Proceedings of the 7th International Conference on Educational Innovation and Psychological Insights
DOI: 10.54254/2753-7048/2026.36059

Signal Tracking and Self-Emotion Coupling in Empathic Misinterpretation of Naturalistic Emotional Narratives

Yutong Hu

College of Arts & Science, New York University, New York, USA
yh4872@nyu.edu

Abstract. Why do observers misread other people's emotions in naturalistic narratives? One possibility is that observers fail to track the target's affective signal; another is that their own emotional response becomes coupled to their inference. This paper tested these accounts using a secondary behavioral analysis of the Stanford Emotional Narratives fMRI Dataset/SENDv1, combining target self-ratings, observer inference ratings, observer own-feeling ratings, and live target-observer tracking metrics. Observer inference was associated with both target actual emotion and observer own feeling, indicating that judgments contained target-related information while also reflecting self-emotion coupling. Self-other affective mismatch predicted empathic error beyond live tracking performance. Signed-error analyses further showed that inference errors were directionally aligned with observers' own emotional deviations from the target, and within-video permutation tests indicated that these associations were not reducible to distance-geometry overlap alone. These findings support a conservative dual-component account of empathic inaccuracy in naturalistic emotional narratives: empathic errors reflect both difficulty tracking the target's affective signal and directionally self-consistent use of the observer's own emotional response.

Keywords: Empathic accuracy, Emotion inference, Self-other mismatch, Naturalistic narratives, Mixed-effects modeling

1. Introduction

Why do people misread another person's emotion in naturalistic situations? This paper asks whether empathic inaccuracy in emotional narratives reflects two contributors: difficulty tracking the target's affective signal and directionally self-consistent coupling to the observer's own emotional response. This question matters because everyday emotion inference rarely depends on a single static cue. Observers must integrate facial expression, voice, language, narrative context, and their own affective response while another person's emotion unfolds over time.

Naturalistic emotional narratives are well suited for studying this problem because they preserve temporal dynamics and contextual richness that are often reduced in simplified laboratory tasks. Recent work on dynamic and context-rich affective stimuli has emphasized that emotion understanding depends on temporally unfolding information rather than isolated cues [1, 2]. The Stanford Emotional Narratives Dataset (SENDv1) was developed to support research on emotion in complex stories. It contains unscripted emotional narrative videos and time-varying affective ratings, allowing researchers to model how emotional meaning develops during naturalistic narratives [3]. Prior SEND-based work has mainly focused on dynamic emotion modeling, multimodal affect prediction, and continuous valence estimation from narrative, facial, acoustic, and linguistic signals. For example, attention-based models have been used to predict emotional trajectories in naturalistic narratives, treating human ratings as annotations or benchmarks for computational emotion recognition [4].

This computational work shows that naturalistic narratives contain rich emotional signals. However, it leaves a psychological question less fully developed: why do human observers sometimes deviate from the target's actual emotional experience? Human observers are not only annotators of target emotion. They are also emotionally responsive perceivers whose own feelings may be coupled to what they infer about the target. Existing SEND-based studies have typically asked how target emotion can be predicted over time; fewer analyses have asked whether observer errors reflect poor target-signal tracking, self-emotion-consistent inference, or both.

Research on empathic accuracy provides one basis for this question. Classic work has examined how accurately perceivers infer another person's thoughts and feelings [5], and social-neuroscience studies have linked empathic accuracy to both mentalizing and affect-sharing processes [6]. More recent naturalistic neuroimaging work has extended this tradition by examining how emotional intent and observer inference align during social consensus [7]. At the same time, affective resonance is not uniformly beneficial for emotion understanding. Evidence from emotion-recognition research suggests that facial mimicry and affective empathy can support recognition in some contexts but do not guarantee accurate inference [8, 9].

A related literature on self-projection, egocentric anchoring, and emotional egocentricity bias suggests why observer own emotion may matter. People often use their own experiences and affective states as anchors when reasoning about others [10-12]. Emotional egocentricity bias further shows that one's own affective state can interfere with judgments about another person's affective state, particularly when self and other experiences conflict [13, 14]. In naturalistic emotion inference, own feeling may therefore have a dual role: it may help when observer and target emotions are aligned, but it may contribute to misinterpretation when they diverge.

The research gap is therefore not whether empathic accuracy exists, or whether self-based inference exists. The open question is whether empathic inaccuracy in naturalistic emotional narratives can be decomposed within the same dataset into two empirically separable components: difficulty tracking the target's emotional signal and directionally self-consistent coupling to the observer's own emotional response. The Stanford Emotional Narratives fMRI Dataset is well suited for this question because it includes target self-ratings, observer inference ratings, observer own-feeling ratings, live target-observer tracking metrics, and fMRI data.

The present paper conducts a secondary behavioral analysis of this dataset. The current manuscript focuses on behavioral evidence; the status of the fMRI data is discussed as a limitation rather than as part of the main empirical argument.

This paper makes three contributions. First, it separates target-signal tracking from observer own-feeling coupling in a naturalistic emotional narrative dataset. Second, it tests whether self-other affective mismatch predicts empathic error beyond live tracking performance. Third, it uses signed-error, permutation, and exploratory tracking-quality analyses to distinguish directionally self-consistent inference from mathematical overlap among distance-based variables.

This paper tested three hypotheses:

1. Observer inference is predicted by both target actual emotion and observer own feeling.

DOI: 10.54254/2753-7048/2026.36059

2. Empathic error increases when observer own feeling diverges from target actual emotion, even after accounting for live tracking performance.

3. If observer own feeling is systematically coupled to inference, signed inference errors should align directionally with observer own emotional deviations from the target.

An additional exploratory analysis examined whether self-emotion coupling varied across observer-level live-tracking quartiles. This analysis was treated as a boundary-condition analysis rather than a primary confirmatory test.

2. Method

2.1. Analytical framework

This study treats empathic inaccuracy as a behavioral outcome that can reflect at least two separable components. The first component is target-signal tracking: observers may be more or less able to follow the target's emotional state. The second component is self-emotion-consistent inference: observers' judgments may be coupled to their own emotional responses. This distinction follows the empathic-accuracy tradition of comparing observer judgments with target-based criteria, while also incorporating work on self-based inference and emotional egocentricity bias.

The framework does not assume that these components are independent psychological modules. Instead, it uses observed ratings to make a more modest analytic distinction. Target-signal tracking asks whether observer inference varies with the target's actual emotion. Self-emotion coupling asks whether observer inference and error direction vary with the observer's own feeling. The signed-error analysis is especially important because absolute error only measures how far the observer was from the target, whereas signed error shows whether the observer overestimated or underestimated the target in a direction consistent with the observer's own emotion.

2.2. Dataset and sample

This analysis used behavioral data from the Stanford Emotional Narratives fMRI Dataset and associated Stanford Emotional Narratives/SENDv1 materials. The broader dataset contains naturalistic emotional narrative videos, target self-ratings, observer ratings, and fMRI data. The present manuscript focuses on behavioral ratings and live target-observer tracking metrics.

The behavioral files were drawn from the associated OSF materials and OpenNeuro dataset record:

- OSF repository: https://osf.io/cs3km/
- OpenNeuro accession: ds006111
- OpenNeuro DOI: doi:10.18112/openneuro.ds006111.v1.0.0
- OpenNeuro dataset version: 1.0.0
- Dataset accessed: June 2026

The public dataset record reports 100 observers. The final merged behavioral analysis table contained 784 usable observer-video observations from 98 observers and 23 videos. Two observers were not retained in the final behavioral merge because the required observer-video rows were not available in the assembled analysis table. Each retained observer contributed ratings for 7 to 8 videos. Each video was rated by 29 to 65 observers in the final merged table. Live tracking RMSE was available for all 784 observations. Live tracking correlation was available for 779 observations; observations without an estimable time-series correlation were treated as missing for models requiring this variable.

DOI: 10.54254/2753-7048/2026.36059

2.3. Variable sources and interpretation

The analysis distinguished between dataset-provided task ratings and study-specific derived variables. This distinction is important because not all variables should be interpreted as established psychometric scales. The dataset-provided variables were treated as task-specific behavioral ratings from the Stanford Emotional Narratives materials. The derived variables were constructed for the present research question.

Dataset-provided variables:

oTargetFeels: observer inference of how the target felt. This was treated as the primary observer inference rating. In the analytic table it ranged from 0 to 100.

oObsFeels: observer's own emotional response to the video. This was treated as the primary own-feeling rating. In the analytic table it ranged from 0 to 100.

TargetActual: target's self-reported actual emotion. This was treated as the target-based criterion variable. It is video-level in the present merged table.

rnEmbody: normalized embodiment-related rating. This was included as a covariate and ranged from 0 to 1.

Compassion: compassion rating. This was included as a covariate and ranged from 0 to 100.

Study-specific derived variables:

$$i = observer, v = video \tag{1}$$

$$Inference_{iv} = oTargetFeels_{iv} \tag{2}$$

$$OwnFeeling_{iv} = oObsFeels_{iv} \tag{3}$$

$$TargetActual_{v} = TargetActual_{v} \tag{4}$$

$$EmpathicError_{iv} = |Inference_{iv} - TargetActual_{v}| \tag{5}$$

$$SelfOtherMismatch_{iv} = |OwnFeeling_{iv} - TargetActual_{v}| \tag{6}$$

$$InferenceOwnDistance_{iv} = |Inference_{iv} - OwnFeeling_{iv}| \tag{7}$$

$$SignedError_{iv} = Inference_{iv} - TargetActual_{v} \tag{8}$$

DOI: 10.54254/2753-7048/2026.36059

$$SelfTargetSigned_{iv} = OwnFeeling_{iv} - TargetActual_{v} \tag{9}$$

EmpathicError indexed overall empathic inaccuracy as absolute distance from the target's self-report. SelfOtherMismatch indexed how far the observer's own feeling diverged from the target's actual emotion. InferenceOwnDistance indexed how close the observer's inference was to the observer's own feeling. SignedError indicated whether the observer overestimated or underestimated the target's emotion. SelfTargetSigned indicated whether the observer's own emotion was higher or lower than the target's actual emotion.

The signed variables were central to the analysis. A positive relation between SelfTargetSigned and SignedError would mean that observers tended to overestimate the target when their own emotion was higher than the target's actual emotion, and underestimate the target when their own emotion was lower. This directional test provides a stronger behavioral test of self-emotion-consistent inference than absolute error alone.

2.4. Live tracking metrics

Live target-observer tracking was indexed using two time-series similarity metrics: root mean squared error (RMSE) and Pearson correlation. RMSE captures the average magnitude of moment-by-moment deviation between observer and target time series. Correlation captures whether the observer followed the temporal pattern of the target's emotional trajectory. These two metrics are complementary: RMSE is sensitive to absolute distance, whereas correlation is sensitive to temporal co-fluctuation.

The analysis used the time-series rating matrices as provided in the source materials. No additional smoothing or interpolation was applied in the present secondary analysis. For each observer-video pair:

$$RMSE_{iv} = \sqrt{mean_t((ObserverRating_{ivt} - TargetRating_{vt})^2)} \tag{10}$$

$$LiveCorrelation_{iv} = cor\left(ObserverRating_{ivt}, TargetRating_{vt}\right) \tag{11}$$

Here, t indexes time points within a video. Higher RMSE indicates poorer live target tracking. Higher live correlation indicates stronger dynamic alignment with the target's emotional trajectory. The present analysis used live tracking metrics extracted from the available time-series rating matrices and merged them with the post-scan behavioral table. Five observer-video pairs did not yield an estimable live correlation after merging, most likely because valid paired time-series variance was insufficient for a Pearson correlation. The analysis did not treat live tracking metrics as standardized clinical measures; they were behavioral indices of dynamic target-signal tracking.

2.5. Primary statistical models

The data have a crossed repeated-measures structure: each observer rated multiple videos, and each video was rated by multiple observers. Ordinary least squares models that treat all observations as independent would therefore underestimate the dependence induced by repeated observers and

DOI: 10.54254/2753-7048/2026.36059

shared videos. The primary analyses used crossed random-intercept mixed-effects models, a standard approach for participant-by-stimulus data structures in behavioral research [15].

General model form:

$$Y_{iv} = \beta_0 + \beta_1 X_{iv} + u_i + w_v + error_{iv} \tag{12}$$

$$u_i = observerrandomintercept \tag{13}$$

$$w_v = videorandomintercept \tag{14}$$

All focal continuous predictors were z-scored before model estimation. The first model tested whether observer inference reflected both target actual emotion and observer own feeling:

$$Inference_{iv} = \beta_0 + \beta_1 TargetActual_v + \beta_2 OwnFeeling_{iv} + u_i + w_v + error_{iv} \tag{15}$$

The second model tested whether self-other mismatch predicted empathic error beyond live tracking performance and covariates:

$$EmpathicError_{iv} = \beta_0 + \beta_1 SelfOtherMismatch_{iv} + \beta_2 InferenceOwnDistance_{iv} \tag{16}$$

$$+\beta_3 LiveRMSE_{iv} + \beta_4 LiveCorrelation_{iv} + \beta_5 Embodiment_{iv}$$

$$+\beta_6 Compassion_{iv} + u_i + w_v + error_{iv}$$

The third model tested whether signed inference errors were directionally aligned with observer own-emotion deviations:

$$SignedError_{iv} = \beta_0 + \beta_1 SelfTargetSigned_{iv} + \beta_2 LiveRMSE_{iv} \tag{17}$$

$$+\beta_3 LiveCorrelation_{iv} + \beta_4 Embodiment_{iv} + \beta_5 Compassion_{iv}$$

$$+u_i + w_v + error_{iv}$$

DOI: 10.54254/2753-7048/2026.36059

The fourth model examined whether the cost of inference-own closeness depended on self-other mismatch:

$$EmpathicError_{iv} = \beta_0 + \beta_1 SelfOtherMismatch_{iv} + \beta_2 InferenceOwnDistance_{iv} \tag{18}$$

$$+\beta_3 \left(SelfOtherMismatch_{iv} \times InferenceOwnDistance_{iv}\right)$$

$$+\beta_4 LiveRMSE_{iv} + \beta_5 LiveCorrelation_{iv} + u_i + w_v + error_{iv}$$

This interaction model was included because closeness between inference and own feeling should not be assumed to be uniformly harmful. Its cost should depend on whether own feeling and target actual emotion are aligned or mismatched.

2.6. Robustness and mathematical-overlap checks

Several derived variables share common rating components, especially TargetActual. This creates a risk that some associations could partly reflect mathematical overlap rather than a meaningful behavioral pattern. Two steps were used to address this concern.

First, two-way cluster-robust OLS models clustered by observer and video were fit as robustness checks for the mixed-effects results. Second, within-video permutation tests were used to evaluate whether the main associations exceeded what would be expected from distance geometry alone. In these tests, oObsFeels was shuffled within each video. This preserved the video-level target emotion while disrupting observer-specific own-feeling alignment.

Permutation logic:

Observed statistic:

cor(EmpathicError, SelfOtherMismatch)

cor(SignedError, SelfTargetSigned)

Permutation procedure:

for each video:

shuffle oObsFeels across observers

recompute SelfOtherMismatch and SelfTargetSigned

recompute the two correlations

Number of permutations = 5000

The empirical p value was computed by comparing the observed statistic with the permutation null distribution.

DOI: 10.54254/2753-7048/2026.36059

2.7. Exploratory tracking-quality analysis

An exploratory analysis examined whether self-emotion coupling varied as a function of observer-level live tracking quality. Observers were grouped into quartiles based on their mean live tracking RMSE, with lower RMSE indicating stronger live tracking. Q1 represented the strongest live tracking group and Q4 represented the weakest live tracking group.

The quartile analysis was treated as exploratory for two reasons. First, quartile grouping reduces a continuous tracking measure into categories. Second, the sample was not designed to identify stable observer subtypes. The quartile results were therefore used as descriptive boundary-condition evidence rather than as a primary confirmatory test.

Quartile moderation models:

$$EmpathicError_{iv} = \beta_0 + \beta_1 SelfOtherMismatch_{iv} + \beta_2 TrackingQuartile_i \tag{19}$$

$$+\beta_3 \left(SelfOtherMismatch_{iv} \times TrackingQuartile_i\right) + controls_{iv} + error_{iv}$$

$$SignedError_{iv} = \beta_0 + \beta_1 SelfTargetSigned_{iv} + \beta_2 TrackingQuartile_i \tag{20}$$

$$+\beta_3 \left(SelfTargetSigned_{iv} \times TrackingQuartile_i\right) + controls_{iv} + error_{iv}$$

2.8. Software and reproducibility

Analyses were conducted in Python 3 using pandas 1.5.3, numpy 1.24.3, scipy 1.10.1, statsmodels 0.14.0, matplotlib 3.7.1, and seaborn 0.12.2. Mixed-effects models were estimated using statsmodels. Permutation tests used a fixed random seed of 20260608. Processed analysis tables, scripts, and generated figures were saved locally in the project analysis directory.Your paper must provide the email address of the corresponding author. Use the asterisk (*) symbol to denote the corresponding author in the author list and prior to the email address following the affiliations.

3. Method

3.1. Descriptive statistics and data structure

A descriptive summary is necessary because the analysis combines observer-level, video-level, and observer-video-level variables. The final analytic sample contained 784 observer-video observations from 98 observers and 23 videos. Each observer contributed ratings for 7 to 8 videos, and each video was rated by 29 to 65 observers in the final merged table. Live tracking RMSE was available for all 784 observations, whereas live tracking correlation was available for 779 observations.

Across the analytic sample, mean empathic error was 16.77 rating units, mean self-other mismatch was 16.85 rating units, and mean inference-own distance was 8.92 rating units. Descriptive statistics for all primary observed and derived variables are shown in Table 1. Zeroorder associations among observer inference, observer own feeling, target actual emotion, error measures, and live tracking metrics are shown in Figure 1.

Unless otherwise noted, inferential results below use standardized focal predictors in crossed random-intercept mixed-effects models. The coefficients therefore describe associations in standard-deviation units while accounting for repeated observers and shared videos.

Table 1. Descriptive statistics of behavioral and tracking variables

Variable	N	Mean	SD	Min	Max
Observer inference	784	48.293	32.274	0.0	100.0
Observer own feeling	784	46.554	28.003	0.0	100.0
Target actual emotion	784	52.858	24.993	3.029	84.444
Empathic error	784	16.773	12.359	0.028	65.121
Self-other mismatch	784	16.853	12.568	0.015	66.03
Inference-own distance	784	8.916	10.884	0.0	57.032
Signed error	784	-4.565	20.336	-56.153	65.121
Self-target signed difference	784	-6.304	20.064	-66.03	62.97
Live tracking RMSE	784	20.89	9.432	3.388	62.164
Live tracking correlation	779	0.334	0.471	-0.923	0.962
Embodiment	784	0.276	0.314	0.0	1.0
Compassion	784	64.697	25.244	0.0	100.0

3.2. Observer inference reflected both target emotion and own feeling

The inference model tested the first part of the analytic framework. In empathic-accuracy research, a perceiver's judgment is evaluated against a target-based criterion; therefore, a positive association between target actual emotion and observer inference provides evidence of target-signal tracking. At the same time, a positive association between observer own feeling and observer inference provides evidence of self-emotion coupling.

In the crossed random-intercept mixed-effects model, observer inference was significantly associated with target actual emotion, $b = 0.264$, $SE = 0.020$, 95% CI [0.225, 0.304], $z = 13.15$, $p < .001$. Observer inference was also significantly associated with observer own feeling, $b = 0.712$, $SE = 0.020$, 95% CI [0.673, 0.752], $z = 35.45$, $p < .001$ (Table 2). These results indicate that observer judgments contained target-related information while also being strongly coupled to observers' own emotional responses.

The own-feeling coefficient was larger than the target-actual coefficient. This comparison should be interpreted cautiously because target actual emotion varied primarily at the video level, whereas observer own feeling varied at the observer-video level. The conclusion is therefore not that own feeling is necessarily more important than target emotion, but that observer inference reflected both information sources in the present naturalistic rating structure.

DOI: 10.54254/2753-7048/2026.36059

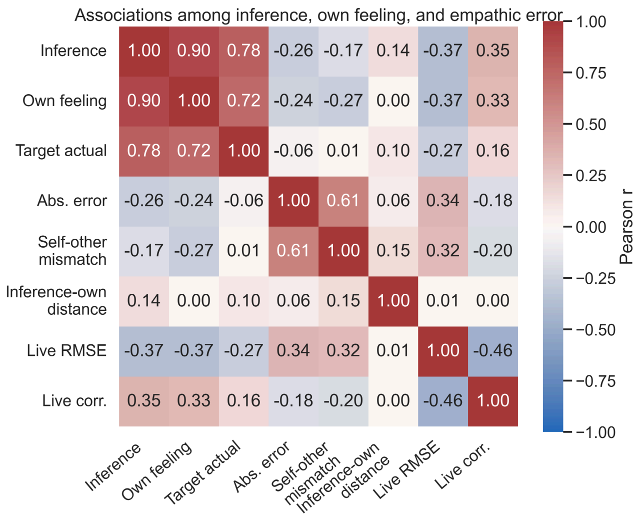

Figure 1. Correlation matrix of core behavioral variables

3.3. Observer inference reflected both target emotion and own feeling

The inference model tested the first part of the analytic framework. In empathic-accuracy research, a perceiver's judgment is evaluated against a target-based criterion; therefore, a positive association between target actual emotion and observer inference provides evidence of target-signal tracking. At the same time, a positive association between observer own feeling and observer inference provides evidence of self-emotion coupling.

In the crossed random-intercept mixed-effects model, as shown in Figure 2, observer inference was significantly associated with target actual emotion, $b = 0.264$, $SE = 0.020$, 95% CI [0.225, 0.304], $z = 13.15$, $p < .001$. Observer inference was also significantly associated with observer own feeling, $b = 0.712$, $SE = 0.020$, 95% CI [0.673, 0.752], $z = 35.45$, $p < .001$ (Table 2). These results indicate that observer judgments contained target-related information while also being strongly coupled to observers' own emotional responses.

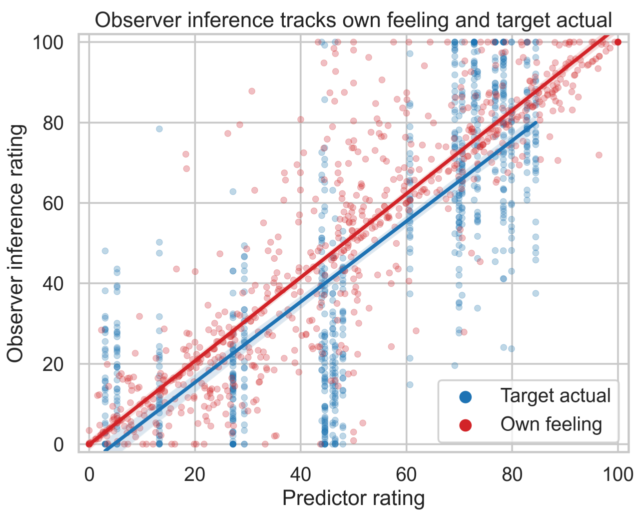

Figure 2. Observer inference as a function of target actual emotion and observer own feeling

DOI: 10.54254/2753-7048/2026.36059

The own-feeling coefficient was larger than the target-actual coefficient. This comparison should be interpreted cautiously because target actual emotion varied primarily at the video level, whereas observer own feeling varied at the observer-video level. The conclusion is therefore not that own feeling is necessarily more important than target emotion, but that observer inference reflected both information sources in the present naturalistic rating structure.

Table 2. Mixed-effects models predicting observer inference and empathic error

Model	Predictor	b	SE	z	p
Inference model	Target actual emotion	0.264	0.02	13.15	< .001
Inference model	Observer own feeling	0.712	0.02	35.45	< .001
Empathic error model	Self-other mismatch	0.563	0.03	19.07	< .001
Empathic error model	Inference-own distance	0.006	0.029	0.22	0.827
Empathic error model	Live tracking RMSE	0.148	0.032	4.64	< .001
Empathic error model	Live tracking correlation	-0.006	0.031	-0.19	0.851
Empathic error model	Embodiment	0.077	0.03	2.54	0.011
Empathic error model	Compassion	0.11	0.03	3.69	< .001

3.4. Self-other mismatch predicted empathic error beyond live tracking

The empathic-error model tested whether final misinterpretation could be explained by dynamic target-tracking difficulty alone. If empathic error only reflected failure to follow the target's emotion over time, then self-other mismatch should add little after accounting for live RMSE and live correlation. Conversely, if observer own emotion contributes additional structure to error, self-other mismatch should predict empathic error beyond live tracking.

Self-other mismatch significantly predicted empathic error, as shown in Figure 3, $b = 0.563$, $SE = 0.030$, 95% CI [0.505, 0.621], $z = 19.07$, $p < .001$ (Table 2). Live tracking RMSE also predicted empathic error, $b = 0.148$, $SE = 0.032$, 95% CI [0.085, 0.210], $z = 4.64$, $p < .001$. Live tracking correlation was not significant in this model, $b = -0.006$, $SE = 0.031$, 95% CI [-0.066, 0.055], $z = -0.19$, $p = .851$.

Inference-own distance did not show a significant main effect, $b = 0.006$, $SE = 0.029$, 95% CI [-0.051, 0.064], $z = 0.22$, $p = .827$. This nonsignificant result is theoretically important. It suggests that closeness between inference and own feeling was not uniformly harmful; its consequences depend on whether the observer's own feeling was aligned with or divergent from the target's actual emotion. Embodiment and compassion were positively associated with empathic error in this model, but these covariate associations should be interpreted cautiously because they may reflect affective engagement or response tendencies rather than independent causal effects.

Together, these results support the dual-component account: empathic error was associated with both dynamic target-tracking difficulty and self-other affective mismatch.

Figure 3. Empathic error increases with self-other mismatch

3.5. Signed errors were directionally aligned with own-emotion deviations

The signed-error model tested the strongest behavioral implication of self-emotion-consistent inference. Absolute error shows how inaccurate an observer was, but it does not show the direction of misinterpretation. Signed error distinguishes overestimation from underestimation and therefore tests whether observer errors move in the direction of the observer's own emotional deviation from the target.

Observer self-target signed difference strongly predicted signed inference error, as shown in Figure 4, b = 0.722, SE = 0.023, 95% CI [0.677, 0.768], z = 30.93, $p < .001$ (Table 3). This means that when observers' own emotion was higher than the target's actual emotion, they tended to overestimate the target's emotion; when their own emotion was lower than the target's actual emotion, they tended to underestimate the target's emotion.

Figure 4. Signed-error alignment between own-emotion deviation and inference error

DOI: 10.54254/2753-7048/2026.36059

Live tracking RMSE was also associated with signed error, b = -0.054, SE = 0.025, 95% CI [-0.104, -0.005], z = -2.15, p = .031. Live tracking correlation was positively associated with signed error, b = 0.147, SE = 0.026, 95% CI [0.096, 0.197], z = 5.72, p < .001. Embodiment was not significantly associated with signed error, whereas compassion showed a small negative association. The central result is the self-target signed difference effect, which provides directional evidence consistent with self-emotion-consistent inference.

Table 3. Signed-error model and permutation-test results

Analysis	Test/Predictor	Estimate	SE / null SD	Statistic	p
Signed-error mixed model	Self-target signed difference	0.722	0.023	30.93	< .001
Signed-error mixed model	Live tracking RMSE	-0.054	0.025	-2.15	0.031
Signed-error mixed model	Live tracking correlation	0.147	0.026	5.72	< .001
Signed-error mixed model	Embodiment	-0.018	0.023	-0.81	0.419
Signed-error mixed model	Compassion	-0.061	0.023	-2.67	0.008
Within-video permutation test	Abs. error vs self-other mismatch	0.614	0.03	null mean r = 0.141	< .001
Within-video permutation test	Signed error vs self-target signed difference	0.761	0.023	null mean r = 0.345	< .001

3.6. Permutation tests addressed mathematical coupling

The main derived variables share common rating components, especially target actual emotion. This creates a legitimate concern that associations among error measures may partly reflect distance geometry. The permutation tests were therefore used as a robustness check against a purely mathematical-overlap explanation.

For the absolute-error analysis, the observed association between empathic error and self-other mismatch was r = 0.614. The within-video permutation null distribution had a mean association of r = 0.141, with null SD = 0.030. The observed association was substantially larger than expected under this null distribution, empirical p < .001.

For the signed-error analysis, the observed association between signed inference error and self-target signed difference was r = 0.761. The within-video permutation null distribution had a mean association of r = 0.345, with null SD = 0.023. The observed association was again substantially larger than expected under the permutation null distribution, empirical p < .001.

These results do not eliminate all interpretive constraints created by shared measurement components, but they indicate that the main associations exceeded what would be expected from distance-geometry overlap alone.

DOI: 10.54254/2753-7048/2026.36059

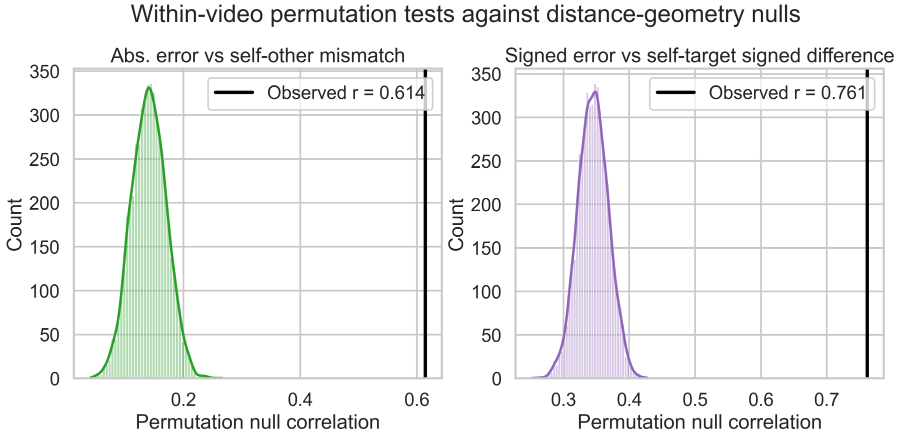

Figure 5. Within-video permutation tests

3.7. Exploratory context-dependent cost of self-emotion coupling

The interaction model tested a boundary condition implied by the analytic framework. If own feeling can support inference when self and target are aligned but mislead inference when they diverge, then the relation between inference-own distance and empathic error should depend on the degree of self-other mismatch.

The interaction between self-other mismatch and inference-own distance was significant, b = -0.392, SE = 0.017, 95% CI [-0.425, -0.359], z = -23.27, p < .001. The negative interaction indicates that the positive relation between self-other mismatch and empathic error was stronger when inference was closer to the observer's own feeling. Put differently, closeness between inference and own feeling was most costly when the observer's own feeling diverged from the target's actual emotion. The model-implied pattern is shown in Figure 5. Because this interaction was not the primary confirmatory test and simple-slope estimates were not specified a priori, it is interpreted as exploratory evidence for a context-dependent account of self-emotion coupling rather than as a standalone mechanistic result (see Figure 6).

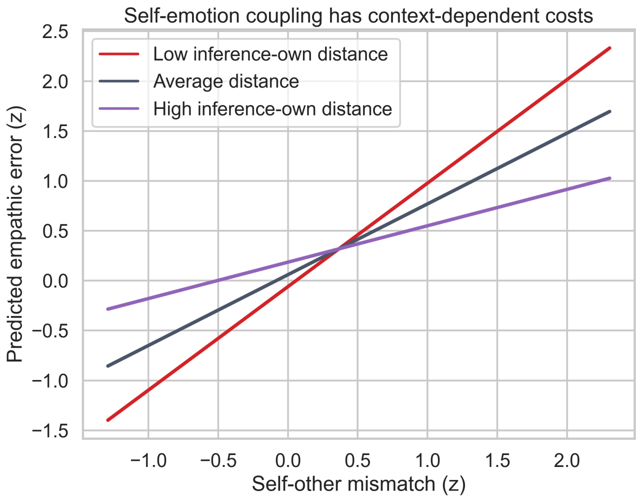

Figure 6. Exploratory interaction between self-other mismatch and inference-own distance

DOI: 10.54254/2753-7048/2026.36059

3.8. Exploratory tracking-quality moderation

The final analysis examined whether self-emotion coupling varied across observer-level live tracking quality. This analysis was exploratory because quartile grouping reduces a continuous variable and because the sample was not designed to identify stable observer subtypes (see Figure 7 and Table 4).

Observers were grouped by mean live tracking RMSE, with Q1 representing the strongest tracking group and Q4 representing the weakest tracking group. Descriptively, mean empathic error was lowest in the strongest live-tracking group and higher in the weaker tracking groups. However, the moderation of absolute error by self-other mismatch was not statistically reliable across quartiles. The signed-error model showed a more specific but non-monotonic pattern: the association between self-target signed difference and signed inference error was stronger for Q3 relative to Q1, but not for Q2 or Q4.

Therefore, the quartile results do not support a strong claim that poorer tracking uniformly increases reliance on own emotion. They suggest that tracking quality may shape self-emotion coupling in some observer groups or contexts, but this pattern should be treated as hypothesis-generating.

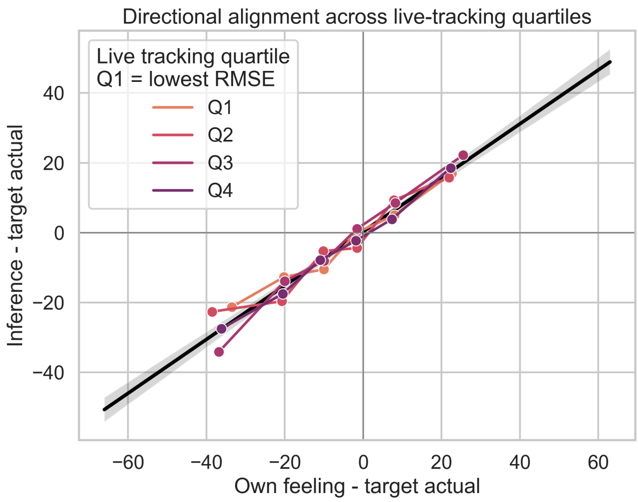

Figure 7. Exploratory live-tracking quartile analysis

Table 4. Live-tracking quartile moderation results

Analysis	Predictor	b	SE	t	p
Empathic error by live-tracking quartile	Live-tracking quartile Q2	0.127	0.081	1.56	0.119
Empathic error by live-tracking quartile	Live-tracking quartile Q3	0.22	0.076	2.89	0.004
Empathic error by live-tracking quartile	Live-tracking quartile Q4	0.151	0.084	1.81	0.071
Empathic error by live-tracking quartile	Self-other mismatch	0.512	0.076	6.76	< .001

DOI: 10.54254/2753-7048/2026.36059

Table 4. (continued)

Empathic error by live-tracking quartile	Self-other mismatch x Q2	-0.06	0.113	-0.53	0.593
Empathic error by live-tracking quartile	Self-other mismatch x Q3	0.143	0.101	1.42	0.155
Empathic error by live-tracking quartile	Self-other mismatch x Q4	0.083	0.101	0.82	0.414
Signed error by live-tracking quartile	Live-tracking quartile Q2	0.051	0.065	0.79	0.430
Signed error by live-tracking quartile	Live-tracking quartile Q3	0.069	0.06	1.16	0.248
Signed error by live-tracking quartile	Live-tracking quartile Q4	0.071	0.065	1.08	0.278
Signed error by live-tracking quartile	Self-target signed difference	0.668	0.059	11.28	< .001
Signed error by live-tracking quartile	Self-target signed difference x Q2	-0.04	0.087	-0.46	0.642
Signed error by live-tracking quartile	Self-target signed difference x Q3	0.158	0.07	2.28	0.023
Signed error by live-tracking quartile	Self-target signed difference x Q4	0.049	0.076	0.65	0.515

3.9. Robustness checks

Two-way cluster-robust OLS models clustered by observer and video were used as robustness checks for the primary mixed-effects models. This check is appropriate because it evaluates whether the focal effects persist under an alternative way of handling repeated observers and shared stimulus videos. The key inference model effects remained positive and statistically reliable for target actual emotion and observer own feeling. The self-other mismatch effect also remained positive and statistically reliable in the empathic-error model. The signed-error association between self-target signed difference and signed inference error remained large and statistically reliable. These robustness checks support the main behavioral conclusions while preserving the caution that the study is correlational and based on derived variables.

4. Conclusion

The results support three main conclusions. First, observer inference was associated with both target actual emotion and observer own feeling, consistent with both target-signal tracking and self-emotion coupling. Second, self-other mismatch predicted empathic error even after accounting for live tracking performance, suggesting that empathic inaccuracy was not reducible to tracking failure alone. Third, signed errors were directionally aligned with observer own-emotion deviations from the target, providing directional evidence consistent with self-emotion-consistent inference.

DOI: 10.54254/2753-7048/2026.36059

Exploratory tracking-quality analyses did not show a consistent monotonic moderation pattern and should therefore be interpreted cautiously.

This secondary behavioral analysis supports a dual-component account of empathic inaccuracy in naturalistic emotional narratives. Observers' inferences were grounded in the target's actual emotional state, but were also strongly coupled to observers' own emotional responses. Empathic error increased when observer own feeling diverged from target actual emotion, and this effect remained after controlling for live tracking performance.

The signed-error analysis strengthens the interpretation. Rather than only showing that observers made larger errors when self and other emotions diverged, signed errors showed that the direction of misinterpretation followed the observer's own emotional deviation. This pattern is consistent with self-emotion-consistent inference.

The findings also suggest that self-emotion coupling may have a dual role. It may support empathic inference when self and other emotions are aligned, but become a source of error when they diverge. This framing avoids treating self-projection as uniformly maladaptive and instead situates it as a context-dependent inference strategy.

The exploratory quartile analysis adds a modest boundary-condition result. Observers with weaker live tracking did not show a uniformly stronger absolute-error cost of self-other mismatch. However, signed-error coupling was stronger in one lower-tracking quartile, suggesting that poorer target tracking may sometimes coincide with greater directional reliance on one's own emotional state. Because this pattern was not monotonic, it should not be overinterpreted as evidence for a discrete observer subtype. Its main value is to refine the behavioral account: self-emotion-consistent inference appears robust overall, whereas its dependence on tracking quality is more tentative.

The present analysis is correlational. The signed-error results are consistent with self-emotion-consistent inference, but they do not establish that observer own feeling causally produced inference errors. Causal claims would require experimental manipulation of observer affect or target information.

Several variables are mathematically related because they are derived from the same three rating values. This issue was addressed with within-video permutation tests, but this remains an interpretive caveat.

The quartile analysis was exploratory and should be treated as hypothesis-generating. Quartiles simplify a continuous live-tracking measure, and the interaction pattern was not strictly monotonic. Future work should test tracking-quality moderation using preregistered continuous models or independent samples.

Observer own feeling was analyzed at the video/post-scan level, not as a fully continuous moment-by-moment own-feeling time series during scanning. Therefore, the current evidence supports video-level self-emotion coupling, not necessarily moment-by-moment self-projection during live inference.

The raw fMRI analysis is currently limited by missing anatomical T1w NIfTI files for most participants in the downloaded OpenNeuro data. The current local dataset contains:

BOLD NIfTI files: 200

BOLD JSON files: 200

T1w JSON files: 100

T1w NIfTI files: 1

Thus, the functional data appear broadly available, but full standard fMRIPrep preprocessing cannot proceed for all participants until the anatomical T1w images are obtained or the OpenNeuro release issue is resolved. For this reason, fMRI is treated as a limitation and future-data issue rather than as part of the main behavioral argument.

References

[1] Scheibe, S., Grundmann, F., Kranenborg, B., & Epstude, K. (2023). Empathising with masked targets: Limited side effects of face masks on empathy for dynamic, context-rich stimuli. *Cognition and Emotion, 37*(4), 835-849. https://doi.org/10.1080/02699931.2023.2193385

[2] Ortega, J., Murai, Y., & Whitney, D. (2023). Modeling emotional cue integration of context-rich and dynamic stimuli reveals Bayesian as well as anti-Bayesian properties. *Journal of Vision, 23*(9), 5221. https://doi.org/10.1167/jov.23.9.5221

[3] Ong, D. C., Wu, Z., Zhi-Xuan, T., Reddan, M., Kahhale, I., Mattek, A., & Zaki, J. (2021). Modeling emotion in complex stories: The Stanford Emotional Narratives Dataset. *IEEE Transactions on Affective Computing, 12*(3), 579-594. https://doi.org/10.1109/TAFFC.2019.2955949

[4] Wu, Z., Zhang, X., Zhi-Xuan, T., Zaki, J., & Ong, D. C. (2019). Attending to emotional narratives. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. https://doi.org/10.1109/ACII.2019.8925497

[5] Ickes, W. (1997). *Empathic Accuracy*. Guilford Press.

[6] Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science, 19*(4), 399-404. https://doi.org/10.1111/j.1467-9280.2008.02099.

[7] Reddan, M. C., Ong, D. C., Wager, T. D., Mattek, S., Kahhale, I., & Zaki, J. (2025). Neural signatures of emotional intent and inference align during social consensus. *Nature Communications, 16*, 6156. https://doi.org/10.1038/s41467-025-59931-8

[8] Holland, A. C., O'Connell, G., & Dziobek, I. (2021). Facial mimicry, empathy, and emotion recognition: A meta-analysis of correlations. *Cognition and Emotion, 35*(1), 150-168. https://doi.org/10.1080/02699931.2020.1815655

[9] Israelashvili, J., Sauter, D., & Fischer, A. (2020). Two facets of affective empathy: Concern and distress have opposite relationships to emotion recognition. *Cognition and Emotion, 34*(6), 1112-1122. https://doi.org/10.1080/02699931.2020.1724893

[10] Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11*(2), 49-57. https://doi.org/10.1016/j.tics.2006.11.004

[11] Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*(3), 327-339. https://doi.org/10.1037/0022-3514.87.3.327

[12] Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General, 142*(1), 151-162. https://doi.org/10.1037/a0028232

[13] Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *The Journal of Neuroscience, 33*(39), 15466-15476. https://doi.org/10.1523/JNEUROSCI.1488-13.2013

[14] Riva, F., Triscoli, C., Lamm, C., Carnaghi, A., & Silani, G. (2016). Emotional egocentricity bias across the life-span. *Scientific Reports, 6*, 32418. https://doi.org/10.1038/srep32418

[15] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005