Promises and Threats in Persuasion

Marco Guerini [1] and **Cristiano Castelfranchi** [2]

Abstract. In this paper we analyse Promises and Threats (P/T) use in persuasion. Starting from a general definition of P/T based on the concepts of speech act and social commitment we focus on Conditional Influencing P/T (CIP/T): those incentive-based P/T used to persuade the addressee, rooted on dependence and power relations. We argue that in CIP/T class the concepts of promise and threat are strictly connected: the promise act is necessarily accompanied by a threat act and vice versa. Thus we discuss the problem of why the CIP/T are credible even if the speaker is supposed to be a rational agent and analyse some asymmetries between CIP and CIT. We also identify - beyond the rhetorical presentation - a deeper difference between substantial promises and substantial threats. Throughout the article is given a pre-formal model of these concepts.

1 INTRODUCTION

In this paper (based on a bigger research on P/T [8]) the concepts of promises and threats are analysed in order to gain some insight on their nature and their relations. The aim is to study P/T use in persuasion.

Starting from the concepts of speech act and social commitment we briefly show that not all P/T are for persuasion or conditional in their nature (like in "*if you do your homework I will bring you to the cinema*"): four different typologies of P/T are possible.

We then focus on Conditional Influencing P/T (CIP/T): those P/T used to persuade the addressee. In our analysis CIP/T are incentive-based influencing actions, rooted on dependence and power relations. These communicative actions affect the practical reasoning of the receiver by adding "artificial" consequences to the required action.

Finally we argue that in CIP/T class the concepts of promise and threat are two faces of the same coin. The deep logical form of these social acts is an IFF: the promise act is always and necessarily accompanied by a threat act ("*if you do not do your homework I will not bring you to the cinema*"), and vice versa.

Thus we discuss the problem of why the CIP/T are credible even if the speaker is supposed to be a rational agent and analyse some asymmetries between CIP and CIT. We also identify - beyond the rhetorical presentation - a deeper difference: a substantial threat, consisting in a choice between two losses, compared with substantial promises where the choice is between a gain and a missed-gain.

Throughout the article is given a pre-formal model for a computational treatment of these concepts. We adopt the Beliefs, Desires, Intentions (BDI) model as a reference framework [9, 10]. In the context of negotiating agents some simplified formalizations of CIP/T has been put forward, see for example [16, 1, 23]. Still, here we will focus on the implicit negotiational nature of CIP/T and not on their use in negotiation.

Hereafter variable x indicates the sender, and variable y the receiver, of the message.

2 PROMISES AND THREATS

2.1 What is a 'promise'

A Promise is, from a general point of view, a speech act that consists in the declaration, by x, of the *intention* of performing a certain action ax, under the pre-condition that ax is something wanted by y, with the aim of entering into an obligation (*social commitment*) of doing ax [20, 2, 22, 18]. A similar definition can be also found in the Webster Dictionary.

Intention = the notion of internal-commitment (intention) as defined by Bouron [3] establishes a relation between two entities: the agent x and the action ax.

$$INTEND(x\ ax) = GOAL(x\ DOES(x\ ax)) \quad (1)$$

This formula defines the intention of x to perform ax as the goal of x to perform the action in the next time interval (for a thorough definition see [10]).

Social commitment = the notion of social commitment (S-commitment) [5] involves four entities: the agent x, the action ax (that x has the intention to perform, for which he takes the responsibility), the agent y for which action ax has some value, and an agent z before whom x is committed (the witness).

$$S - COMMITED(x\ y\ ax\ z) \quad (2)$$

In the definition of S-commitment the key point is that x is committed to do ax because y is interested in ax. So a S-commitment is a form of goal adoption[3], and P/T are a particular form of social commitment.

When x promises something (ax) to y she is committing herself to do ax. This is not simply an internal commitment that stabilize x's choices and actions [4], and it is not simply a 'declaration of a personal intention'. In intention declaration x is committed about the action only with herself and she can change her mind. Instead in

[1] Itc-Irst, Istituto per la Ricerca Scientifica e Tecnologica, 38050 - Trento, ITALY, email: guerini@itc.it

[2] National Research Council - ISTC - Institute of Cognitive Sciences and Technologies via San Martino della Battaglia 44, 00185 - Roma, ITALY, email: c.castelfranchi@istc.cnr.it

[3] By '(Social) Goal-Adoption' we mean the fact that x comes to have a goal because and until she believes that it is a goal of y. x has the goal to 'help' y, or better (since 'help' is just a sub-case of social goal-adoption) x has the goal that y realizes/obtains his goal $GOAL(y\ p)$, thus decides to act for y by generating $GOAL(x\ p)$. This can be for various motives and reasons: personal advantages (like in exchange), cooperation (common higher goals), altruism, norms, etc. [11].

promises she is committed with the other, x has an interpersonal obligation - $OBL(x\ y\ DOES(x\ ax))$ - and creates some 'rights' in the other (entitled expectation & reliance/delegation, checking, claiming, protesting).

Moreover, being sincere in promising (i.e. being internally committed) is not necessary for a P/T to be effective. This commitment has an interpersonal and non-internal nature, there is a real created and assumed 'obligation' (see also [24]).

Let us better represent these features of a Promise:

a) x declare to y his intention to do ax

$$UTTER(x\ y\ INTEND(x\ ax)) \quad (3)$$

b) that is assumed to be in y's interest and as y likes,

$$GOAL(y\ DOES(x\ ax)) \quad (4)$$

c) in order that y believes and expects so

$$BEL(y\ INTEND(x\ ax)) \quad (5)$$

d) and y believes also that x takes a commitment to y, an obligation to y to do as promised.

$$BEL(y\ S-COMMITED(x\ y\ ax)) \quad (6)$$

e) The result of a promise is y's belief about ax, the public 'adoption' by x of a goal of y, y's right and x's duty about x doing ax.

$$BEL(y\ DOES(x\ ax)) \quad (7)$$

Finally, a promise presupposes the (tacit) agreement of y to be effective, i.e. to create the obligation/right. It is not complete and valid, for example, if y refuses (see section 2.4).

2.2 What is a 'threat' and P/T asymmetry in commitments

A threat is, from a general point of view, the declaration, by x, of the intention of performing a certain action ax, under the pre-condition that ax is something not wanted by y. Analytically, the situation is similar to promises apart from:

b1) ax is assumed to be against y's interest and what y dislikes,

$$GOAL(y\ \neg DOES(x\ ax)) \quad (8)$$

d1) x takes a commitment, an obligation to y to do as threaten.

In the threatening case, ax is something y dislikes (b1), and the consent or agreement of y is neither presupposed nor required. It is important to note that it is not strictly necessary that conditions (b) and (b1) hold before the P/T utterance. It is sufficient that ax is wanted (or not wanted) after that the P/T is uttered: P/T can be based on the elicitation or activation of a non-active goal of y[4].

P creates an obligation of x toward y, and corresponding rights of y about x's promised action. But this looks counter intuitive for T cases where ax is something y does not want[5]. To find an answer, we have to differentiate the two S-commitments that P creates.

[4] We thank Andrew Ortony for suggesting us to make this explicit and clear. On goal-activation see [6].

[5] One might also claim - for the sake of uniformity and simplicity - that in fact there are such a 'right' for y and such an obligation for x, but y will never exercise his rights and claim for them. One might support the argument with the example of the masochist (E2): if pain is a pleasure for y he can expect for x's 'promised' bad action, and can in fact claim for it, since x has committed himself on it.

S1) A S-commitment about the **truth** of what x is declaring (he takes responsibility for this) and this is the kernel of 'promising'

S2) A S-commitment on a future event under x's control. This is about the action that x has to accomplish in order to **make true** what he has declared.

In T the first commitment (S1) is there: y can blame and make fun of x for not keeping his word on what threatened: the reputation of x is compromised. But for the second more important social-commitment to do ax, there is an important asymmetry between P and T (conditions (d) and (d1)) that we will adjust in section 4.3.

2.3 Promises as public goal adoption

Our analysis, so far, basically converges with Searle's one, but in our view Searle missed the "adoption" condition, which is entailed by the notion of S-commitment (condition (d)). In order to have a promise, it is not enough (as seems compatible with his 4^{th} condition and not well expressed in his 5^{th} condition) that:

- x declares (informs y) to have a give intention to do action ax - condition (a) of our analysis
- x and y believe that y likes (prefers) that x does such an action - condition (b) of our analysis.

This is not a promise. For example:

E1) for his own personal reasons x has to leave, and informs y of his intention, and he knows that y will be happy for this; but this is not a 'promise' to y, since x do not intend to leave because y desires so.

While promising something to y, x is adopting a goal/desire of y. x intends to do the action since and until she believes that it is a goal for y; x's intention is "relativized" to this belief (see formula below).

$$REL-GOAL(x\ DOES(x\ ax) GOAL(y\ DOES(x\ ax))) \quad (9)$$

2.4 Y's agreement

The commitment, and the following 'obligations', of x to do ax is relativised to ax being a goal of y. So, for a felicitous promise the (tacit) acceptance of y is crucial; it is this (tacit) agreement that actually creates the obligation and the obligation vanishes if y does no (longer) desires/requires ax (condition (b)). This analysis is also valid for the threatening case, but in a reverse sense: the consent/acceptance is presupposed not to be given. The paradoxical joke of the sadist and the masochist, in example E2, points out clearly this case:

E2) Sadist: *"I will spank you!"* Masochist: *"Yes please!"* Sadist: *"No"*

But y, in declaring she does not want x to perform ax, is not necessarily negating her need for ax: there are different reasons that can bring y to reject x's help (e.g. not to feel in debt).

2.5 The notion of persuasion

There is a strong relation between P/T and persuasion; P/T are often used as persuasive means. We think there is a lack of theory on their relation. To analyse it we need a theory of persuasion (some preliminary ideas can be found in [15, 14]).

According to Perelman [19], persuasion is a skill that human beings use in order to make their partners perform certain actions or collaborate in various activities, see also [17]. This is done by modifying - through communication (arguments) - the other's intentional attitudes. In fact, apart from physical coercion and the exploitation of stimulus-response mechanisms, the only way to make someone do something is to change his beliefs [6].

We propose two different formalizations of "goal of persuading" (formulae 10 and 11). Formula 10 implies formula 11 when y is an autonomous agent (i.e. every action performed by an agent follows from an intention).

$$PERSUADE(x\ y\ ay) \rightarrow INTEND(x\ DOES(y\ ay)) \quad (10)$$

$$PERSUADE(x\ y\ ay) \rightarrow INTEND(x\ INTEND(y\ ay)) \quad (11)$$

Considering formula 11, 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 dissuasion as persuasion to not perform a given action.

$$DISSUADE(x\ y\ ay) \rightarrow PERSUADE(x\ y\ ay) \quad (12)$$

In analyzing the notion of 'intention', three cases must be considered. The intention of performing ay (formula 13), the intention of not performing ay (formula 14), and the lack of intention (formula 15).

$$INTEND(y\ ay) \quad (13)$$

$$INTEND(y\ \neg ay) \quad (14)$$

$$\neg INTEND(y\ ay) \quad (15)$$

Following the definitions from 13 to 15 we can model two different notions of persuasion and dissuasion:

- the *weak notion* captures the idea that the receiver is not already planning to perform the required action (formula 15);
- the *strong notion*, captures not only the idea that y is not already planning to perform ay, but also that he has some specific barriers against the action (y has some reason for not doing ay).

The terms "barriers/reasons" indicate those dispositions - of the receiver - that are against ay. In our approach barriers are modelled as contrary intentions: for any given action ay, the contrary intention is the intention of performing $\neg ay$ (formula 14). P/T, when used as persuasive means, refer only to the strong cases of persuasion (see section 3.3).

2.6 The main classes of P/T

There are four main classes of promises and threats. The distinction can be made along two dimensions: (a) presence of a conditional part in the P/T message, (b) presence of a persuasive aim in x (see table 1).

a) Some promises are conditional in their nature (e.g. "*If tomorrow is sunny I will bring you to the zoo*", "*If you do your homework I will bring you to the cinema*"). This dimension refers to the presence or the absence of a conditional part in the message

b) The second dimension refers to the presence or the absence, in the speaker of the intention to influence the hearer. If the predicate $PERSUADE(x\ y\ ay)$ holds, we are in the influencing class. This dimension is the most important in the division of P/T. In this paper we will focus on conditional-influencing class, central from a persuasive perspective.

	INFLUENCING	NON-INFLUENCING
CONDITIONAL	"*If ay then ax*" (CIP/T)	"*If c then ax*" (CP/T)
NON-CONDITIONAL	"*I will ax*" (IP/T)	"*I will ax*" (P/T)

Table 1. Main classes of promises and threats

3 THE INFLUENCING CLASSES

3.1 General Structure

The key question is: why should x perform an action positive or negative for y? And why x should want to communicate this to y?

This is done exactly with the aim of inducing y to perform (not to perform) some other action (ay). This is obtained by artificially linking a new effect (ax) to the action ay. This is the very nature of Influencing P/T (IP/T).

The two classes of IP/T can be considered both as conditional, because this is entailed by the influencing nature of IP/T, and we will refer to both as CIP/T. In non conditional cases, simply, x leaves implicit the conditional part for pragmatic reasons. The structure of the utterance 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 act.

- 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)*".

Generically, a CIP has a higher goal that ay, and the message is aimed at this goal. More precisely: when x utters the sentence, he has the goal that y believes that x is going to favour him ($G1$) with the super-goal ($G2$) to induce in y the intention to do ay. Finally $G2$ has another super-goal ($G3$) to induce y to perform ay (which is the ultimate goal of CIP/T). The cognitive structure is depicted in figure 1.

A CIT has the same structure, except that the influencing goals ($G2$ and $G3$) are the opposite of the condition of the utterance: $\neg ay$ and ay (for additional important differences in the plan, see section 2.7). The distinction between goals $G2$ and $G3$ is motivated by the two definitions of PERSUADE: to induce someone to act (formula 11), by creating the corresponding intention (formula 10). This distinction is necessary in those cases where CIP/T are used only to create an intention, as in example E3.

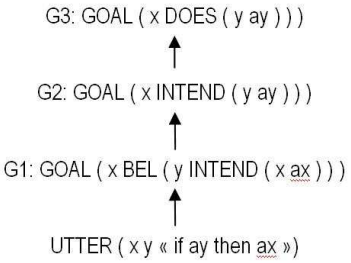

Figure 1. The goal structure of a CIP speech-act

E3) x, a lackey of a Mafia boss, promises to y, another lackey of the boss, to give him a huge money reward (ax) if he kills the boss (ay). But x wants to show to the boss that y is not loyal. The overall goal of his promise is just that y intends to kill the boss ($G2$), and not that he actually does it ($G3$).

3.2 The relation between persuasion/dissuasion and IP/T

In common sense, promises are for persuading and threats are for dissuading (see for example [12, 25]), but this is not true. The complete spectrum is depicted in table 2 ("+" means a benefit for y, "-" means a disadvantage).

	A. Persuading PP: $\neg INTEND(y\ ay)$ Gx: $INTEND(y\ ay)$	**B. Dissuading** PP: $INTEND(y\ ay)$ Gx: $\neg INTEND(y\ ay)$
1. Promise: y prefers ax	*"If ay then ax+"* (CIP/T)	*"If not ay then ax+"* (CP/T)
2. Threat: y prefers $\neg ax$	*"If not ay then ax -"* (IP/T)	*"If ay then ax -"* (P/T)

Table 2. The relation between Persuasion/Dissuasion and IP/T

In 1A and 1B, x is meaning: "*if you change your mind, I will give you a prize*"; i.e. the condition of the CIP is the opposite of the presupposition. While in 2A and 2B x is meaning: "*if you persist, do not change your mind, I will punish you*"; i.e. the condition of the CIT coincides with the presupposition.

3.3 CIP/T as "commissive requests"

Using Searle's terminology, CIP/T represent a *request* speech act by means of a *commissive* [22]. A set-based description of the various classes is given in figure 2.

There are different communicative acts (like "asking for", argumenting) with different "costs" that can be used to persuade. CIP/T are the most "expensive". In fact, given that every action has a cost, if y carries out ay, then x is committed to carry out ax (on this, see section 3.6). Why not simply asking for ay, or argumenting on the advantages, for y, to perform ay? If successful, x does not have any additional cost.

The answer relies on the necessity (following x) of using rewards (defined as "incentives", see section 3.4) and on the different presuppositions that lead to different persuasive acts.

1. In a simple request (lowest cost for x) y is presupposed to have no contrary intentions on ay (or that y's internal reward - like satisfaction, reciprocation - may suffice for overcoming y's barriers)

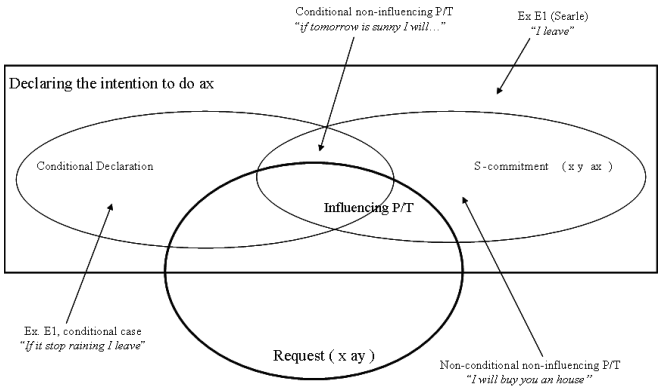

Figure 2. A set based description of the various classes of P/T and related concepts

2. In argumenting the presupposition is that, even if y can have some contrary intentions, when he will know all the outcomes of ay he will perform it.
3. In P/T (highest cost for x) instead the presupposition is not only that y has some contrary intentions, but also that there is no purely argumentative way to make him change his mind.

So, an influencing promise is a sort of combination between two different (linguistic) acts, an *offer* (*commissive* offer) of ax and a *request* for *ay*. In particular the offer is conditioned to the request.

3.4 Artificial consequences and incentives

In argumentation x can persuade y by prospecting "natural" positive or negative consequences of ay. But in CIP/T x has additional ways to persuade y to do ay:

- through the prospect of positive outcomes (whose acquisition is preferable) due to x's intervention (ax), not natural consequence of ay
- through the prospect of negative outcomes (whose avoidance is preferable) due to x's intervention (ax), not natural consequence of ay[6].

In CIP/T outcomes are linked to ay in an artificial way: "artificial" means that the consequence is under the control (direct or indirect) of x and will not happen without his intervention. With CIP/T arguments are "built" and not "found". This definition includes also the case in which ax is performed by a third, delegated, agent z. The fact is that this third agent will perform ax only if requested, and because delegated, by x. Let us consider the following examples:

E4) y's schoolmate: "*if you finish your homework your mother will bring you to the cinema*"
E5) y's mother: "*if you finish your homework I will tell your aunt to bring you to the cinema*"

These two examples show that being *natural* or *artificial* is strictly context dependent and the presence of an agent in the delivering of the outcome does not discriminate the two cases. In example E4 the same consequence of E5 (to be bring to the cinema) is used by the

[6] It is important to remark that 'not doing a' is an action (when is the output of a decision). Thus x can induce y to not doing something.

speaker in an argumentative way, by making the other believe or consider some benefits coming from her own action.

We consider CIP/T as social acts based on the prospect of incentives, where "incentives" are precisely those artificial consequences that are delivered - by x to y - in order to influence y. These incentives can be positive (*prizes*) or negative (*punishments*). In particular:

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

$$GOAL(y\ ax) \rightarrow PRIZE(ax) \tag{16}$$

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

$$GOAL(y\ \neg ax) \rightarrow PUNISHMENT(ax) \tag{17}$$

In table 3, we have a summary of the different typologies of outcomes of ay with the corresponding term to indicate them (similar to the distinction proposed in [12] between *conditionals inducements* and *conditional advices* classes). Incentives, promises and threats are on line B; prospected natural outcomes, instead, are on line A.

	POSITIVE OUTCOMES	NEGATIVE OUTCOMES
A. Natural Consequences	Advantages	Disadvantages/ Drawbacks
B. Artificial Consequences	**Prizes**	**Punishments**

Table 3. Different typologies of ay outcomes

3.5 Credibility, preferability pre-conditions and the power of x

Many pre-conditions of the P/T act have to be met in order to have a felicitous communication: a P/T must be *credible* and convincing (*preferable*).

1) *Credibility pre-conditions*: The fact that the loss or gain for y is due to x's decision and intervention, explains why, in order to have a "credible" promise or threat, it is crucial that y believes that x is in condition to favour or to damage her. Thus when x announces his promise or threat he also has the goal that y believes that x has the "power of" ax; this belief is y's "trust" in x and it can be based on x reputation, on previous experience, on some demonstration of power, etc.[7]

Thus in order to have true promises or threats, x must have some power over y; the power of providing to y incentives (or at least y must believe so). More analytically:

- x has some *power of* doing ax

$$CAN - DO(x\ ax) \tag{18}$$

- y depends on x, and more precisely on his action ax, as for achieving some goal Gy;

$$DOES(x\ ax) \rightarrow Gy \tag{19}$$

$$DEPEND(y\ x\ ax\ Gy) \tag{20}$$

This means that:

- x gets a *power over* y's goal Gy, the power of giving incentives or not to y by the realization of Gy;

$$POWER - OVER(x\ y\ Gy) \tag{21}$$

- both x and y believe so[8];

$$BMB(x\ y\ POWER - OVER(x\ y\ Gy)) \tag{22}$$

on such a basis:

- x gets a power of influencing y to do ay while using the promise of Gy (performing ax) as an incentive [9].

$$PERSUADE(x\ y\ ay) \tag{23}$$

$$PRIZE(ax) \tag{24}$$

That is, x can make y believe that "if y performs ay (adopts the goal of x) then x will reward her by performing ax (adopting y's goal)".

2) *Preferability pre-conditions*: The above conditions represent the applicability conditions for P/T, but there is still another condition to be met in order to make CIP/T effective:

- If x has the power to jeopardise (or to help achieve) a goal Gy of y, and the goal has a higher value than the value of the action (ay), then x can threaten y to jeopardise the goal if he does not perform ay (or promise to help him realise his goal if he performs ay).

$$V(Gy) > V(ay) \tag{25}$$

Preferability conditions regard only the effectiveness of the message. "*If you carry that heavy bag for five kilometres I will give you 20 cents*": this is a true and credible promise, but ineffective (not preferable), because x has the *power of* giving 20 cents to y but the value of ay (carrying the heavy bag for five kilometres) is much greater the value of Gy (gaining 20 cents).

3.6 Scelling's plan asymmetry and inefficacy paradox in CIP/T

Plan asymmetry: in order to be efficacious the promised or threatened action ax must have an higher value than the requested action ay (in y's perspective)[10]: $V(ax) > V(ay)$. On the other side (in x's perspective), the promised action ax (that is: x's cost) has to have less value than ay: $V(ax) < V(ay)$. It represents x's costs. However, there is an asymmetry between P and T under this respect (considering those P/T where ax is an action to be performed and not the abstaining from an action).

- In Promises, x - if sincere - plans (intends) to do ax in order to obtain ay. In case of a successful P it is expected that x performs ax.
- In Threats, x plans the *non* execution of ax. It should be executed only in case of failure and y's refusal[11].

[7] This is why a mafia's warning is not usually limited to a simple (verbal) message, but is a concrete harm (beating, burning, etc.). This is a 'demonstrative' act (that is communication) but with the advantage to directly show and make credible the threatening power of the speaker [7]. On the use of fear and scare tactics in threats see also [26].

[8] We do not address here the problem of false P/T, like in the case of an armed robbery with a fake gun.

[9] The power of influencing y to do something can based not only on incentive power, but also on imitation, reactive elicitation, normative endowment, etc.

[10] $V(ax)$ for y is equivalent to $V(Gy)$ since $ax \rightarrow Gy$

[11] This is the genial intuition of Schelling [21] (p.36, especially note 7, p. 123) but within an not enough sophisticated theory of P/T.

This difference is especially important in substantial P vs. substantial T (see later). Under this respect a T looks more convenient than a P: a successful T has only communication/negotiation costs.

Though, there are serious limits in this 'convenience', not only from the point of view of social capital and collective interest, but also from x's point of view. In fact in those kinds of relationships y is leaning to exit from the relation, to subtract herself from x (bad) power and influence. It requires a lot of control and repression activity for maintaining people under subjection and blackmail.

Inefficacy paradox: in threats, ax (detrimental for y) should be executed only in case of failure/inefficacy of the threat, but why x should perform it and having useless costs? [21]. Surely not for achieving the original goal - $DOES(y\ ay)$ -. Thus, it seems irrational to do what has been threatened.

Moreover, that this action would be useless for x should be clear also to y, and this makes x's threat non credible at all: y knows that x (if rational) will not do as threaten if unsuccessful; so why accepting?

Analogously, the promised action (beneficial for y) usually[12] has to be performed by x in case of success, so why should x spend his resources when he already obtained his goal? But this is known by y and should make x's promise not very credible.

As Shelling suggests, threats (and promises) should be performable in steps: the first steps are behavioural messages, demonstration of the real power of x, warnings or "lessons". However, this is just a sub-case; the general solution of this paradox has to be found in *additional and different reasons and motives of* x.

Let's consider threats. In keeping threats after a failure, x aims at giving a "lesson" to y, at making y learning (for future interactions with x or with others) that (x's) threats are credible. This can be aimed also at maintaining the reputation of x as a coherent and credible person. Another motive can be just rage and the desire of punishing y; TIT for TAT. In keeping promises after success - a part from investing in reputation capital - there might be 'reciprocation' motives, or fairness, or morality, etc.

If these additional motives are known by y, they make x's P/T credible; but it is important to have clarified that:

- if x performs what he promised it is **not** in order to obtain what he asked for.

4 THE JANUS NATURE OF CIP/T

4.1 Logical form of CIP/T

No P/T of the form "*if ay I will ax*" would be effective if it does not also mean "*if not ay I will not ax*", that is: if it would mean "*if ay I will* ax*, and also if not ay*". x can either plan for persuading y to ay or for dissuading y from not ay. He can say: "*if ay I will give you a positive incentive*" (promise) or "*if not ay I will give you a negative incentive*" (threat).

In these cases, one act is only the implicit counterpart of the other and the positive and negative incentives are simply one the negation of the other ("*I will do ax*" vs. "*I will not do ax*"). Also for this reason, one side can remain implicit. A threat is aimed at inducing an avoidance goal, while a promise is aimed at eliciting attraction, but they co-occur in the same influencing act[13]. Though the two P/T are not an identical act they are two necessary and complementary parts of the same communicative plan.

Despite the surface IF-THEN form of CIP/T, our claim is that the deep logical form is an IFF[14]. There is no threat without promise and vice versa. In the (intuitive) equivalence between: "*if you do your homework I will bring you to the cinema*" (promise) and "*if you do not do your homework I will not bring you to the cinema*" (threat), the logical IF-THEN interpretation doesn't work:

$$(ay \rightarrow ax) \neq (\neg ay \rightarrow \neg ax) \tag{26}$$

while this is the case for the IFF interpretation:

$$(ay \leftrightarrow ax) = (\neg ay \leftrightarrow \neg ax) \tag{27}$$

4.2 Deep and surface CIP/T

Only a pragmatic difference seems to distinguish between P and T as two faces of the same act (here we will not address the problem of how x decides which face to show). 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 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 4, 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),
- deep promises (gain): a choice between a gain (greater then the cost) or a missed gain.

On the rows we have the surface form of the corresponding communicative acts: in the case of surface promise what is promised is a missing loss or a gain, while in the case of surface threat 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.

What is explained in table 4 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 "aggressive" ones (in which ay is something wanted by x).

[12] There are promises of this form: "*I will do ax if you promise to do ay*". In this case the promised action ax has to be performed before ay. In such conditions there is no reason for x to defeat.

[13] It is also possible to have independent and additional positive and negative incentives, in a strange form of double Threat-Promise act like the following one: "*If you do your homework I will bring you to movie; if you do not do your homework I will spank you*".

[14] We mean that the correct logical representation of the intended and understood meaning of the sentence is an IFF. One can arrive to this either via a pragmatic implicature [13] or via a context dependent specialized lexical meaning (see later).

	Deep T: Loss (scenario A)	**Deep P: Gain** (scenario B)
Surface Promise	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	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 4. Deep and surface P and T

4.3 CIP/T and their commitments

The analysis just introduced on the logical structure of CIP/T allows us, now, to define the different kinds of commitments entailed by promises and threats (points d and d1 of our analysis, see section 1.3). As we already saw (section 2.2 and note 5) apparently, threats seem to fall out of our analysis in terms of S-commitment. In threats the committed action is not, superficially, a y's goal. If x does not keep his commitment, y won't protest. But, given that every threat entails a promise - at least for CIP/T - the asymmetry can be solved: the S-commitment in threats is taken on the corresponding promise form. So:

- Promise: $(COMMITTED\ x\ y\ \underline{ax}\ z)$ where ax is "I will bring you to the cinema"
- Threat: $(COMMITTED\ x\ y\ \underline{\neg ax}\ z)$ where ax is "I will spank you"

In the first case y can protest if x does not perform the action, in the second, instead, y can protest if x performs the action[15].

But the commitment structure of CIP vs. CIT is even more complex: we need the concept of "Pact" - or "Mutual S-commitment" - in which the commitment of x with y is conditioned to the commitment of y with x and vice versa. In fact any P presupposes the 'agreement' of y (see section 2.4), a tacit or explicit consent, or a previous request by y. This means that y takes a S-Commitment toward x to accept his 'help' and to rely on his action [5]. x will protest (and is entitled to) if y solves the problem on his own or ask someone else.

In our view an accomplished promise is a Multi-Agent act, it requires two acts, two messages and outputs with two commitments. It seems necessary to go - thank to the notion of conditional reciprocal goal-adoption - beyond the enlightening notion of Reinach [20] (cited and discussed in [18]) of 'social act' as an act which is etherodirected, that needs the listening and "grasping" of the addressee.

Moreover, there's the need of a distinction between "negative pacts" (based on threats) and "positive pacts" (based on promises), they entail different S-commitments.

- In CIP x proposes to y to 'adopt' her goal (ax) if y adopts his own goal (ay); he proposes a **reciprocal goal-adoption**, and exchange of favors.
- In prototypical CIT we have the complementary face. x is proposing to y an **exchange of abstentions from harm** and disturb. The reciprocal S-commitments are formulated and motivated by avoidance, in both x and y.

[15] Even from a threatening point of view is counterproductive for x not to respect the "promise" after a successful threat. In fact x would be perceived as unfair if she were to spank the kid after he did his homework.

5 CONCLUSIONS

In this paper we analysed the persuasive use of Promises and Threats. Starting from the definition of P/T as "speech acts creating social-commitments" and the definition of persuasive goal, we showed that not all P/T are for persuasion or conditional in their nature.

We then focused exactly on those conditional P/T that are intended to influence - persuade - the addressee (CIP/T). In our analysis CIP/T are incentive-based influencing actions for overcoming y's resistance to influence; they are based on x's power over y's goals.

We claimed that in CIP/T class the concepts of promise and threat are two faces of the same coin: a promise act is always and necessarily accompanied by an act of threat, and vice versa.

We also identified - beyond the rhetorical presentation - a deeper difference: a substantial threat and a substantial promise (independent of the presented 'face'). A plan asymmetry between P/T and a paradox of CIP/T, that should be non-credible in principle, were also introduced.

The aim of this work was to give a pre-formal model of P/T as a basis for a computational treatment of these concepts.

ACKNOWLEDGEMENTS

This work was partially supported by the HUMAINE project.

We would like to thank the referees for their comments which helped improve this paper.

REFERENCES

[1] 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).
[2] J.L. Austin, *How to do things with words*, University Press, Oxford, 1962.
[3] T. Bouron, *Structures de Communication et dOrganisation pour la Coopration dans un Univers Multi-agent*, Ph.D. dissertation, Universit Paris 6, 1992.
[4] M. Bratman, *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge, Mass., 1987.
[5] C. Castelfranchi, 'Commitments: from individual intentions to group and organizations', in *Proceedings of the First International Conference on Multi-Agent Systems*, S. Francisco, (1995).
[6] C. Castelfranchi, 'Reasons: Beliefs structure and goal dynamics', *Mathware & Soft Computing*, **3(2)**, 233–247, (1996).
[7] C. Castelfranchi, 'Silent agents. from observation to tacit communication.', in *Proceedings of the workshop 'Modeling Other Agents from Observations' (MOO 2004).*, NY, USA, (2004).
[8] C. Castelfranchi and M. Guerini, 'Is it a promise or a threat?', Technical report, ITC-Irst Technical report T06-01-01, (January 2006).
[9] P. R. Cohen and H. J. Levesque, *Intentions in Communication*, chapter Rational interaction as the basis for communication, 221–256, The MIT Press, Cambridge, MA, 1990.
[10] P. R. Cohen and H. J. Levesque, *Intentions in Communication*, chapter Persistence, Intention, and Commitment, 33–69, MIT Press, Cambridge, MA, 1990.
[11] R. Conte and C. Castelfranchi, *Cognitive and Social Action*, UCL Press, London, cognitive and social action edn., 1995.
[12] J. St. B. T. Evans, 'The social and communicative function of conditional statements', *Mind & Society*, **4(1)**, 97–113, (2005).
[13] H. P. Grice, *Speech Acts*, chapter Logic and conversation, 4158, New York: Academic Press, 1975.
[14] M. Guerini, *Persuasion models for multimodal message generation*, Ph.D. dissertation, University of Trento., 2006.
[15] M. Guerini, O. Stock, and M. Zancanaro, 'Persuasion models for intelligent interfaces', in *Proceedings of the IJCAI Workshop on Computational Models of Natural Argument*, Acapulco, Mexico, (2003).

[16] S. Kraus, K. Sycara, and A. Evenchik, 'Reaching agreements trough argumentation: a logic model and implementation', *Artificial Intelligence Journal*, **104**, 1–69, (1998).
[17] B. Moulin, H. Irandoust, M. Belanger, and G. Desordes, 'Explanation and argumentation capabilities: Towards the creation of more persuasive agents', *Artificial Intelligence Review*, **17**, 169–222, (2002).
[18] K. Mulligan, *Speech Act and Sachverhalt*, chapter Promises and others social acts: Constituents and Structure, 29–90, Dordrecht, 1987.
[19] C. Perelman and L. Olbrechts-Tyteca, *The new Rhetoric: a treatise on Argumentation*, Notre Dame Press, 1969.
[20] A. Reinach, *Samtliche Werke, 2 Bde*, chapter Die apriorischen Grundlagen des burgerlichen Rechtes, Philosophia, Munchen, 1989.
[21] T. Schelling, *The Strategy of Conflict*, Harvard University Press, Cambridge, 1960.
[22] J. Searle, *Speech Acts*, Cambridge University Press, Cambridge, 1969.
[23] 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).
[24] M.P. Singh, 'Social and psychological commitments in multiagent systems', in *Proceedings of Knowledge and Action at Social & Organizational Levels, AAAI Fall Symposium Series*, ed., California Menlo Park, pp. 104–106. American Association for Artificial Intelligence, Inc., (1991).
[25] V. A. Thompson, J. St. B. T. Evans, and S. J. Handley, 'Persuading and dissuading by conditional argument', *Journal of Memory and Language*, **53(2)**, 238–257, (2005).
[26] D. Walton, *Scare Tactics: Arguments that Appeal to Fear and Threats*, Kluwer, Dordrecht, 2000.