Mitigation Theory: An Integrated Approach

Bilyana Martinovski (martinovski@ict.usc.edu)
Institute for Creative Technologies, University of Southern California
13274 Fiji Way, Marina del Rey, CA 90292

Wenji Mao (mao@ict.usc.edu)
Institute for Creative Technologies, University of Southern California
13274 Fiji Way, Marina del Rey, CA 90292

Jonathan Gratch (gratch@ict.usc.edu)
Institute for Creative Technologies, University of Southern California
13274 Fiji Way, Marina del Rey, CA 90292

Stacy Marsella (marsella@isi.edu)
Information Science Institute, University of Southern California
4676 Admiralty Way, Marina del Rey, CA 90292

Abstract

The purpose of this paper is to develop a theoretical model of mitigation by integrating cognitive and discourse approaches to appraisal and coping. Mitigation involves strategic, emotional, linguistic, and Theory of Mind processes on different levels of consciousness. We emphasize that discourse analysis can assist our understanding of these processes.

Keywords: Mitigation; Discourse analysis; Attribution; Appraisal; Coping; Cognition.

Introduction

Social judgment processes such as blame and mitigation involve human abilities to perceive and appraise one's own cognitive states and processes as well as the cognitive states and processes of others (Mead, 1993). Mitigation is of particular interest because it is a form of coping with stress and is typically realized in discourse. Spoken language discourse integrates different levels of consciousness (Allwood, 1996; Chafe, 1995) and is one of our few empirical entries to cognition (Edwards, 1997). Mitigation involves argumentation (Toulmin, 1958; van Eemeren and Grootendorst, 2004) thus it is also a way we practice and learn social judgment, which is involved in the creation, maintenance and change of social institutions including the concept of self.

In this paper we develop a theory of mitigation by integrating psychological, cognitive and discourse approaches to responsibility, blame (Shaver, 1985), and coping (Lazarus, 1999). We explore how judgments of blame and defense are realized in institutional discourse such as court trials and what is the relation between the mitigation and the discourse structure. We identify linguistic features, which can be used to recognize mitigation reasoning in discourse.

The paper starts with a description of background studies such as Shaver's model of blame assignment, Lazarus' coping strategies, and linguistic approaches to mitigation. We then proceed with an integrating description of mitigation theory, analysis of data, and a conclusion.

Approaches to Mitigation

Mitigation is a cognitive but also a linguistic and a social phenomenon. It is applied to describe both expressions of politeness and reactions to stressors, such as blame. Below we describe relevant approaches to it, which independently arrive at compatible observations.

Psychological Studies on Blame Attribution

Attribution theory has explored people's assessments of the accountability of social behavior for decades. The attribution approach focuses on the descriptive features of behavior explanation by identifying the broad features people use in determining cause, responsibility and blameworthiness. Among them, the models of Shaver (1985) and Weiner (1995) are the most influential. The judgment process underlying their models relies on several conceptual variables. Take Shaver's model as an example. First one assesses *causality*, distinguishing between personal causality (i.e., human agency) versus impersonal causality (i.e., situational factor). If human agency is involved, the judgment proceeds by assessing whether the actor possessed the *foreknowledge* about the action and its consequence; whether the actor *intended* to produce the action consequence; whether the actor had choices or acted under *coercion*? Causality and coercion determine who is responsible for the outcome, while intention and foreknowledge determine the degree of responsibility assigned. Finally, one takes mitigating circumstances into account. Two kinds of mitigating factors come into play in the judgment process. *Excuse* seeks to

change the variable values in order to effect a judgment result. It is given when one admits to having caused the negative event, but does not accept (or fully accept) responsibility for it. *Justification*, on the other hand, seeks to change the beliefs and values of the accuser. It is offered when one accepts responsibility, but refuses that it was a bad thing to have done, so as to avoid blame, and perhaps gain credit.

Coping

Mitigation can be described in the light of coping (Martinovski and Marsella, 2003, 2005). Appraisal theory (e.g. Scherer et al., 2001) has argued that emotions arise from a person's appraisal of the event. To deal with the resulting emotions, particularly dysphoric emotions, people employ in turn a wide range of coping strategies (Lazarus, 1999). These various strategies can be characterized into several broad classes. Lazarus mentions and elaborates the idea of two main coping strategies in psychological research: problem-focused coping and emotion-focused coping. Problem-focused strategies include taking action (actively addressing the stressor), planning and seeking instrumental support. Emotion-focused coping strategies include suppression, seeking emotional support, restraint, acceptance, turn to religion, denial, and disengagement.

These strategies raise questions on the Theory of Mind processes involved in coping: coping with own emotions/attitudes and coping with other's emotions/attitudes. Martinovski and Marsella's (2003) analysis suggests that the coping process as such is two directional: when faced with stress such as blame people work simultaneously on changing other's mental models of them and on maintaining their own mental model of their own selves.

Linguistic Studies on Mitigation

Within the study of discourse, mitigation has been defined broadly as weakening or downgrading of interactional parameters, which affects allocation and shuffling of rights and obligations (Caffi, 1999), as a way "to ease the anticipated unwelcome effect" (Fraser, 1999, pp. 342) or as the "reduction of vulnerability" (Martinovski, 2000). Discourse mitigation is also distinguished from legal mitigation. In the first case, mitigation is mainly directed to face-work (Brown and Levinson, 1987), whereas in the legal context mitigation is related mainly to defense, credibility and guilt issues.

An attempt to relate concrete verbal behavior, cognition and argumentation in courts is Martinovski's (ibid.) framework for analysis of mitigation. Within this framework there are three main reactions to allegation called defense moves: admission, avoidance (prolepsis), and denial (including counter-attack). These realize different mitigation arguments, such as lack of intention, lack of agency, shared agency (or responsibility), common knowledge, authority, and credibility. In fact, Martinovski's analysis of mitigation in courts is based only on discourse analysis but arrives at categories similar to

Shaver's general psychological variables related to blame (see previous sections).

Mitigation involves argumentation, which is a more general strategy for practice and change of social judgment (Eemeren and Grootendorst, 2004). Both in mitigation and in argumentation, the speaker aims at choosing arguments to alter the hearer's standpoint, but argumentation focuses on the themes of proving the correctness of the conclusion and the logical forms. Argumentation techniques can help mitigating, but in general, the mitigation process is driven by the internal motive of the speaker in avoiding being held blameworthy/guilty for the wrongdoing, rather than the truth of arguments (Goldman, 1997; Martinovski, 2000).

How do we describe the relationship between the concrete linguistic expression in interaction and the cognitive level of appraisal? This is what the next two sections are about.

Integration into a Theory of Mitigation

In order to give a holistic view of the social judgment process, the theoretical model presented here integrates a number of components and distinctions involved in the cognitive process of mitigation[1] such as politeness/defense, potentiality/actuality, Theory of Mind processes, attribution/appraisal, coping/mitigation, cognitive/pragmatic patterns, and social judgment as learning.

To begin with, we need to integrate the two main pragmatic aspects of mitigation: one related to saving face and politeness and the other related to defensive reaction to accusations. We capture these by distinguishing between acceptability (primarily related to face work) and responsibility (primarily related to defense and accusation) (see Figure 1 below), on one hand and between potential and actual triggers of mitigation, on the other.

Mitigation prior to the triggering event can be described as a result of a mentally simulated appraisal processes, which foresees a dis-preferred outcome/s of a future action and offers a mitigated version of that action before it happens. Another way of integrating potentiality and actuality is the distinction between preventive and active coping strategies/discourse moves (see below).

The arguments and coping strategies available might be the same when saving own face and saving other's face but the procedures arriving at such mitigations require different Theory of Mind inferences or simulation processes. When saving face of other one must first establish a model of the other's mind, both actual state and desired state and then relate to that desired state, where the relation will be determined by a number of factors such as relevant memory, emotion, future goals, biological state, and situation. Saving own face involves mental simulation of or inferences about the models others have of one, which is then matched to desired model of self as well as

---

[1] Here we will just touch on each component respecting the space restrictions.
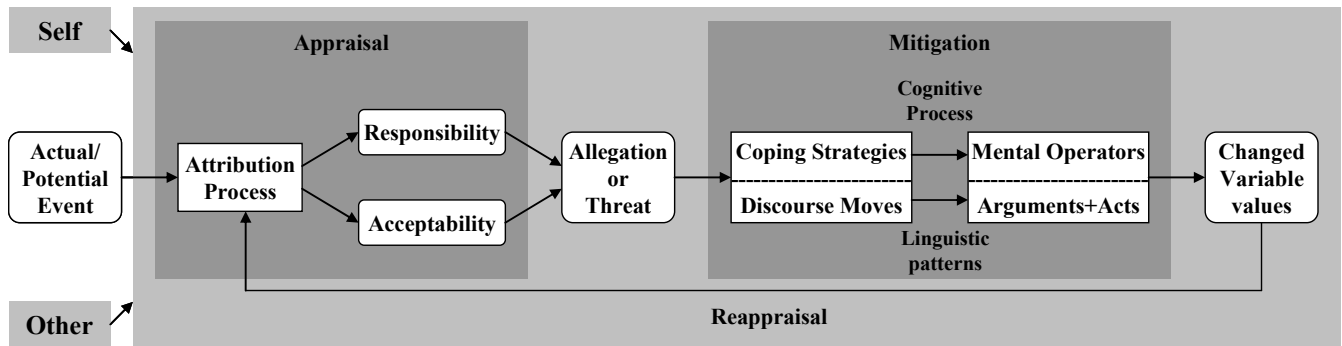
Figure 1 Mitigation Processes

matching of own model of present and historical self with own model of desired self. These cognitive procedures apply also to mitigation as reaction to accusation. As mentioned earlier, they appear to be realized simultaneously on discourse level. These complex relations between self and other are schematically represented in Figure 1, as all-influencing 'self' and 'other' categories.

Furthermore, we need to elaborate on the relationship between attribution theory and the process of appraisal. In our framework, the initial appraisal consists of the cognitive attribution process involved in assignment of blame. It consists of two main criteria, *acceptability and responsibility*. Acceptability of consequence and action refers to local or universal values such as right or wrong, good or bad and thus involve the face work functions of mitigation. Responsibility involves *attribution variables* such as *cause, sharedness, intention, foreseeability, volition, coercion,* and *capacity* (Shaver, 1985; Martinovski, 2000; Martinovski and Marsella, 2003). These are then used in mitigating arguments expressed by justifications and excuses, where justifications concern wrongdoing and excuses concern responsibility. These variables are related to each other, which can be seen even in discourse (see Example E2 in next section).

The outcome of the initial appraisal is a predicted/expected threat (which one thinks other/s might direct to one or other) or an actual allegation (which self or other's direct to one). The appraisal process generates emotions related to it, which influence the outcome (Scherer et al., 2001). Together they contribute to the conscious or not so conscious choice of coping strategy. The classification of coping strategies, such as in Lazarus (1999), corresponds roughly to Martinovski's mitigation moves. We may distinguish between coping by *facing the stressor and dealing with it, coping by avoidance of/preventing the stressor, and coping by acceptance of stressor*. These categories are not exclusive i.e. some coping strategies (see section 'coping' above) such as 'focusing on and venting' can be seen as a combination of both active coping and acceptance of stressor. Prevention is a strategy, which deals with expected or anticipated threats

or allegations. One may prevent by accepting, avoiding or by attacking.

The relation between cognitive and discourse processes is described as a mapping of cognitive and linguistic patterns. The analogy between the discourse moves of mitigation and the psychological strategies for dealing with a stressor helps us to trace the discursive realization of the involved cognitive processes in interaction (see Martinovski and Marsella, 2005). Mental operators are the cognitive correspondents to arguments and communicative acts. Discourse moves correspond to coping strategies, employ much larger context and realize across larger amounts of talk or reasoning whereas communicative acts realize as utterances or as few adjacent utterances. Arguments are linguistic formulations, which use attribution variables for the purpose of mitigation. Besides the acceptability and responsibility attribution variables listed earlier there are also other mitigating arguments such as *reference to commonsense (*or *shared knowledge)* and *authority*. These are used to support reasoning towards the acceptability of consequence or action and to support credibility (see next section). If there is no memory or no trust in the speaker it is unlikely that his/her excuses or justifications will be acceptable no matter what variables they concern.

Communicative acts, mitigation moves and arguments are expressed into concrete linguistic patterns. On the discourse surface we may find recognizable combinations of communication features, such as elliptic clauses, cut-off words, self-repetitions, pauses, lower voice, gestures, modal expressions, etc., which seem to be independent of language, culture and legal system (Martinovski, 2000). For instance, denials tend to be formulated as correction. In contrast, admissions typically consist of confirmations followed by accounts.

Finally, the outcome of the mitigation process consists of changed variables (i.e. change in degree and/or relevance) and these new values/variables go through a reappraisal process, which also influences the emotional coloring of the experience. The whole procedure may be repeated number of times before final consensus, personal

or interpersonal, is reached. The reversibility of the procedure reflects also the learning character of the process.

In the next section, we will illustrate the categories referred to in the theoretical exposition with a few short and representative examples. The multilingual corpus used here consists of trial data and thus deals with already stated allegations[2]. The fact that similar patterns appear in different languages, legal systems and cultures inspires us to think that the mitigation theory we present here is not a local phenomenon.

Discussion and Analysis of Data

Allwood (1976) suggests that mental acts (or operators) are directed at different kinds of objects of consciousness, which in our case are concepts such as intention, agency, cause etc. These objects may be accompanied by different attitudes or emotions, which can be reflected in the information structure of e.g. an utterance or in the voice or gesture. The background of these acts may consist of different degrees of Theory of Mind models of self and other. There can be different levels of consciousness, which can also be reflected in the linguistic realization of discourse, in the speech order, the tone, the gesture. Spoken language discourse is multimodal and thus invite realization on different levels of awareness or intentionality, which is reflected in Allwood's distinction between indication (when we convey information without intending to do so), display (when we intend to show something to somebody) and signal (when we signal that we are displaying something). These levels of consciousness can be traced in discourse (Edwards, 1997). Allwood distinguishes also between different types of agent of consciousness, namely individual or collective. In that sense, dialogue can be seen as a tool for collective thinking and focus.

Mitigation in courts is, in this context, a process of collective re-evaluation and re-appraisal. In interaction all the coping strategies and the mental operators can be realized simultaneously. Lets look at some examples.

Avoidance, common knowledge, and admission

In the following example the Prosecutor (P) tries to establish if the Defendant (D) has been drinking the night of the crime; in case he has the final sentence increases in severity (J stands for judge, Pl – for plaintiff, DC – defense counsel, / stands for pause, <> defines the scope of the described feature, (…) stands for inaudiable speech, capital letters for emphatics, + for cut-offs, : stands for prolonged speech).

Example E1:
1. P: but that particular evening afternoon evening you had you been drinking alcohol then /

2. D: < (yes) beer I mostly run on beer> <quiet>

The evasive answer on the second line[3] above refers to a general state of affairs thus display desire to avoid acceptance of guilt but at the same time providing generally but not necessarily specific truthful information. The mitigation argument is *commonsense or common knowledge* i.e. speaker displays or signals that what he is speaking about is a principle fact or a matter-of-fact or a belief shared by the members of the socio-cultural community to which he belongs and is thus understandable to 'everyone', including the participants in the discursive activity. It is presented by the idiomatic metaphorical expression 'run on' (i.e. he runs on beer as cars run on gas) and formulated as a correction with topicalization of the mitigating evidence, i.e. he drank beer not harder alcohol. The entire speech act of admission is uttered in a very low voice and the initial confirmation 'yes' is almost inaudible, which indicates the desire to avoid the admission of blame. Line two above realizes the coping strategies of avoidance, acceptance, and countering at the same time where acceptance is most prominently expressed although mitigated. Of course, there are complex Theory of Mind processes involved too (see Martinovski and Marsella, 2003 for detail).

Countering, no-agency/no-intention, and correction

In the next example we have the discourse move of countering realized with no-agency and no-intention arguments presented as a correction (as a speech act). The *no-agency mitigation* accepts action but denies or silences agency. It is typically realized by topicalization and emphatics (indicated below with capitals). In the example below the defense counsel (DC) puts an inference-based declarative question, which indicates agency on the behalf of the witness. In his answer the witness rebuts this inference by topicalizing and emphasizing the lack of agency.

Example E2:
1. DC: so you moved towards Nilsson then
2. Pl: THE DOOR moved towards Nilsson I moved inside inside the car /

This example shows the close relation between *no-agency* and *no-intention mitigations*: correcting the ascription of agency he distances himself from the ascription of intention of hitting 'Nilsson' with the door of the car. Intentionality is understood differently in different cultures and legal systems. Having no intentions to do wrong is an acceptable justification in Roman law.

On discourse level, no-agency arguments are typically formulated as corrections and counters i.e. as mitigated forms of denial, which corresponds to Shaver's analysis of justification. In contrast, no-intention mitigations are introduced suggestively after the admission of the alleged

---

[2] Here we will use only few examples from the Swedish data. Martinovski (2000, 2003, 2005) uses also Bulgarian and English data.

[3] This is the last instance of a long sequence of evasive answers by the same witness on the topic of his drinking 'that night'.

action, which corresponds to Shaver's analysis of excuses i.e. even on the linguistic surface the responsibility/causality is dealt first with.

Acceptance, shared responsibility/no-intention and admission

In the next example we have mitigation through reference to s*hared responsibility* (i.e. Shaver's multiple causality) or partial shift of blame, which is also a coping strategy associated more with active dealing with stressor than with acceptance or avoidance of stressor. After the acceptance of wrongdoing follows mitigation by contesting aspects and degrees of responsibility using the attribution variables. That is, we have two coping strategies realized in the same turn. The judge (J) is explicitly requesting the defendant's consent or admission (line 1).

Example E3:
1. J: alright what is it said / there than
2. D: yes it is like this of course that Bengt felt he he was with me when I made the deal so then he <pay+> paid too a certain amount of <mon+> money didn't he / and then it was this that we we register the car on me eh eh what it <de+> depended on other circumstances / and eh / that I made myself guilty of this is is OK I think I I have made a mistake then

The elicited admission starts with a positive feedback item, which serves as a receiver of the given turn. This item is followed by a narrative, which ends with an explicit declaration of the speaker's understanding of the wrongdoing and approval of the accusation. However, the narrative refers to co-authors of the crime and by the use of the modal expression 'of course' the defendant signals the self-given or known character of the information he gives. The defendant minimizes his own guilt before presenting the admission in this way initiating negotiation of guilt and at the same time acting cooperatively.

See a similar pattern in another extract:

Example E4:
1. DC: against / Nilsson then
2. Pl: <yes exactly> he WAS STANDING in front of the door of the car <quiet>

Here the defense counsel alleges on line one that the plaintiff has hit intentionally the defendant by opening aggressively the door of a car. On line two the plaintiff responds with the same pattern but uses a different argument, namely no-intention and instead of topicalization we have emphatics.

Credibility

Credibility and certainty are two variables, which are not listed in Shaver but are very productive value shifters in mitigation. They are often associated with arguments such as reference to authority, references to memory and common knowledge. The witness in the next example has felt

the pressure of accusations with regard to his sobriety during the actual day of the crime. He has already used number of mitigating strategies but the examiner still formulates his inferential declarative question (a very suggestive formulation, thus we are not surprised he talks to his own client) on line one in a way, which implies that the witness is often drunk or drinking.

Example E5:
1. P: so you were after all in somewhat good CONDITION you mean
2. Pl: YES: / I was of course // after all the police THOUGHT so when i was at the police station

The defendant is eager to be believed which he displayed, indicated and probably even signaled by the emphatic initial confirmation, modal expression of known states-of-affairs 'of course' (Swedish 'ju') and after a pause a reference to the indisputable statement of the authority. All these features realize his desire to mitigate the perceived accusation of non-credibility, which he infers from the formulation of the previous inference.

In short, admission in courts have the following reoccurring pattern: confirmation item (e.g. 'yes') + account (justification, excuse, narrative) + mitigation items (e.g., voice, hesitation, self-repetition, pauses) + admission. The general preferred pattern realizing coping dominated by avoidance and acceptance is:

confirmation + account [arguments:
　　　　　　　　commonsense
　　　　　　　　authority
　　　　　　　　shared responsibility…]

A preferred pattern for actively dealing with stressor or countering is:

No confirmation, direct correction with emphasis or topicalization [argument: no intention
　　　　　　　　　　no agency…]

The data analysis suggests that the procedural cognitive strategies and arguments are realized in dialogue simultaneously in the same utterance and/or turn, which indicates that cognitive processes are either not organized linearly and/or that the speed of processing is high in comparison to speed of speech.

Conclusion

The integrated theory of mitigation presented here described mitigation as a dynamic appraisal process, in which cognitive and linguistic procedures and variables are mapped into each other. The dialogue functions as a tool for collective thinking, re-evaluation, and focus. Mitigation involves both face saving and reaction accusation processes. The predicted or real result of appraisal activates coping strategies, which are influenced by self and other models and goals. These coping strategies in-

volve a set of attribution variables such as cause, intention, credibility etc. which are then realized in recognizable patterns of discursive moves, arguments and communicative acts. The model describes how the meaning of these concepts and the institutions that sustain them is publicly negotiated.

In the future, we intend to analyze more data from different social contexts and explore further discourse patterns. One area of application for the theory is to inform the formal modeling and simulation of human behavior in agent-based systems, where it would mediate agent interactions. Specifically, it could be incorporated within virtual humans, software agents that look like, act like and interact with humans within a virtual world (Rickel et al., 2002). Such incorporation will create a bridge between natural language/dialogue module and task planning/emotion module (Gratch & Marsella, 2004; Mao & Gratch, 2004).

References

Allwood, J. (1976). *Linguistic Communication as Action and Cooperation.* Gothenburg Monographs in Linguistics 2, University of Goteborg, Department of Linguistics.

Allwood, J. (1996). Some comments on Wallace Chafe's "How Consciousness Shapes Language". *Pragmatics and Cognition,* 4(1).

Brown, P. and Levison, S. C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge University Press.

Caffi, C. (1999). On Mitigation. *Journal of Pragmatics*, 31.

Edwards, D. (1997). *Discourse and Cognition*. London: SAGE Publications.

van Eemeren, F.H. and Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The Pragma-Dialected Approach*. Cambridge University Press.

Fraser, B. (1999). Mitigation. *Journal of Pragmatics*, 31.

Goldman, A. I. (1997). Argumentation and Impersonal Justification. *Argumentation*, 11, 155-164.

Gratch, J. and Marsella, S. (2004). A Domain-Independent Framework for Modeling Emotion. *Journal of Cognitive Systems Research*, 5(4), 269-306.

Lazarus, R. S. (1999). *Stress and Emotion*. New York: Springer Publishing Company, Inc.

Mao, W. and J. Gratch (2004). Social Judgment in Multi-agent Interactions. *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*.

Martinovski, B. (2000). *The Role of Repetitions and Reformulations in Court Proceedings: A Comparison between Sweden and Bulgaria*. Ph. D. Thesis. Goteborg University Press.

Martinovski, B. and Marsella, S. (2003). Coping processes in discourse: dynamic reconstruction of selfhood. *Proceeding of the 4th International Conference on Cognitive Science*, Sydney, Australia.

Martinovski, B. and Marsella, S. (2005). Theory of Mind and Coping in Discourse. *Artificial Intelligence and the Simulation of Behavior (AISB) 2005 Symposia on Mind-Minding Agents*.

Mead, G.H. (1993). *Mind, self, and society*. Chicago: University of Chicago Press.

Rickel, S. Marsella, J. Gratch, R. Hill, D. Traum, and W. Swartout. (2002). Toward a New Generation of Virtual Humans for Interactive Experiences. *IEEE Intelligent Systems,* 17(4), 32-38.

Scherer, K. R., Schorr, A., and Johnstone, T. (2001). *Appraisal Processes in Emotion*. Oxford University Press.

Shaver, K. G. (1985). *The Attribution of Blame: Causality, Responsibility and Blameworthiness*. Springer-Verlag.

Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge University press.

Weiner, B. (1995). *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. The Guilford Press.