

Why Digital Exposure Can Backfire: A Contact-Architecture Framework for Social Media and AI

Isabella Chau
ichau27@priorypanther.com

Keywords / Index Terms

Intergroup contact, digital contact, algorithmic bias, artificial intelligence, prejudice reduction

1. Introduction

1.1 Background

The algorithms and artificial intelligence programs behind social media platforms influence which social groups people interact with digitally, as well as how such groups are represented. Although contact theory suggests that these encounters can mitigate prejudice, digital interaction may instead worsen stereotypes, deepen divisions, and silence marginalized voices.

1.2 Problem

In this paper, the core question addressed is: What design elements differentiate digital contacts that combat prejudice from those that perpetuate it? This research proposes a contact-architecture approach, taking psychological literature on contact theory and applying it to technological algorithms.

2. Investigation

2.1 Methodology

In this systematic review, seven focal sources were analyzed using narrative analysis, selected from a larger pool of forty papers [1]–[7]. The evidence set included two meta-analyses of intergroup contact, a theoretical assessment, a social media field experiment, a

language-model bias audit, a meta-analysis of digital contact, and a controlled AI-facilitated contact experiment. The sources provided quantitative evidence on prejudice, its mediators and mechanisms, and technologically mediated interaction.

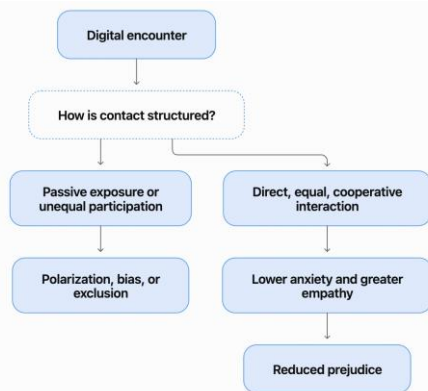
Information extracted from every source included the design of the study, the sample size, the context of contact, the intensity of participant engagement, the psychological variables involved, and the presence of four established contact conditions: equal status, cooperative goals, institutional support, and repeated meaningful contact. Interactions mediated via technological devices were further subdivided into passive and structured direct contact. More emphasis was placed on studies involving controlled experiments than on those based on correlational analysis. Finally, patterns identified across the different sources were condensed into Table 1 by translating each factor into corresponding requirements applicable to online platforms.

2.2 Results

The connection between intergroup contact and decreased levels of prejudice has been demonstrated across multiple analyses, including 515 separate studies and 713 individual sample groups. In addition, anxiety

reduction and increased empathy and perspective-taking were shown to mediate this effect. In relation to technology use, digital evidence showed distinctions based on mere exposure versus active interaction. For instance, digital contact decreased prejudice among a sample group of 9,385 participants across 88 separate studies (effect sizes $d = 0-0.25$). Exposure to conflicting viewpoints led to high levels of polarization in one study of 1,652 individuals, while another study found that a language-model AI reproduced biases against minority dialects. On the other hand, a mediated AI contact experiment lowered interracial prejudice and anxiety among 128 participants compared with non-mediated control groups. Ultimately, technology's effects depended less on exposure itself than on how the encounter was structured.

Figure 1. How contact architecture shapes digital outcomes



3. Conclusion

3.1 Related Work

Past research has demonstrated both general and digital contact effects [1], [4]. This study extends that work by converting psychological conditions into platform design requirements and measurable outcomes.

3.2 Limitations

Because this was not a preregistered systematic review, source selection may be incomplete, and the proposed framework has not yet been directly tested across platforms or demographic groups.

3.3 Societal Impact

This framework could potentially assist developers, schools, and Bay Area institutions in differentiating meaningful contact from exposure that may worsen prejudice. Balanced participation and consistent moderation of platforms could benefit marginalized users, while poorly designed systems may amplify harassment, stereotypes, or exclusion. The framework should therefore be tested for effectiveness, accessibility, privacy, and unequal outcomes before implementation.

Table 1. Proposed solution: contact conditions translated into testable platform requirements.

Evidence-based condition	Technical or policy requirement	Measurable outcome
Equal status	Balance visibility and speaking power	Participation gaps, exposure diversity
Shared goals and cooperation	Build collaborative tasks, not solo debate	Reciprocity, task completion, empathy
Institutional support	Publish rules and enforce moderation consistently	Harassment, safety, appeal outcomes
Repeated meaningful contact	Support persistent groups and follow-up	Trust and attitude durability

4. References and Additional Details

4.1 References

- [1] T. F. Pettigrew and L. R. Tropp, “A meta-analytic test of intergroup contact theory,” *Journal of Personality and Social Psychology*, vol. 90, no. 5, pp. 751–783, 2006, doi: 10.1037/0022-3514.90.5.751.
- [2] T. F. Pettigrew and L. R. Tropp, “How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators,” *European Journal of Social Psychology*, vol. 38, no. 6, pp. 922–934, 2008, doi: 10.1002/ejsp.504.
- [3] E. L. Paluck, S. A. Green, and D. P. Green, “The contact hypothesis re-evaluated,” *Behavioural Public Policy*, vol. 3, no. 2, pp. 129–158, 2019, doi: 10.1017/bpp.2018.25.
- [4] L. P. da Costa, K. Bierwiazek, and M. Bianchi, “Does digital intergroup contact reduce prejudice? A meta-analysis,” *Cyberpsychology, Behavior, and Social Networking*, vol. 27, no. 7, pp. 440–451, 2024, doi: 10.1089/cyber.2023.0591.
- [5] C. A. Bail et al., “Exposure to opposing views on social media can increase political polarization,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 37, pp. 9216–9221, 2018, doi: 10.1073/pnas.1804840115.
- [6] V. Hofmann, P. R. Kalluri, D. Jurafsky, and S. King, “AI generates covertly racist decisions about people based on their dialect,” *Nature*, vol. 633, pp. 147–154, 2024, doi: 10.1038/s41586-024-07856-5.
- [7] S. Sahab et al., “E-contact facilitated by conversational agents reduces interethnic prejudice and anxiety in Afghanistan,” *Communications Psychology*, vol. 2, art. no. 22, 2024, doi: 10.1038/s44271-024-00070-z.