 DOI: https://doi.org/10.12345/xdjyjz.v4i2.36148

The Psychological Mechanism of Extreme Online Speeches under the Framework of Social Identity-Based on the Theoretical Model of Strategic Expression Motivation and Group Interaction

Tianya Chen

Purdue University West Lafayette, Hangzhou, Zhejiang, 310000, China

Abstract

The radicalization of online discourse is evolving from sporadic individual expressions into a normalized phenomenon amplified through group interactions. While existing research primarily examines this phenomenon through platform algorithmic mechanisms, it fails to explain why individuals still opt for high-intensity expressions despite the absence of direct conflicts of interest. Building on social identity theory, this study constructs a six-stage theoretical model encompassing "social identity—strategic expression motivation—group interaction," demonstrating that radicalization results not from intensified attitudes but from low-risk expression strategies rationally chosen in visibility competition. The model further differentiates the divergent pathways of KOL-led and decentralized group interaction structures operating under identical psychological mechanisms. This research provides a theoretical foundation for subsequent quantitative studies and intervention design.

Keywords

Social identity; Strategic expression motivation; Online speech radicalization; Group interaction; Cognitive load

社会认同框架下网络言论极端化的心理机制——基于策略性表达动机与群体互动的理论模型

陈天雅

普渡大学，中国 · 浙江 杭州 310000

摘　要

网络言论极端化正从个体的偶发表达演变成可以被群体互动不断放大的表达常态。现有研究多从平台算法机制的角度去解释该现象，但难以解释为何在缺乏直接利益冲突的情况下，个体仍然选择高强度的表达方式。本研究基于社会认同理论，构建了一个包含“社会认同—策略性表达动机—群体互动”六阶段理论模型，尝试解释极端化并非态度激化的结果，而是在可见性竞争中被理性选择的低风险的表达策略。模型进一步区分了KOL引导型与去中心化群体互动结构在相同心理机制下的差异化作用路径。本文为后续定量研究与干预手段的设计提供了理论基础。

关键词

社会认同；策略性表达动机；网络言论极端化；群体互动；认知负荷

1 引言

网络言论极端化可以从两方面进行解释，一是态度的极端化，即个体在认知和情感上的极端立场；二是表达策略的极端化，即个体如何在与社会互动中主动选择极端化的表达。

既有研究多从信息茧房的角度去解释网络言论极端化的成因。但在 Christopher A. Bail 所设计的“走出回声室”的实验中显示，暴露于对立信息并不必然缓和立场，反而在许多情况下立场更坚定或更极端 (Bail, 2018)。这表明，认知茧房在情绪化表达、信息过滤、群体极化等理论在阐述极端化现象时的局限性。

因此本文将建构一个包含”KOL- 追随者 - 群体内互动”的模型，分析不同互动结构下，策略性表达动机在网络言论极端化的过程中发挥的作用，并回答以下 4 个关键问题：

为什么在缺乏直接利益冲突的情况下，个体仍会选择高攻击性的表达？

为什么极端表达会在某些群体中成为常态，而非边缘现象？

为什么理性推理在这些场景中失效，甚至主动回避？

为什么在符合社会朴素正义的价值观的场景，仍然会演化出高度对抗性？

【作者简介】李芹芹（1995—），女，中国浙江杭州人，本科，从事社会与应用心理学研究。

2 理论框架

2.1 社会认同理论框架下的极端化

社会认同理论（Social Identity Theory，SIT）认为社会认同由类化、认同和比较三个过程组成。个体倾向于通过群体分类将世界分为“我们”与“别人”，并在群体比较的过程中追求积极的社会认同 (Tajfel, 2001)。

在 SIT 理论框架下，群体认同通过影响个体的态度和立场发挥作用，往往表现为态度鲜明、非黑即白的二元立场。个体会更倾向于采纳符合群体规范的态度以维持积极的社会认同和自我评价，从而解释为什么群体内会出现内群体偏好和外群体贬抑的现象。这一理论框架能够部分解释为什么即便不存在直接的现实利益冲突（如对稀缺资源的争夺）的情况下，群体内仍出现稳定的对立结构。

然而，SIT 理论的分析重点集中在群体态度的形成；对于个体在互动的过程中如何通过不断增强表达强度来表明立场，解释相对有限。这种个体从态度一致到表达极端化的问题，是本文进一步讨论的核心问题。

2.2 群体互动：从态度到表达

在网络互动的情境中，个体的态度并非总是发生显著变化，但表达方式却可能呈现系统性的强度升级，最终导致极端化。这一现象表明，“态度是否极端”和“表达是否极端”并非同一问题。个体的对表达方式的策略性选择会受到可见性、反馈与群体回应的持续影响，也会受到社会认同、压力、认知负荷等影响。基于此，本文将引入策略性表达动机的心理机制模型，对表达强度的升级的过程进行分阶段分析。

2.3 策略性表达动机导致极端化的阶段性心理机制模型

2.3.1 第一阶段：社会认同显著性提升

当个体进入高度社会化的网络互动情境时，个体首先经历社会认同显著性的提升。个体开始以群体身份来定义自我，比如“女性”、“某种 MBTI”、“某某粉丝”，这种定义自我的方式使得个体能够快速将世界分成“我们”与“别人”。在这一阶段表达开始承担起确认归属、强化边界的功能。说什么、怎么说不再只是观点选择，而是一种传递是否是“自己人”的信号。

2.3.2 第二阶段：群体规范敏感性增强

随着互动的持续，个体意识到不同的说法会收到不同的反馈，包括回应、质疑、忽视。这时候表达的动机就从“表达态度”逐渐转化成了“如何表达更容易被看见的表达”。从心理机制上看，这一阶段体现了个体对群体规范的敏感性增强。

群体规范并非以明文规定、白纸黑字的形式出现，而是互动反馈中被学习和内化。群体性事件在线互动中情绪参与反馈机制对于表达强度的提升起到显著影响 (王思尹，2024)。在认知资源有限的情境下，个体倾向于以最小认知努力完成判断任务，从而偏好信息明确、处理成本较低的选项 (Fiske, 2020)。因此跟随群体规范被视为一种降低认知负荷的心理策略。

2.3.3 第三阶段：群体规范稳定后的可见性需求与表达动机改变

当群体中的表达规范稳定之后，个体之间的态度与表达方式差异缩小。在这时候，与群体保持一致的表达，已不足以获得回应和注意，个体面临着可见性下降的心理压力。为了缓解被忽视的不确定性焦虑，个体就会倾向于通过提高表达的强度来制造差异，从而恢复可见性。

2.3.4 第四阶段：表达的工具化

当个体逐渐认知到，不同的表达会带来不同的互动结果时，表达的目的就从单纯的观点表达转为了结果导向。个体在表达之前，会隐约地有这种表达方式被注意、回应、反馈的心理预期，并据此调整表达方式与强度。从而把表达变成一种可控的、可被策略性调节的行为手段。这是一种“预期—校准”的过程。

2.3.5 第五阶段：极端化表达的策略性选择与互动强化

在表达被当作一种实现结果的工具后，个体仍面临选择何种表达强度的问题。在这种情境下，个体会表现出明显的风险厌恶的倾向。相较于中立的态度、温和的表达，高强度的表达更加明晰地确认边界，从而降低被误解的可能。从决策心理的角度看，这一选择并非追求表达效果的最优解，而更接近与一种“齐当别”决策，即把一个多维度的复杂比较，简化成单一维度的辨别过程 (李纾 , 2016)。该选择同样延续了认知经济性原则，与第二阶段中成员向群体规范靠拢的动机一致。

2.3.6 第六阶段：表达规范的稳定化和常态化

当个体在不同的群体互动中做出一致的表达选择后，个体层面的策略偏好会进一步汇聚成群体的偏好。此时极端化表达不再是个体的策略性选择，而是在群体互动的模仿过程中被不断强化的。

整个极端化的过程中，表达规范并没有被明文规定，而是在互动反馈中自然稳定。当极端化表达更常被看见和回应时，温和表达则被边缘化，极端化因此能够在无需态度变化的刺激下，实现稳定化与常态化。

3 不同群体结构下的策略性表达动机的比较分析

3.1 KOL 引导下的极端化加速机制

在 KOL 引导的群体中，极端化表达的扩散、形成、与稳定具有加速的特征。KOL 作为意见领袖，因为其高可见性的优势，被视为一种权威线索，可以加速自上而下纵向的表达规范稳定化与常态化的过程。

在社会认同提升阶段（2.3.1），KOL 的表达具有放大效应。群体成员无需经历漫长的身份区分的试探和协商的流程，KOL 可以迅速确立群体身份的立场边界。

在群体规范敏感性增强和稳定的阶段（2.3.2-2.3.3），KOL 承担了规范制定者的角色。而 KOL 所呈现的表达方式和强度被跟随者视为安全、无需额外解释的表达范式，触发了认知外包的过程，从而显著降低了跟随者的认知负荷。而 KOL 本身的表达强度也被视作群体的表达强度的基准。

在表达工具化与策略选择阶段（对应 2.3.4-2.3.5），高强度表达在风险厌恶和认知经济性的作用下，被反复安全地采用。群体成员可以避免遭受“忠诚不绝对就是绝对不忠诚”的质疑。

最终在表达规范的稳定化和常态化阶段（对应 2.3.6），极端化表达通过 KOL 的持续稳定的表达和跟随者的不断模扩散。最终极端化的表达不再依赖外界的持续刺激，而可以在群体层面获得稳定。

3.2 群体内互动下的极端化协商机制

在没有明显 KOL 的群体中，极端化的形成具有竞争性、缓慢推进、协商确定的特征。群体规范通过成员内持续的横向互动逐步形成。

在社会认同提升阶段（对应 2.3.1），由于去中心化的互动缺乏权威示范，个体得通过使用标签、表达立场、收获反馈的互动流程，慢慢完成“我们”与“别人”的区分。

在群体规范敏感性增强和稳定的阶段（对应 2.3.2-2.3.3），成员们通过不断试探发觉到不同的表达会收获不同的反馈，在成员之间会产生横向比较。为了避免被忽视和被质疑的风险，个体倾向于压缩表达的空间，从而推动群体规范敏感性的形成。但由于个体成员不存在高度集中的意见，规范形成过程可能反复。

在表达工具化与策略选择阶段（对应 2.3.4-2.3.5），高强度表达在风险厌恶和认知经济性仍发挥了重要的心理作用。但在这个基础上，由于个体的竞争性需要，高强度表达更是处于对内部声望需要的考量，个体存在对自己成为 KOL 的期望。

最终在表达规范的稳定化和常态化阶段（对应 2.3.6），由于成员内部对极端化表达的相互模仿和强化，群体内部会形成基于横向互动的稳定的结构。

3.3 两种情境的比较和整合

两种结构在权力结构、规范形成路径上存在系统性差异，两者的核心差异在表 1 中体现。

表 1 KOL 引导结构与群体内互动结构的比较

	KOL 引导结构	群体内互动结构
权力结构	纵向、集中	横向、分散
规范形成速度	自上而下形成，速度快	自下而上反复，速度慢
极端化推动力	示范 - 模仿机制	竞争 - 模仿机制
极端化表达来源	KOL 引领	成员间竞争
群体身份边界	边界明确	边界妥协
认知负荷	低	高

4 综合讨论

本文从社会认同理论出发，认为言论极端化是在群体互动结构中的策略性表达行为，而非通过情绪激化所导致。

在缺乏直接利益冲突的情况下，个体仍会选择高攻击性的表达，并非因为情绪对立，而是因为极端化表达在身份边界的确认上可以降低被误解的风险。

极端化之所以在部分群体中常态化，是因为群体成员会基于社会认同中可见性的需求，压缩温和表达的空间，并通过“KOL 示范 -- 跟随者模仿”和“群体内互相竞争 -- 互相模仿”的形式被稳定地复制学习。

理性推理在部分群体中失效，源于表达的选择并非以保证表达的质量为标准，而是基于风险厌恶和认知经济性的原则，个体倾向于选择通过模仿表达范式以降低认知负荷，理性推理反而可能增加表达风险，从而被系统性地回避。

而在符合社会朴素正义的价值观的场景中，立场也是一种身份立场。会受到如一般身份立场一样经历同样的系统性的极端化的过程。

5 结语

本文通过构建“社会认同—策略性表达动机—群体互动”的六阶段模型，尝试解释极端化表达如何在确认身份边界、可见性竞争、减轻认知负荷、表达风险管理等机制下，在不同群体结构中被激活、放大、规范化、稳定化的。

本文区分了 KOL 引导的群体互动的结构和去中心化互动的两类互动结构，说明相同的心理机制在不同互动结构下会呈现出一定的差异化和共通性，给该模型的适用范围提供了更多可能。

理解极端化表达的生成逻辑，对个人有助于理解自己的观点所处的社会相对位置，对管理机制有助于在不同阶段适配不同的干预手段延缓社会极端化形成的进程。

本文的不足点在于，本文以理论模型和探索性粉丝为主，并没有考虑平台算法机制、个体差异等变量。未来研究可以结合各类平台数据、针对不同特定问题的实验，对问题所述的心理机制进行实证验证。

参考文献

[1] Bail C A, Argyle L P, Brown T W, et al. Exposure to opposing views on social media can increase political polarization[J]. Proceedings of the National Academy of Sciences, 2018, 115(37): 9216-9221.

[2] FISKE S T, TAYLOR S E. Social cognition: From brains to culture[M]. 4th ed. Thousand Oaks: Sage Publications, 2020.

[3] Tajfel H, Turner J, Austin W G, et al. An integrative theory of intergroup conflict[J]. Intergroup relations: Essential readings, 2001: 94-109.

[4] 李纾. 决策心理：齐当别之道[M]. 上海：华东师范大学出版社， 2016.

[5] 张爱军,王思尹.网络群体性事件的社会情绪演进逻辑与动态调节策略[J].信阳师范学院学报(哲学社会科学版), 2024(6).