

How context dependent are potentially face-damaging actions? Effects on rapport and response production

Sara Amido

Universitat Pompeu Fabra
sara.amido@upf.edu

Vladislav Maraev

University of Gothenburg
vladislav.maraev@gu.se

Christine Howes

University of Gothenburg
christine.howes@gu.se

Abstract

Previous corpus research identified three conversational actions that recurrently resulted in face damage in confrontational contexts. We experimentally inserted these actions into text-based conversations to investigate whether their effects persist outside their original context, assuming these would be reflected in participants' rapport evaluations and conversational behaviour. Although the interventions did not significantly affect rapport, responses to interventions were longer and contained more subjectivity markers. The mixed results thus do not fully corroborate context-independent face-damaging effects of these actions.

1 Introduction

The notion of *face* is central to studying different tones of interaction, such as confrontational versus non-confrontational conversation. Face is the mental representation of how an individual is perceived publicly (Brown and Levinson, 1987; Goffman, 1955). It can be defined in terms of wants: the want to be approved by others and the want to be free to act as one wishes (Brown and Levinson, 1987, 61). Face effects (damage and enhancement) happen as a result of actions in conversation and are thus shaped in interaction (e.g., Eelen 2001; Mills 2003; Watts 2003, among others). When actions impede face wants, this damages face; when actions support face wants, this enhances face.

Previous corpus work on confrontational conversations from the *The Only Way is Essex* (TOWIE) reality TV series identified three candidate actions which recurrently resulted in face damage (Amido, 2025)¹. However, because these actions occurred in confrontational contexts, it remains unclear whether their effects are observable across

other contexts. We therefore ask how context dependent are the face-damaging effects of these actions, as reflected by perceptions of interpersonal rapport and subsequent conversational behaviour across contexts.

Because face damage concerns participants' face wants, we hypothesised that repeated face-damaging actions would negatively affect participants' global evaluations of the relationship with their interlocutor, operationalised here as rapport. Likewise, we assumed face-damaging effects of these actions would be observable through conversational behaviour: recipients would respond by escalating the conversation through counter-attacks or by de-escalating through hedging their response.

2 Method

Participants and task 40 participants were paired up and prompted to discuss a moral dilemma via chat on Telegram for 20 minutes.² All participants subsequently completed a survey scoring rapport with their conversational partner based on the chat (see Appendix A). Additionally, participants were given the option to provide a written justification for each score.

Manipulations For the 10 pairs in the experimental condition, interventions representing three action types were inserted into the chat via the Dialogue Experimental Toolkit (DiET) chat tool, a text-based chat interface that allows real-time experimental interventions (Healey et al., 2003)³. Interventions were produced through participant-blind manipulation so they were visible only to their recipient. The three action types were not equally frequent because their production conditions differed and interventions were generated automatically when those conditions were met. No in-

¹TOWIE corpus https://osf.io/6npj/overview?view_only=ca565bcabf3d45baae0431be108c00ba and corpus study https://osf.io/a2k6p/overview?view_only=eb85317832504c0680e757b72077bdd9.

²Data and materials https://osf.io/3en4f/overview?view_only=1460ea3646674850a055a25ef32910db.

³<https://dialoguetoolkit.github.io/chattool/>

interventions were produced in the control condition. Criteria, examples, and operationalisation of interventions for the three potentially face-damaging actions follow.

Addressee-centred statements (1)⁴ are when a speaker makes a direct unhedged statement about the addressee and it is the addressee who has epistemic authority on the matter (Heritage and Raymond, 2005). The intervention changes “I” to “you”; the production condition is that the turn contains “I”. Metacommunicative directives (2) are directives such as commands or requests that serve to manage how the conversation unfolds, such as directing the addressee’s attention, rather than contributing to its topical content. The intervention adds “Listen,” utterance-initially; the production condition is that the turn does not end with “?”. Address terms (3) are when a speaker uses the addressee’s name at the end of the utterance, not for turn-management purposes. The intervention adds “, p1” or “, p2” (for the sake of anonymity) utterance-finally; the production condition is that the turn does not end with “?”.

- (1) Meg: I react by screaming and shouting. Like, I can’t help it. It’s who I am.
Chloe: Okay, well. **You can help it. You can.**
Meg: Well, I can’t! It’s the way I am when I’ve been hurt!
- (2) Yaz: **Now let’s be honest**, you message me as well.
Lockie: **Listen.** No no no, **listen.**
- (3) Lockie: Me and you just didn’t work out, **Yaz.**
Yaz: I know that, you haven’t got to tell me that!

3 Results

Rapport scores The analysis of rapport scores for each question and overall revealed no significant differences between the experimental and control groups (see Table 1, Appendix B).

Responses to interventions There were no high level differences between the control and experimental dialogues in terms of number or average of turns or words (see Table 2, Appendix C). We

therefore examined whether interventions had local effects on responses to turns containing them. We treated the next turn produced by the other participant as a response to the intervention (despite the limitations, see Section 5). Turns which were a response to an intervention contained more words than average (see Table 3, Appendix C). T-tests confirm this, with response turns containing more words than both other turns in the experimental condition ($t(103.44) = 3.66, p < 0.001, d = 0.52$) and turns in the control condition ($t(116.48) = 2.55, p = 0.012, d = 0.29$). A visual inspection of the data suggested that these differences might be driven by participants doing more face work in the form of hedging in turns following an intervention. We therefore used simple pattern matching to annotate turns which contained subjectivity markers such as “I feel” and “I mean”. A chi-squared test showed a significant difference, such that people were nearly twice as likely to use these phrases in responses to interventions than in the control dialogues or other turns in the experimental condition ($\chi^2_{(2,1188)} = 7.43, p = 0.024, V = 0.08$).

Participant feedback In the optional feedback, two participants referred to the interventions specifically. Participant 5 justified low rapport scores referring to their interlocutor’s use of “p1” describing it as “strange” and “not really respectful”, and also referred to their use of “listen”. Participant 10 justified low rapport scores referring to their interlocutor’s use of “p2” and “listen”, associating these with a “bossy” and “disrespectful” attitude.

4 Discussion

The interventions did not result in a significant effect on the overall rapport scores, but were associated with local changes in immediate response production. Both patterns observed in the responses are compatible with recipients managing face-damaging actions and de-escalating the conversation: longer responses may provide space for hedging response prefaces (Pomerantz 1984; Sacks 1987), while subjectivity markers can weaken the speaker’s commitment to a claim and/or not assume their interlocutor shares the same views (Brown and Levinson, 1987). These mixed results do not fully support context-independent face-damaging effects of these actions. Future work should assess whether response analysis can robustly capture face damage in interaction and whether these actions elicit similar response patterns across contexts.

⁴Examples are adapted from the TOWIE corpus and potentially face-damaging actions are in boldface.

5 Limitations

One main methodological limitation of this study is that rapport perceptions are not a direct measure of face damage incurred throughout a dialogue. Several factors may have confounded scores such as task duration, interlocutor engagement (e.g., delayed responses), diverging moral values, and diverging attitude (serious or playful) towards the task. Another limitation of the rapport evaluations is that participants were recruited within the same university campus and so a potential compromise in anonymity may have influenced the scores.

Other limitations in this study concern the implementation of interventions, like automated intervention production resulting in infelicitous utterances. In addition, using “p1/p2” likely had a different effect from using participants’ names. In addition, the chat modality of the task differed from the corpus’ multimodal data on which the study was based so the effect of the conversational actions may not be comparable. Furthermore, each intervention type may have had a distinct effect and the occurrences of each were not equally distributed as seen in Table 4 (Appendix C).

Regarding the response length analysis, we assumed next other user turn to be the response to the previous turn. Since turn production and receipt is necessarily asynchronous in text based chat, this means that some of the included turns may not have been direct responses to the intervention.

References

- Sara Amido. 2025. *You’re offended, I’m offended: A face-based analysis of confrontational conversation*. Ph.D. thesis, Universitat Pompeu Fabra.
- Penelope Brown and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press.
- Gino Eelen. 2001. *A Critique of Politeness Theories*. St. Jerome Publishing, Manchester.
- Erving Goffman. 1955. On face-work: An analysis of ritual elements in social interaction. *Psychiatry*, 18(3):213–231.
- Jonathan Gratch and Sin-Hwa Kang. 2015. [Gratch lab rapport measures](#). online.
- P. G. T. Healey, Matthew Purver, James King, Jonathan Ginzburg, and Greg Mills. 2003. Experimenting with clarification in dialogue. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Boston, Massachusetts.
- John Heritage and Geoffrey T. Raymond. 2005. The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly*.
- Sara Mills. 2003. *Gender and Politeness*. Cambridge University Press, Cambridge.
- Anita Pomerantz. 1984. Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In J. Maxwell Atkinson and John Heritage, editors, *Structures of Social Action: Studies in Conversation Analysis*, Studies in Emotion and Social Interaction, pages 57–101. Cambridge University Press, Cambridge.
- Harvey Sacks. 1987. [On the preferences for agreement and contiguity in sequences in conversation](#). In Graham Button and J. R. E. Lee, editors, *Talk and Social Organisation*, pages 54–69. Multilingual Matters, Clevedon.
- Richard Watts. 2003. *Politeness*. Cambridge University Press, Cambridge.

A Rapport survey

Rapport survey ([Gratch and Kang, 2015](#))

1. I felt I had a connection with the listener.
2. I think that we understood each other.
3. The listener was warm and caring.
4. The listener was respectful to me.
5. I felt I had no connection with the listener. (reverse coded)
6. The listener created a sense of closeness or camaraderie between us.
7. The listener created a sense of distance between us. (reverse coded)
8. The listener communicated coldness rather than warmth. (reverse coded)
9. I wanted to maintain a sense of distance between us. (reverse coded)
10. I tried to create a sense of closeness or camaraderie between us.
11. I tried to communicate coldness rather than warmth. (reverse coded)

Participants evaluated how strongly they dis/agreed with the 11 statements on a 5 point Likert scale.

B Rapport scores

Question	Experimental		Control	
	mean	s.d.	mean	s.d.
1	3.25	1.16	3.20	1.06
2	3.65	1.23	3.95	1.00
3	3.10	1.17	3.25	1.02
4	4.20	1.24	4.70	0.57
5 (RC)	3.60	1.43	3.65	1.23
6	3.05	1.15	3.15	0.88
7 (RC)	3.45	1.23	3.65	1.09
8 (RC)	3.55	1.28	3.60	1.14
9 (RC)	4.05	1.10	3.60	1.14
10	3.60	0.94	3.65	0.88
11 (RC)	4.45	0.76	4.20	1.06

Table 1: Rapport scores: means and standard deviations.

C Response length

Group	Turns, total	Turns, avg. per participant	Words, avg. per turn
Control	571	28.55	10.56
Experimental	617	30.85	9.95

Table 2: Number of turns and average turns and words per participant by condition.

Group	N	mean	s.d.	% <i>I feel</i>
Control	571	10.56	10.83	11.2
Experimental	529	9.33	7.88	13.2
I-Responses	88	13.67	10.65	21.6

Table 3: Number of words in turns. *I-Responses* are turns which respond to or follow an intervention; *Experimental* is all other turns produced by participants in the experimental condition (including original turns on which an intervention was performed); *I feel* are turns containing the subjectivity markers “I feel”, “I mean”, “I think”, “I guess”, “I believe”, “personally”, “to/for me” and “in my opinion”.

Intervention	#	Avg. per participant	Words in response turn, mean
<i>You</i> -statements	13	0.65	15.15
<i>Listen</i> ,	34	1.7	14.21
<i>,p1/p2</i>	41	2.05	12.76

Table 4: Number of interventions and mean number of words in turns that are a response to an intervention.