The logical handling of threats, rewards, tips, and warnings

Leila Amgoud Jean-Francois Bonnefon Henri Prade

Institut de Recherche en Informatique de Toulouse (IRIT)
118, route de Narbonne,
31062 Toulouse Cedex 4 France
{amgoud, bonnefon, prade}@irit.fr

Abstract. Previous logic-based handling of arguments has mainly focused on explanation or justification in presence of inconsistency. As a consequence, only one type of argument has been considered, namely the explanatory type; several argumentation frameworks have been proposed for generating and evaluating explanatory arguments. However, recent investigations of argument-based negotiation have emphasized other types of arguments, such as *threats*, *rewards*, *tips*, and *warnings*. In parallel, cognitive psychologists recently started studying the characteristics of these different types of arguments, and the conditions under which they have their desired effect. Bringing together these two lines of research, we present in this article some logical definitions as well as some criteria for evaluating each type of argument. Empirical findings from cognitive psychology validate these formal results.

Keywords: Argumentation, Negotiation, Threats/Rewards, Tips/Warnings.

1 Introduction

Argumentation is an established approach for reasoning with inconsistent knowledge, based on the construction and the comparison of arguments, and it may also be considered as an alternative method for handling qualitative uncertainty. A basic idea behind argumentation is that it should be possible to say more about the certainty of a particular fact than just assessing a certainty degree in $[0, 1]$. In particular, it should be possible to assess the reason why a fact holds, under the form of arguments, and combine these arguments for evaluating the certainty of the fact they support. This combination process can be viewed as determining the most acceptable among arguments.

Various argument-based frameworks have been developed in defeasible reasoning [1, 6, 8, 20, 22], for generating as well as for evaluating arguments. However, in that explanation-oriented perspective, only one type of argument has been considered, namely the *explanatory* type (reasons for believing, explanations for states of affairs). Yet, another line of work [2, 15, 19] suggests that argumentation can also play a key role in negotiation: E.g., an offer supported by a good argument has a better chance of being accepted by another agent.

Argumentation may also lead an agent to change its goals, or may impose a particular response onto an agent.

In addition to the explanatory arguments studied in classical argumentation frameworks, the above works have emphasized other types of arguments such as inducements, deterrents, and pieces of advice. For example, if an agent receives a *threat*, it may accept an offer even though this offer has no particular appeal, so as not to jeopardize the truly important goals targeted by the threat. In parallel, cognitive psychologists have studied in recent years the characteristics of these different types of arguments, and the conditions under which they have their desired effect. Bringing together these lines of research, we present in this article the formal definitions of the four basic non-explanatory arguments (threats, rewards/promises, warnings, and tips[1]), as well as some criteria for evaluating them. Empirical findings from cognitive psychology validate our formal results.

2 The Four Basic Non-Explanatory Arguments

It has been pointed out that it is not possible to present an exhaustive classification of arguments, because arguments operate within a particular context and domain [24]. For example, when inferring from inconsistent knowledge bases, arguments aim at finding the most supported beliefs. But during a negotiation, the exchange of arguments may lead the agent which receives them to change its goals or preferences.

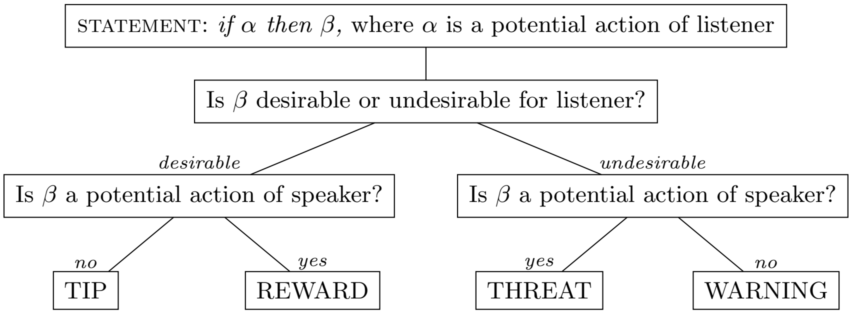
STATEMENT: *if* α *then* β, where α is a potential action of listener
Is β desirable or undesirable for listener?
desirable
undesirable
Is β a potential action of speaker?
Is β a potential action of speaker?
no
yes
yes
no
TIP
REWARD
THREAT
WARNING

Fig. 1. A decision tree for classifying arguments, adapted from [17].

Nonetheless, some typologies exist that consider the kinds of arguments thought to have persuasive force in human negotiation, both in artificial intelligence [15] and in cognitive psychology [17]. Building on previous research [9, 16], López-Rousseau and Ketelaar recently tested a simple yet remarkably efficient algorithm for predicting whether people will think of a given conditional

[1] Although the term ‘tip’ can evoke a small piece of heuristic information for making something better, it must be understood here in the sense of a *recommendation*.

statement as expressing a threat, a reward, a tip, or a warning (see Figure 1). Consider that a speaker is telling a listener: “If you do α, β will happen.” (Note that α is necessarily a potential action of the listener.) The algorithm of [17] focuses on two characteristics of β, namely: Is β something the speaker will do, or something that will happen independently of the speaker? Is β something good for the listener, or something bad? Quite remarkably, this simple algorithm correctly predicted 92% of the classifications made by human subjects.

In parallel, other authors [4] elaborated an in-depth psychological analysis of the motivational structure of such statements, and emphasized that the action α itself, inasmuch as rewards or threats are concerned, should have positive or negative consequences for the speaker. Indeed, why would the speaker attempt to bribe the listener into doing α if the speaker had no interest in seeing that α is done? Likewise, why would the speaker attempt to scare the listener out of doing α if the speaker had no interest in seeing that α is not done?

Our goal in this article is to organize psychological analysis and empirical results in a formal framework that will do justice to the psychological state of the art. To our knowledge, this framework is the first to address all four types of non-explanatory arguments, as well as the first to be entirely grounded in experimental research.

3 Formal Definitions

In what follows, $\mathcal{AC}$ denotes a set of actions. $\{a_1, \ldots, a_n\}$ is a set of agents involved in a discussion. In addition to this set of agents, we suppose that we have a neutral agent, denoted by a_0, that may stand for impersonal powers such as Nature itself. Let $\mathcal{AG} = \{a_0, a_1, \ldots, a_n\}$ be the set of all agents. Each agent is supposed to have the control over a subset of actions of $\mathcal{AC}$. This captures the fact that an agent is able to do some actions but not others. The function

$$\mathcal{F} : \mathcal{AG} \longrightarrow 2^{\mathcal{AC}}$$

retuns the actions under the control of each agent.

Each action performed by a given agent (including the neutral agent) is supposed to have consequences for all agents. These consequences can be good, neutral, or bad, and can be good or bad to different degrees. This notion of consequence is captured by the following function:

$$\texttt{Cons} : \mathcal{AC} \times \mathcal{AG} \longrightarrow \{-n, \ldots, 0, \ldots, +n\},$$

where n is an integer that denotes the extremity of the consequence of some action to some agent. Positive values of Cons denote good consequences, the higher the value of Cons the better. Negative values of Cons denote bad consequences, the lower the value of Cons the worse. The value 0 is attached to neutral consequences. This simple, ordinal scale can be generalized to more sophisticated scales, providing that they include a neutral point. Throughout the paper, we suppose that agent S, the speaker, addresses a negotiation move (a statement) to agent L, the listener, with $S, L \in \{a_1, \ldots, a_n\}$.

Definition 1 (Argument) *An argument is an expression of the form* $(a_i, \alpha) \longrightarrow (a_j, \beta)$ *such that:*

1. $a_i, a_j \in \mathcal{AG}$,
2. $\alpha, \beta \in \mathcal{AC}$

The meaning of the above expression is that if agent a_i performs action α, the agent a_j will perform action β.

3.1 Threats

Threats are used to coerce an agent into behaving in a certain way, by emphasizing the unpleasant measures the speaker would take otherwise. Different linguistic expressions of threats are possible. The conditional expression is canonical, but threats can easily be reformulated as conjunctions or disjunctions [11].

i) If you do α, I will do β,
ii) Do α and I will do β,
iii) Do not do α otherwise I will do β.

Example 1 (Tantrum)

i) If you throw a tantrum, I'll ground you.
ii) Throw a tantrum and I will ground you.
iii) Don't throw a tantrum, otherwise I will ground you.

Definition 2 (Threat) *An argument of type threat, or a* threat*, is an argument* $(a_i, \alpha) \longrightarrow (a_j, \beta)$ *such that:*

1. $a_i = L$
2. $\alpha \in \mathcal{F}(a_i)$
3. $a_j = S$
4. $\texttt{Cons}(\alpha, a_j) < 0$
5. $\texttt{Cons}(\beta, a_i) < 0$

Since $S, L \in \{a_1, \ldots, a_n\}$, neither can be the neutral agent. Points 1 and 2 are common to the definition of threats, rewards, tips, and warnings. They ensure that α is an action under the control of the listener; otherwise, the threat (reward, etc.) would be *useless*. Point 3 ensures that β is an action of the speaker, a characteristic feature of threats and warnings.[2] Point 4 ensures that the speaker does not attempt to prevent something that would actually be beneficial; otherwise, the threat would be *irrational*. Finally, Point 5 ensures that β is something unpleasant to the listener; otherwise, the threat would be *misplaced*.

In Example 1, α is meant to be an action of the listener (a child), namely, throwing a tantrum. This action (or lack thereof) is presumed to be under the control of the child. In contrast, β is meant to be an action of the speaker (the mother), namely, grounding the child. The child throwing a tantrum is something unpleasant to the mother, and being grounded is something unpleasant to the child. The statement meets all the criteria in the definition of a threat.

[2] We will return in section 5 to the fact that Definition 2 does not feature the condition $\beta \in \mathcal{F}(a_j)$.

3.2 Rewards

Rewards are used to encourage another agent to behave in a certain way, by emphasizing the pleasant measures the speaker will take in response. There are two main linguistic expressions of rewards. As for threats, the conditional expression of rewards is canonical, but the conjunctive reformulation is possible. Unlike threats, the disjunctive paraphrase is awkward [11].

i) If you do α, I will do β.
ii) Do α and I will do β.

Example 2 (Free CDS)

i) If you buy this computer, I'll throw in a box of free CDs.
ii) Buy this computer and I'll throw in a box of free CDs.
iii) Don't buy this computer, otherwise I'll throw in a box of free CDs.

Definition 3 (Reward) *An argument of type reward, or a* reward*, is an argument* $(a_i, \alpha) \longrightarrow (a_j, \beta)$ *such that:*

1. $a_i = L$
2. $\alpha \in \mathcal{F}(a_i)$
3. $a_j = S$
4. $\texttt{Cons}(\alpha, a_j) > 0$
5. $\texttt{Cons}(\beta, a_i) > 0$

Note that, due to the fact that $S, L \in \{a_1, \ldots, a_n\}$, neither can be the neutral agent. Points 1 and 2 serve the same function as in the definition of threats. Point 3 ensures that β is meant to be an action of the speaker, a characteristic feature of threats and rewards. Point 4 ensures that the speaker does not attempt to bring about something that would actually be detrimental; otherwise, the reward would be *irrational.* Finally, Point 5 ensures that β is something pleasant to the listener; otherwise, the reward would be *misplaced.*

In Example 2, α is an action under the control of the listener of the listener (a customer), namely, buying a computer. In contrast, β is an action of the speaker (the salesperson), namely, throwing in a box of free CDs. The customer buying a computer is something desirable for the salesperson, and being given a box of free CDs is something desirable to the customer. The statement meets all the criteria in the definition of a reward.

3.3 Warnings

Warnings are addressed to another agent in an attempt to discourage a given course of action, by emphasizing the unfortunate consequences that would follow. In contrast to threats, these unfortunate consequences are not within the control of the speaker [10], and the speaker has no particular stake in preventing the course of action to occur [18]. Just as threats, warnings can be formulated conditionally, conjunctively, or disjunctively.

Example 3 (Computer Virus)

i) If you open this file, your computer will crash.
ii) Open this file and your computer will crash.
iii) Don't open this file, otherwise your computer will crash.

Definition 4 (Warning) *An argument of type warning, or a* warning*, is an argument* $(a_i, \alpha) \longrightarrow (a_j, \beta)$ *such that:*

1. $a_i = L$
2. $\alpha \in \mathcal{F}(a_i)$
3. $a_j \neq S$
4. $\mathtt{Cons}(\beta, a_i) < 0$

Points 1 and 2 are common to all four definitions. Point 3 states β should be an action of another agent than the speaker (possibly an action by the impersonal agent). This is characteristic of tips and warnings. Point 4 ensures that β is an unpleasant consequence for the listener; otherwise, the warning would be *misplaced.* Note that the definition of a warning differs in two important respects from that of a threat. First, β is not an action of the speaker; second, it is not necessary (though not excluded, either) that α harms the speaker.

In Example 3, α is meant to be an action of the listener (a computer user), namely, opening a file. This action is under the control of the user. Action β (namely, crashing the computer) is not meant to be an action of the speaker (some hotline operator), but an action of a 'neutral' agent, the computer virus. Finally, whilst the crashing of the computer is certainly undesirable to the listener, the opening of the file is of no concern to the speaker. The statement meets all the criteria in the definition of a warning, but not the criteria in the definition of a threat (or a reward, of course).

3.4 Tips

Tips are addressed to another agent in an attempt to encourage a given course of action, by emphasizing the positive consequences that would follow. In contrast with rewards, these positive consequences are not within the control of the speaker, and the speaker has no particular stake in seeing that the course of action is taken. Just as rewards, tips can be formulated conditionally or conjunctively, but sound awkward when paraphrased disjunctively.

Example 4 (Revise and Resubmit)

i) If you revise the paper, the editor will accept it.
ii) Revise the paper and the editor will accept it.
iii) Don't revise the paper, otherwise the editor will accept it.

Definition 5 (Tip) *An argument of type tip, or a* tip*, is an argument* $(a_i, \alpha) \longrightarrow (a_j, \beta)$ *such that:*

1. $a_i = L$
2. $\alpha \in \mathcal{F}(a_i)$
3. $a_j \neq S$
4. $\texttt{Cons}(\beta, a_i) > 0$

Points 1 and 2 are common to all four definitions. Point 3 states β should be an action of another agent than the speaker (possibly an action by the impersonal agent). This is characteristic of tips and warnings. Point 5 ensures that β is indeed a pleasant consequence for the listener; otherwise, the tip would be *misplaced.* Note that the definition of a tip differs from that of a reward in two different respects. First, β is not an action of the speaker; second, it is not necessary (though not excluded, either) that α benefits the speaker.

In Example 4, α is meant to be an action of the listener (a graduate student), namely, revising a paper. This action is under the control of the student. Action β (namely, accepting the paper) is not meant to be an action of the speaker (a post-doctoral student met at a conference), but an action of a third agent, the editor. Finally, the acceptance of the paper is of course desirable to the listener, but the speaker has no particular stake in seeing that the paper is revised. The statement meets all the criteria in the definition of a tip, but not the criteria in the definition of a reward (or a threat, or a warning).

4 General Properties

We assume *symmetrical control:* An agent who controls action α also controls $\neg\alpha$: $\alpha \in \mathcal{F}(a_i) \iff \neg\alpha \in \mathcal{F}(a_i)$.[3] Furthermore, we assume *bipolar consequences:* When an action α has positive consequences for an agent, $\neg\alpha$ has negative consequences for this same agent: $\texttt{Cons}(\alpha, a_i) > 0 \iff \texttt{Cons}(\neg\alpha, a_i) < 0$.

Proposition 1 (Exclusive Definitions) *An argument $(a_i, \alpha) \longrightarrow (a_j, \beta)$ is either a threat, or a reward, or a tip, or a warning, or none of these. All* or*s in the previous sentence are exclusive.*

It follows trivially from the definitions we have given that an argument can only meet the criteria in one definition, but not two. An example of an argument that does not satisfy any of the four definitions is $(a_i, \alpha) \longrightarrow (a_0, \beta)$, where $a_i \in \mathcal{AG} \backslash \{S, L\}$. E.g., 'If my CEO admits the fraud, her stocks will go down.' We will get back to this kind of 'consequential arguments' [7] in the final section of this article. Although Proposition 1 is straightforward, it is a genuine improvement over previous frameworks that failed to give non-overlapping definitions of threats, rewards, tips, and warnings [3, 12].

Proposition 2 (From Threats to Rewards) *If $(a_i, \alpha) \longrightarrow (a_j, \beta)$ is a threat, then $(a_i, \neg\alpha) \longrightarrow (a_j, \gamma)$ is a reward for any γ such that* $\texttt{Cons}(\gamma, a_j) > 0$.

[3] Note that $\neg\alpha$ means 'not executing α' and not 'executing some action with the complementary effect of α'.

Proof. If $(a_i, \alpha) \longrightarrow (a_j, \beta)$ is a threat, then $a_i = L$, $\alpha \in \mathcal{F}(a_i)$, and $a_j = S$. The first three criteria for $(a_i, \neg\alpha) \longrightarrow (a_j, \gamma)$ to be a reward are thus satisfied. Furthermore, $\texttt{Cons}(\alpha, a_i) < 0$, which implies, under the assumption we have made, that $\texttt{Cons}(\neg\alpha, a_i) > 0$. The fourth criteria is satisfied. What remains to be satisfied is the fifth criteria, i.e., $\texttt{Cons}(\gamma, a_j) > 0$. Note that this criterion will be automatically satisfied in the particular case where γ is $\neg\beta$.

Proposition 3 (From Rewards to Threats) *If $(a_i, \alpha) \longrightarrow (a_j, \beta)$ is a reward, then $(a_i, \neg\alpha) \longrightarrow (a_j, \gamma)$ is a threat for any γ such that $\texttt{Cons}(\gamma, a_j) < 0$.*

Proof. Proof is similar to that of Proposition 2, and the same remark holds about the particular case where γ is $\neg\beta$.

Example 5 (Threat to reward, and vice versa) *The threat 'If you throw a tantrum, I'll ground you' becomes a reward when its antecedent is negated and its consequent replaced by anything desirable to the listener, e.g., 'If you don't throw a tantrum, we'll come back here another time.' The reward 'If you buy this computer, I'll throw in a box of free CDs' becomes a threat when its antecedent is negated and its consequent replaced by anything undesirable to the listener, e.g., 'If you don't buy this computer, I'll tell your wife about our affair.'*

Proposition 4 (From Warnings to Tips) *If $(a_i, \alpha) \longrightarrow (a_j, \beta)$ is a warning, then $(a_i, \neg\alpha) \longrightarrow (a_j, \gamma)$ is a tip for any γ such that $\texttt{Cons}(\gamma, a_j) > 0$.*

Proof. If $(a_i, \alpha) \longrightarrow (a_j, \beta)$ is a warning, then $a_i = L$, $\alpha \in \mathcal{F}(a_i)$, and $a_j \neq S$. The first three criteria for $(a_i, \neg\alpha) \longrightarrow (a_j, \gamma)$ to be a tip are thus satisfied. What remains to be satisfied is the fourth criteria, i.e., $\texttt{Cons}(\gamma, a_j) > 0$. Note that this criterion will be automatically satisfied in the particular case where γ is $\neg\beta$.

Proposition 5 (From Tips to Warnings) *If $(a_i, \alpha) \longrightarrow (a_j, \beta)$ is a tip, then $(a_i, \neg\alpha) \longrightarrow (a_j, \gamma)$ is a warning for any γ such that $\texttt{Cons}(\gamma, a_j) < 0$.*

Proof. Proof is similar to that of Proposition 4, and the same remark holds about the particular case where γ is $\neg\beta$.

Example 6 (Warning to tip, and vice versa) *The warning 'If you open this file, your computer will crash' becomes a tip when its antecedent is negated and its consequent replaced by anything desirable to the listener, e.g., 'If you don't open this file, you can claim you never received it.' The tip 'If you revise the paper, the editor will accept it' becomes a warning when its antecedent is negated and its consequent replaced by anything undesirable to the listener, e.g., 'If you don't revise the paper, your co-authors will think poorly of you.'*

5 The Strength of Non-Explanatory Arguments

It is a standard perspective in argumentation research to assume that arguments differ in strength, or persuasive force. This makes it possible for an agent to compare arguments and select the strongest one. In [3], different definitions are proposed for computing the strength of threats and rewards. These computations are based on the quality of information used to build the arguments. Within this framework, threats and rewards are built from a knowledge base and a base of goals. Thus, the strength of a threat will depend on the *certainty level* of the beliefs used to build that threat, and on the *importance* of the threatened goal. A threat is strong if it invalidates an important goal according to the most certain beliefs. A threat is weaker if it involves beliefs of low certainty, or if it only invalidates a goal of low importance. This framework, however, does not entirely do justice to the complexity of evaluating threats. Even if a threat does target an important goal of the listener and involves highly certain beliefs, other aspects of the situation can make it weak, as suggested by experimental results available in the cognitive psychology literature. Formally:

Definition 6 (Force of a threat) *A threat* $(a_i, \alpha) \longrightarrow (a_j, \beta)$ *is* strong *iff:*

- $\beta \in \mathcal{F}(a_j)$*, and*
- $\mathtt{Cons}(\beta, a_j) \geq 0$*, and*
- $|\mathtt{Cons}(\beta, a_i)|$ - $|\mathtt{Cons}(\alpha, a_j)| \leq \delta$*, where* δ *is a threshold.*

Otherwise, the threat is weak.

The first condition says that the action β should be under the control of the speaker. If it is not, the threat is 'degenerated' [4], and will have little effect on the listener, as empirically shown in, e.g., [18]. One might try to threaten a journal editor to commit her to a psychiatric ward if one's paper is not accepted, but such a threat is unlikely to be taken seriously, as the speaker is unlikely to have such a power. The second condition says that a threat is stronger if action β has a positive side effect for the speaker, or, at least, does not harm the speaker. The sentence 'If you don't behave, we will leave immediately' has more weight if the speaker is a mother looking forward to going home, than if she is a mother who took her child to an important meeting. The third condition is more subtle. It says that the threat should not be disproportionate, i.e., that the punishment should be balanced to the offense if the threat is to be taken seriously. As shown empirically by [25], a proportionate threat such as 'If you tell your brother that Santa does not exist, I'll ground you' is much more efficient than its disproportionate version 'If you tell your brother that Santa does not exist, we will return all your presents to the store.' Remarkably, this result does not hold for rewards, as reflected in the following definition.

Definition 7 (Force of a reward) *A reward* $(a_i, \alpha) \longrightarrow (a_j, \beta)$ *is* strong *iff:*

- $\beta \in \mathcal{F}(a_j)$*, and*
- $\mathtt{Cons}(\beta, a_j) \geq 0$.

Otherwise, the reward is weak.

Just like threats, a reward is strong only if it is indeed in the power of the speaker to deliver the reward β. The reward is even more convincing if the speaker finds a positive side effect in doing β. Unlike threats, rewards do not have to be proportionate, because unlike threats, rewards engage the speaker [4, 12]. As shown in [25], the reward 'If you behave, I'll give you $100' is just as credible that the reward 'If you behave, I'll let you watch a cartoon tonight.' While individuals may think that promising $100 to a child is not very good parenting, they do not question the fact that the parent will stay true to her promise and deliver the $100 if the child does behave. These preliminary results do not preclude the possibility that a limit may exist beyond which a reward is no longer credible (as a function of the speaker's resources? as a function of the listener's assumptions about what a fair reward should be?). Maybe this limit is only much more flexible for rewards than it is for threats—this is still, however, an open empirical question.

Tips and warnings do not seem to have special requirements to be strong. However, as for threats and rewards, a necessary condition for a tip or a warning to be strong is that it is indeed in the power of the third (possibly neutral) agent a_j to take action β. Note that a tip (resp., a warning) might seem even stronger is $\texttt{Cons}(\alpha, a_j) < 0$ (resp., $\texttt{Cons}(\alpha, a_j) > 0$)—that is, when the speaker suggests a course of action that would be beneficial to the listener but detrimental to the speaker herself, or when the speaker warns against a course of action that would be detrimental to the listener but beneficial to the speaker herself. We do not know, however, of any experimental data that would back up this intuition.

6 Related works

This article does not deal with the notion of threat that is pervasive in research on decision under risk, nor with the notion of threat that is involved in engineering applications such as military target analysis, or intrusion detection in computer security. Such applications revolve around evaluating how certain the threat is, and how important its potential consequences. The use of fuzzy logic-based techniques has been proposed for both applications [5, 13, 14]. Rather, in this article we are concerned by the expression of a threat as a *special type of argument,* and how it is perceived by another agent. We are also interested in the dual notion of reward, and in the other duality represented by tips and warnings. For that purpose, we have proposed a formal and abstract framework, grounded in experimental results, in which these four types of arguments are defined, distinguished, and evaluated.

A relevant line of work in artificial intelligence can be found in [12, 15, 21], although the present approach is substantially different. [15, 21] in particular do not study tips and warnings, and do not consider threats and rewards as arguments. Rather, threats and rewards are considered persuasive particles, *speech acts* having preconditions and post-conditions. The preconditions must be satisfied before sending a particle, and the post-conditions represent the consequences of that particle (more precisely, these consequences amounts to adding new beliefs in the listener's beliefs base). A final and important difference between the two approaches is our reliance on empirical data to ground our definitions and validate our assumptions.

Cognitive psychologists have also explored arguments that are kindred to threats, rewards, tips and warnings, and that our framework should easily handle. For example, [7] showed that a statement such as 'If my CEO admits the fraud, her stocks will go down' is perceived as an argument that the CEO will not admit the fraud. The informal definition given for these 'consequential' conditionals can easily be translated into our formal framework as $(a_i, \alpha) \longrightarrow (a_0, \beta)$, where $a_i \in \mathcal{AG} | \{S, L\}$ and $\texttt{Cons}(\beta, a_i) < 0$. Likewise, the 'persuasion' conditionals studied by [23] (e.g., 'If the Kyoto accord is ratified, greenhouse gas emissions will be reduced') can easily be defined in our formal framework.

7 Conclusion

Different types of arguments are exchanged in negotiation dialogues in addition to explanatory arguments. The most common are threats, rewards, warnings and tips. Although there have been attempts at formalizing threats and rewards, no effort has been done at providing a systematic formalization of all four arguments, as well as the criteria to evaluate their strength. We have proposed such a formalization, in which the differences between these arguments are clearly identified, and their persuasive forces are discussed. Furthermore, in a collaborative effort between psychologists and computer scientists, our formal choices have been systematically guided by recent empirical findings from cognitive psychology [4, 9, 10, 16–18, 25]. As a result, our formalization captures exactly what we know of the way human agents exchange threats, rewards, tips, and warnings.

References

1. L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, Volume 34:197–216, 2002.
2. L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In *Proceedings of the 14th European Conference on Artificial Intelligence*, 2000.
3. L. Amgoud and H. Prade. Handling threats, rewards and explanatory arguments in a unified setting. *International Journal Of Intelligent Systems*, 20:1195–1218, 2005.
4. S. Beller, A. Bender, and G. Kuhnmünch. Understanding conditional promises and threats. *Thinking and Reasoning*, 11:209–238, 2005.
5. A. Berrached, M. Beheshti, A. de Korvin, and R. Al. Applying fuzzy relation equations to threat analysis. In *Proc. 35th Annual Hawaii International Conference on System Sciences, Volume 2*, pages 50–54, 2002.
6. P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128:203–235, 2001.
7. J. F. Bonnefon and D. J. Hilton. Consequential conditionals: Invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30:28–37, 2004.

8. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
9. J. S. B. T. Evans. The social and communicative function of conditional statements. *Mind & Society*, 4:97–113, 2005.
10. J. S. B. T. Evans and J. Twyman-Musgrove. Conditional reasoning with inducements and advice. *Cognition*, 69:B11–B16, 1998.
11. S. Fillenbaum. How to do some things with IF. In J. W. Cotton and R. L. Klatzky, editors, *Semantic factors in cognition*, pages 169–231, Hillsdale, NJ, 1978. Erlbaum.
12. M. Guerini and C. Castelfranchi. Promises and threats in persuasion. In *Proceedings of the 6th Workshop on Computational Models of Natural Argument*, 2006.
13. E. Hamed, J. Graham, and A. Elmaghraby. Computer system threat evaluation. In *Proc. 10th International Conference on Intelligent Systems. Washington, DC, International Society for Computers and Their Applications, Raleigh, NC.*, pages 23–26, 2001.
14. E. Hamed, J. Graham, and A. Elmaghraby. Fuzzy threat evaluation in computer security. In *Proc. International Conference on Computers and Their Applications. San Francisco, CA: International Society for Computers and Their Applications, Raleigh, NC*, pages 389–393, 2002.
15. S. Kraus, K. Sycara, and A. Evenchik. Reaching agreements through argumentation: a logical model and implementation. *Journal of Artificial Intelligence*, 104:1–69, 1998.
16. A. López-Rousseau and T. Ketelaar. "If...": Satisficing algorithms for mapping conditional statements onto social domains. *European Journal of Cognitive Psychology*, 16:807–823, 2004.
17. A. López-Rousseau and T. Ketelaar. Juliet: If they do see thee, they will murder thee: A satisficing algorithm for pragmatic conditionals. *Mind & Society*, 5:71–77, 2006.
18. E. Ohm and V. Thompson. Everyday reasoning with inducements and advice. *Thinking and Reasoning*, 10:241–272, 2004.
19. S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261—292, 1998.
20. J. L. Pollock. How to reason defeasibly. *Journal of Artificial Intelligence*, 57:1–42, 1992.
21. S. D. Ramchurn, N. Jennings, and C. Sierra. Persuasive negotiation for autonomous agents: a rhetorical approach. In *IJCAI Workshop on Computational Models of Natural Arguments*, 2003.
22. G. Simari and R. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Journal of Artificial Intelligence*, 53:125–157, 1992.
23. V. A. Thompson, J. S. B. T. Evans, and S. J. Handley. Persuading and dissuading by conditional argument. *Journal of Memory and Language*, 53:238–257, 2005.
24. S. Toulmin, R. Reike, and A. Janik. An introduction to reasoning. *Macmillan Publishing Company, Inc.*, 1979.
25. S. Verbrugge, K. Dieussaert, W. Schaeken, and W. Van Belle. Promise is debt, threat another matter: The effect of credibility on the interpretation of conditional promises and threats. *Canadian Journal of Experimental Psychology*, 58:106–112, 2004.