induce a response that will be different in the two perspectives.

Emotional mediation: if we have an emotional-persuasion use of CIP/T [3] then our prediction is that the P-form is aimed at eliciting attraction while the T-form bets on eliciting fear. The affective response is due to the considered and focused face of the scenario.

Some of the factors at the **relational and social** level are:

x's presented social image: the T-form (and a Deep-T) provides a bad face of *x*, to *y* and/or to others. Is *x* interested in presenting such an image of herself? If he likes a nice-gentle

guy image she will avoid the T-form, and present himself as a promising person (using P-form and/or a Deep-P).

y's relational response: will *y* be more prone to adhere to the request of a gentle guy or to a hungry one? Is he in an opposition-rebel attitude towards *x*? Or on the contrary he refuses paternalism from *x*? If *x* mainly relies on *y*'s fear, he will prefer the form of threat, focusing *y*'s attention of the possible danger and exhibiting an "aggressive" face (to make more believable his willingness and disposition to damage). If *x* mainly relies on exchange, goodwill, positive reciprocation, sympathy, gratitude, etc. he will bet "benevolent" face on the promise form.

These ideas can be summed up in the following formulae (10 and 13):

10. BENEVOLENT-FACE: Deep-P > Deep-T = If ay then prize > If not ay then punishment

11. BENEVOLENT-FACE: P-form > T-form

12. AGGRESSIVE-FACE: Deep-T > Deep-P = If ay then punishment > If not ay then prize

13. AGGRESSIVE-FACE: T-form > P-form

Linguistic and Pragmatic factors: as we saw threats can be more presuppositionally natural in dissuasion, while promises can be more spontaneous in persuasion. So this is an additional reason for presenting one face or the other.

Premise and presupposition: the chosen form may depend also on the presupposition of the sentence and in particular of possible previous linguistic expression of it.

- Suppose for example that *y* said: "*I leave; I intend to go*". In this context (let's call it: *Affirmative presupposition, AffPP*) a sentence presupposition that recovers, reproduces it seems a bit preferable; i.e. the form "*If you go…*" is better than "*If you do not go…*".
- Suppose now that *y* said: "*I do not leave; I do not intend to go*" (*Negation presupposition, NegPP*) the form "*If you do not go…*" sounds better than "*If you go…*".[4]

In other terms recovering the original expression seems to be better than reversing it:

14. AffPP: affirmative expression > negative expression

15. NegPP: negative expression > affirmative expression

It is important to note that presuppositions are useful not only to select P/T form, but also to establish the class of a P/T act: When *x* promises to *y* that *ax* if *ay*, he assumes (presupposes) – as we argued - that *y* is not already and certainly intending to do *ay*. Otherwise his promising is pragmatically incorrect and redundant (useless); or better is not a conditional-influencing promise but simply a conditional

---

[2] We know that a double CIP/T is also possible with two independent prospected harm and advantage. *x* both disposes of a reward producing act and of a harm producing act and the harm is not just the non-production of the benefit, or the benefit is not simply the non-production of the damage. In this case the positive reward is the sum of the NotT plus the prize; while the negative reward is the sum of T plus NotP. This makes thing more complicated.

[3] That is if *x* tries to induce *y* to do *ay* by eliciting an emotional reaction [Miceli *et al.*].

[4] Obviously the thing is also complicated by the linguistic expression of the verb in negative or positive form: "I leave" → "If you do not leave" = "If you remain here"; "If you leave" = "If you do not remain here".

promise (not aimed at influencing), a free and post hoc prize:

"*if you finish your homework (as you are already doing), I will bring you to the movie*"

On the contrary, when *x* threatens *y* that *ax* if *ay*, he also presupposes that *y* is already and probably intending to do *ay* and wants to dissuade *y* from doing *ay*. Otherwise his promising is pragmatically incorrect and redundant (useless).

* "*if you do not finish your homework (that you are actually already doing), I will beat you*"

Premises and presuppositions regard not only *y*, but also *x*'s "side": the intention *y* expressed can be among *x*'s goals or it can endanger them. So it can be good or bad for *x* (what she likes or dislikes); if bad *x* wants to dissuade *y*; if good *x* wants to reinforce *y*'s intention. In particular, recalling the distinction between *requesting* and *defensive P/T* (where *ay* is something wanted/not wanted by *x*), it appears more natural to prefer (both in the surface and in the deep dimension):

16. REQUESTING-P/T: Deep-P > Deep-T = If ay then prize > If not ay then punishment

17. DEFENSIVE-P/T: Deep-T > Deep-P = If ay then punishment > If not ay then prize

This is based on a "reciprocation" rule[5]: since *x* is asking *y* to do something good for him, *x* should also do something good for him, in exchange (and vice-versa if *y* is doing something bad *x* should reciprocate with a punishment). This reciprocation rule reflects also on the surface form of the P/T: if *y*'s intention of doing *ay* is good for *x* and she wants to reinforce it the formulation "*If you do ay I will prize you*" looks a bit better than the formulation "*If you do not do ay I will punish you*". But of course this should be combined with the *NegPower* or *PosPower* of *x* on *y*; which makes thing more vague and complicated.

- If *x* has *PosPower* (when *x* disposes of a incentive producing act and the benefit is not simply the non-production of the damage):

18. REQUESTING P/T: P-form > T-form = If ay then prize > If not ay then not-prize

19. DEFENSIVE P/T: T-form > P-form = If ay then not-prize > If not ay then prize

- If *x* has *NegPower* (when *x* disposes of a harm producing act and the harm is not just the non-production of the benefit):

20. REQUESTING-P/T: P-form > T-form = If ay then not-punishment > If not-ay then punishment

21. DEFENSIVE-P/T: T-form > P-form If ay then punishment > If not-ay then not-punishment

When there is a distinct Pos and Neg Power, i.e. when *x* both disposes of an incentive producing act - *PRIZE(ax)* - and of a harm producing act - *PUNISHMENT(ax)* - and the harm is not just the non-production of the benefit, or the benefit is not simply the non-production of the damage we have a combination of formulae 18 with 20 and 19 with 21:

22. REQUESTING-P/T: If ay then prize > If not ay then not-prize > If ay then not-punishment > If not ay then punishment

23. DEFENSIVE-P/T: If ay then punishment > If not ay then not-punishment > If ay then not-prize > If not ay then prize

4 Factors Interaction

Obviously these are only questionable intuitions, just speaker's judgments. Moreover all the principles we have introduced interact with each other and it is difficult to discriminate which rule has the precedence above the others.

Let us suppose that persuader wants to show a benevolent face, but in the situation of a defensive P/T: by formulae 10 and 11 (benevolent face) we have that *Deep-P > Deep-T* and *P-form > T-form*. But from formulae 17, 19 and 21 (defensive P/T) we have *Deep-T > Deep-P* and *T-form > P-form*. And we should also consider the presupposition of the utterance (formulae 14 and 15).

Experimental data would be needed for establishing the real context-dependent preferences of speaker between Deep-P vs. Deep-T and P-face vs. T-face. We plan to start an experimental research about the capacity of human subjects to perceive the difference between 'deep' and 'superficial' P/T to evaluate the appropriateness of a given 'form' (P vs. T; negative vs. positive) in a given pragmatic and linguistic context and for a given influencing intention. Experimental data like these are in fact lacking, and we were able in this position paper only to use intuitive criteria. To our knowledge there are really few works addressing evaluation of CIP/T (among these [Verbrugge *et al.*, 04; Verbrugge *et al.*, 05]). Still Verbrugge and his colleagues do not differentiate in the evaluation experiment from deep and surface P/T and do not capture the double face of promise/threats. In our view this has somehow invalidate the results.

5 Conclusions

In this paper, we expanded our theory about Promises and Threats (P/T) by exploring some of the factors that influence the final realization of the P/T utterance. In particular we focused on Conditional influencing P/T. We tried to single out some of the factors that lead to the choice between Deep-P vs. Deep-T and its surface realization (as a promise or as a threat) of the CIP/T. We found *Objective*, *Personality and Emotional*, *Relational and social* and *Pragmatic-linguistic* factors.

---

[5] We can say that CIP/T are promises/threats to 'reciprocate': "if (and only if) you do this (good/bad) to me, I will do the same to you: a good/bad thing". To 'reciprocate' in fact means: 'to give in return'; to give something to *y*, to do something for *y* - because *y* gave/did something to us favorable or harmful - and to give the same (good for good, bad for bad).

6   Acknowledgements

References

[Amgoud & Prade, 04] L. Amgoud and H. Prade, 'Threat, reward and explanatory arguments: generation and evaluation', in *Proceedings of the ECAI Workshop on Computational Models of Natural Argument*, Valencia, Spain, (August, 2004).

[Castelfranchi, 03] Castelfranchi, C. 'The Micro-Macro Constitution of Power', ProtoSociology, An International Journal of Interdisciplinary Research Double Vol. 18-19, 2003 Understanding the Social II – Philosophy of Sociality, Edited by Raimo Tuomela, Gerhard Preyer, and Georg Peter.

[Cohen & Levesque, 90] Cohen P. and Levesque H. Persistence, Intention, and Commitment. P. Cohen, J. Morgan, and M. Pollack (Eds.), *Intentions in Communication*, pages 33--70, MIT Press, Boston, 1990.

[Cohen & Levesque, 87] Cohen P. R., Levesque H. J., *Persistence, intention, and commitment,* MIT Press, Cambridge, 1987.

[Kahneman *et al.*, 82] Kahneman D., Slovic P. and Tversky A. (Eds.) *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press, 1982

[Kraus *et al.,* 98] Kraus S., Sycara K. and Evenchik A. Reaching agreements trough argumentation: a logic model and implementation, *Artificial Intelligence Journal* 104:1--69, 1998.

[Guerini & Castelfranchi, 06] Guerini M. & Castelfranchi C. "Promises and Threats in Persuasion". in Proceedings of the ECAI Workshop on Computational Models of Natural Argument, Riva del Garda, 2006.

[Miceli *et al.*] Miceli, M., Poggi, I., De Rosis, F., & Carofiglio, V. (submitted) Emotional and Non-Emotional Persuasion.

[Sierra *et al.*, 98] C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons, 'A framework for argumentation-based negotiation', *Lecture Notes in Computer Science*, 1365, 177–192, (1998).

[Verbrugge *et al.*, 04] Verbrugge, S.; Dieussaert, K.; Schaeken, W. & Van Belle, W. (2004), 'Promise is debt, threat another matter. The effect of credibility on the interpretation of conditional promises and threats', *Canadian Journal of Experimental Psychology* **58 (2)**, 106--112.

[Verbrugge *et al.*, 05] Verbrugge, S.; Dieussaert, K.; Schaeken, W. & Van Belle, W. (2005),Compelling promises and hollow threats: why you can keep someone to his promise but not to his threat, *in* 'Proceedings of the 27th Annual Conference of the Cognitive Science Society'.