Proceedings of ICADSS 2026 Symposium: Contemporary Visual Arts, AI Creativity, and Interactive Media: Theory and Practice
DOI: 10.54254/2753-7064/2026.BJ38973

From Persona Collapse to Trust Erosion: Attribution, Cognitive Dissonance, and User Reactions in Inconsistent Social Media Expressions

Hao Tang

Department of Social and Behavioural Sciences, City University of Hong Kong, Hong Kong, China
htang72-c@my.cityu.edu.hk

Abstract. Inconsistent expression refers to the situation where an individual's statements, behaviors, or presented image on social media contradict their previous statements, behaviors, or presented images. This review is based on the theories of impression management and context collapse, using attribution theory and cognitive dissonance theory as explanatory mechanisms, and introducing familiarity as a moderating variable to explain how users perceive and respond to such inconsistencies. Firstly, it explains inconsistent expression and how social media contributes to it. Then, it argues that observers explain inconsistencies through causal attribution: external attribution absolves the target, whereas internal attribution views the inconsistency as a manifestation of hypocrisy, which in turn triggers cognitive dissonance, and users then resolve this dissonance by changing their attitudes, reducing trust, decreasing interaction, or making negative evaluations. Finally, it proposes that familiarity plays a moderating role in this entire chain. At the end, it discusses the existing limitations and future directions.

Keywords: Inconsistent self-expression, attribution, cognitive dissonance, familiarity, social media

1. Introduction

It is commonly seen on social media that the actual behaviors and the displayed images of people are inconsistent; for example, some celebrities who advocate for a healthy lifestyle are found to eat junk food occasionally, and some friends who suggest a simple lifestyle usually purchase numerous expensive items. Such discrepancies may lead to the loss of trust, provoke criticisms, or even result in the breakdown of relationships or a career. The aim of this research is to investigate how the observers understand and interpret the contradictory self-presentation of a particular individual, and how their understanding influences their future actions. It is significant to examine this problem since these interpretations have real impacts on the reputation and cause distrust, unfollowing, and negative word-of-mouth towards the individual. Moreover, this subject is theoretically important as it combines the classical theories of attribution and cognitive dissonance with a special digital phenomenon.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7064/2026.BJ38973

Social networks aggravate the situation. They bring people together not only relatives, friends, and colleagues, but also strangers in one environment [1]. Additionally, as the information uploaded by users on social networks is retained for a long time, the present self is always compared with that in the past. Thus, it is difficult to keep the way people try to control others' impression of themselves [2]. Actually, individuals have to deal with the persistent conflict between presenting an idealized self and displaying their real self [3]. Therefore, this paper firstly introduces the concept of inconsistent expression, then discusses its reasons, the responses obtained, and finally investigates the effect of familiarity on the whole process.

2. Inconsistent expression on social media

On social media, people carry out impression management, which means to modify the way others view them [2]. Inconsistent expressions happen when the various identity-related information given off by an individual conflicts with each other. According to the operational definition, inconsistent self-expression is described as a felt contradiction between two or more such pieces of information - between previous and current statements, between the online image and real behavior, or between the self-proclaimed characteristics and the observable actions.

This definition contains certain common classifications that agree with the present study, which also points out the discrepancies within the same account, between public and private accounts, and between online and offline conditions [4]. Temporal inconsistency means the contradiction between the previous and current information, which can be found because the websites keep the recorded materials. Situational inconsistency occurs when a person behaves differently towards different persons, leading to a situation called context collapse [1, 5]. Online-offline inconsistency refers to the difference between the intended image on the internet and the real-life actions. Furthermore, verbal-act inconsistency indicates the discrepancy between what a person says and what he really does. Besides, there is account-level inconsistency where some people have an excellent main account but possess another one with a poor reputation [6, 7]. All these types of inconsistencies may be due to the characteristics of social media sites. The saving of messages, the communication between different groups of people, and the convenience of posting repeatedly result in the possibility of these different identity clues contradicting each other.

The inconsistency may have an influence on the results. Small mistakes may be ignored, but significant discrepancies, particularly those which are intentional, frequent, and connected with one's personality, can produce a stronger feeling of surprise. According to attribution research, when people infer a person's real character, they give the greatest importance to the behavior that is unusual in the situation [8]. Hence, a main and continuous discrepancy is more likely to be considered as untruthfulness.

3. Attribution of inconsistent expression

Attribution refers to the way individuals determine the reasons for their own and others' behaviors, considering whether an action is due to internal (dispositional) or external (situational) reasons. Kelley has described this judgment process in his covariation model, stating that an effect is attributed to the condition which varies with it; in practice, people use the information of consensus, distinctiveness, and consistency to decide if the cause is personal or situational [8].

Thus, when people meet with some contradictory statements, they often wonder whether the disagreement comes from the individual (such as hypocrisy, dishonesty, or instability) or from the surroundings (e.g., changes in the context, misunderstanding of old messages, or the pressure from

DOI: 10.54254/2753-7064/2026.BJ38973

the public). The conclusion they reach influences their evaluation. If they attribute it to the person's own characteristics ("this is just an acting behavior"), they would make severe judgments. On the contrary, if they believe the cause lies outside ("the platform misunderstood it"), they may excuse the individual. In fact, people often commit a fundamental attribution error, overemphasizing the personal reasons and underestimating the social ones; consequently, inconsistencies are usually attributed to the individual's characteristics instead of the platform or the environment. Weiner pointed out that the causes are classified into three aspects—locus, stability, and controllability—with different emotional effects: internal, stable, and controllable causes result in blame and anger, while external, unstable, and uncontrollable causes lead to sympathy [9]. Some recent research using attribution theory on social media has shown that the observers interpret the inconsistent expressions by inferring the inner motives of the expresser and considering the situational factors [4]. The attribution process plays an important role in connecting the perception of inconsistency and the corresponding reaction.

4. Cognitive dissonance and user responses

Cognitive dissonance refers to an unpleasant psychological state that arises when a person holds two or more cognitions that are contradictory [10]. Because this state is uncomfortable, people have the motivation to eliminate it, for example, by changing a belief or adding a compatible one. Festinger further proposed that dissonance can be reduced not only by changing a belief but also by changing the inconsistent behavior, altering one of the conflicting cognitions, or adding new consonant beliefs that help justify the inconsistency [10].

When a user's previous favorable opinion about another is changed by new inconsistent behavior, the conflict arises, i.e., the idea "this person is reliable" disagrees with the idea "this person is dishonest." The user solves this inconvenience by following several ways, which show certain effects: (1) changing attitude; (2) reducing trust; (3) decreasing contact; and (4) negative evaluation or word-of-mouth criticism. These responses do not exclude each other but frequently happen simultaneously, thus one detected inconsistency may gradually weaken the relationship between the observer and the object. Recent studies have confirmed these responses in the electronic environment: when teenagers come across information contradictory to their own experiences, dissonance makes them doubt and lose trust in the contradictory source [11]; similarly, social network users who are not satisfied with the contents usually react by unfollowing the poster, which means lowering the relationship [12]. Moreover, a negative bias suggests that a single important inconsistency can overcome many previous positive indications, because negative information is usually considered more important than positive information [13].

5. Role of familiarity

Familiarity is important in this procedure, since one's previous experiences and knowledge mainly affect their opinions and feelings. Zajonc's "simple exposure effect" shows that just observing a stimulus and repeatedly exposing oneself to it can improve one's attitude towards it, regardless of any reinforcement [14]. In other words, the familiar things make people have a sense of security and comfort. In the evaluation of interpersonal relationships, familiarity gives individuals a way to obtain new information: the more one knows about a person, the more information there is about their usual behavior. This opinion is in accordance with Kelley's "consistency" and "particularity" judgments [8].

DOI: 10.54254/2753-7064/2026.BJ38973

Therefore, the responses to the contradictions of information for both known and unknown things are probably different. When meeting strangers, the observers have little grounds for judgment, thus a single discrepancy may cause a strong internal attribution — "This person is not trustworthy" — and result in a quick negative evaluation. However, when facing familiar persons, the same contradictory information is based on a series of consistent behavioral histories; the subjects might regard it as situational reasons, or they might feel more dissonance since it contradicts the previously established expectations greatly. The recent study on the contradictions confirms this idea, showing that whether the inconsistency is pardoned or criticized largely depends on the observer's familiarity and relationship with the communicator, together with the perceived authenticity and social category cues [4]. Hence, the familiarity can either reduce (by giving excuses) or increase (by breaking the trust) the reaction, according to the gravity of the inconsistency and the intensity of the previous relationship. This dual effect of familiarity can be explained by attribution: familiar observers have more knowledge about the usual behavior of the target, which enables them to ignore a single contradiction as a situational issue, but it also raises their demands, making a serious violation seem like a greater betrayal.

Familiarity can be considered a moderating factor in this process—first, it alters the attribution, then affects the degree of inequality, and ultimately decides the final evaluation and behavioral reaction.

6. Conclusion

This review describes a procedure for transforming irregular statements into other people's responses. In social media, the combination of different publics into one group and keeping the posts for a long time makes it hard to avoid conflicting information. When the observers notice the inconsistency, they try to find out the causes, thinking whether it is due to oneself or the surrounding environment. If the difference is attributed to oneself, it leads to an evaluation of one's own character and results in cognitive dissonance, which can be eliminated by modifying attitudes, reducing trust, decreasing interaction, and giving negative comments. The familiarity also affects this process: the mere exposure increases the favorableness, and the extra knowledge provides consistent information, which decides whether the inconsistency is accepted or regarded as a breach. By using these fundamental mechanisms in a certain digital event, this review offers a general framework that may be beneficial for future empirical studies on online reputation and trust.

Existing research nevertheless has notable limitations. Much of it is qualitative and descriptive rather than causal: recent work has mapped how people perceive and reconcile inconsistent expressions, but the full attribution–dissonance–response chain has yet to be tested quantitatively. Evidence also remains largely Western-centric, cross-sectional, and reliant on self-report or hypothetical scenarios, with small, convenience-based samples. The construct of "degree of inconsistency" is still under-theorized, and familiarity is rarely treated as a formal moderator.

Future research should therefore (1) test the mediated model experimentally, with larger and more diverse samples and behavioral outcomes; (2) measure attributions with validated instruments; (3) manipulate both the severity of inconsistency and the level of familiarity; (4) examine platform affordances, individual differences, and cross-cultural variation; and (5) explore interventions that help users interpret inconsistency more fairly, thereby mitigating the reputational harms of such "image collapse".

DOI: 10.54254/2753-7064/2026.BJ38973

References

[1] Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society, 13*, 114–133. DOI: 10.1177/1461444810365313.

[2] Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*, 34–47.

[3] Kreling, R., Meier, A., & Reinecke, L. (2022). Feeling authentic on social media: Subjective authenticity across Instagram stories and posts. *Social Media + Society, 8*, 1–13. DOI: 10.1177/20563051221086235.

[4] Li, P., Cho, H., & Qin, Y. (2022). (In)consistency matters: An account of understanding the perception of inconsistent expressions on social media. *Frontiers in Psychology, 13*, 885498. DOI: 10.3389/fpsyg.2022.885498.

[5] Chen, Y. A., & Mares, M.-L. (2025). Social media self-presentation of LGBTQ+ youth in the United States: The role of identity exploration, context collapse, and supportive feedback. *Journal of Computer-Mediated Communication, 30*, zmaf020. https://doi.org/10.1093/jcmc/zmaf020

[6] Darr, C. R., & Doss, E. F. (2022). The fake one is the real one: Finstas, authenticity, and context collapse in teen friend groups. *Journal of Computer-Mediated Communication, 27*, zmac009.

[7] Tao, M., & Ellison, N. B. (2023). "It's your finsta at the end of the day…kind of": Understanding emerging adults' self-presentational changes on secondary accounts. *Social Media + Society, 9*, 1-13. DOI: 10.1177/20563051231152812.

[8] Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation, 15*, 192-238.

[9] Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*, 548-573.

[10] Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

[11] Hashemi, R., & Vogel, E. A. (2024). Adolescents' perceptions of substance use messaging in the age of social media: Resolving cognitive dissonance. *Health Education Research, 39*, 1–11. DOI: 10.1093/her/cyad046.

[12] Hu, J. M., & Zhan, E. S. (2026). Networked intimacy via post visibility: Why do people unfollow WeChat friends? *Communication Quarterly, 74*, 223–246. DOI: 10.1080/01463373.2026.2643642.

[13] Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323–370. DOI: 10.1037/1089-2680.5.4.323.

[14] Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.