Proxy-to-Preference Transfer Under Structured Label Mismatch: Tail-Class Collapse in Few-Shot Adaptation from Physics-Derived Proxies to Human Preference Labels

Anonymous authors
Paper under double-blind review

Abstract

Many learning systems must predict a scarce, subjective human target using supervision that is abundant but only a *structured proxy* for that target: a simulator output, a heuristic score, or a physical model. We study the resulting **proxy-to-preference transfer** problem — how to exploit abundant proxy-labeled data without letting the proxy silently redefine the target it approximates — and identify a specific, previously undocumented failure mode of naive adaptation in this setting. We instantiate the problem in multimodal thermal sensation inference: a source domain of 141,330 synthetic state-text pairs labeled with a physics-derived comfort index (Predicted Mean Vote, PMV) and a target domain of 279 real, chronologically ordered human comfort judgments (Thermal Sensation Vote, TSV). The proxy-target relationship here is not merely noisy: PMV is directionally biased and compresses the dynamic range of TSV, and the bias itself is regime-dependent. We pretrain a multimodal encoder on proxy supervision and evaluate zero-shot and few-shot transfer to the human target. Aggregate metrics reproduce a familiar-looking tension: zero-shot transfer is already strong, while few-shot adaptation raises exact agreement but degrades mean absolute error (MAE). Disaggregating by class reveals what the aggregate numbers hide: adaptation does not fail uniformly. On our main split, the adapted model's predictions never fall below the majority label region of its small adaptation set — it stops covering 38% of the true label distribution entirely, in the direction *opposite* the proxy's own bias. We argue this is not an artifact of one dataset but a structural risk whenever (i) the adaptation set is small, (ii) its label distribution is unrepresentative of the deployment distribution — which chronological, non-i.i.d. real-world adaptation windows often are by construction, and (iii) the target label space is discrete and imbalanced. We call this **tail-class collapse under adaptation**, give empirical evidence for the mechanism (not just the symptom), and argue that structured proxy supervision is best used as an inductive bias for representation learning rather than a target substitute, with adaptation procedures evaluated for label-tail coverage rather than aggregate error alone.

1 Introduction

A recurring situation in applied machine learning is this: the label a system is ultimately deployed to predict is a scarce, subjective human judgment, while the label that is cheap and abundant is a structured proxy for it — a simulator score, an expert heuristic, a physics-based index, or a model-generated pseudo-label. The standard response is to treat the proxy as a stand-in for the target, either by training directly on proxy labels or by pretraining on them and fine-tuning on a small amount of real target supervision. This paper is about what goes wrong with the second strategy when the proxy-target relationship is *structured* rather than random, and the target label space is discrete and imbalanced.

We study this question through a concrete instantiation — thermal sensation inference — chosen because it exhibits every ingredient of the general problem in a controlled, quantifiable form: a large physics-derived proxy corpus (Predicted Mean Vote, PMV, computed from measurable environmental variables via the

Fanger model; Fanger, 1970), a small human-labeled target trajectory (Thermal Sensation Vote, TSV, a 6-point subjective scale defined in ASHRAE Standard 55; ASHRAE, 2020), a documented and quantifiable proxy-target mismatch, and a target adaptation set that is chronologically ordered rather than i.i.d. — a property shared by most real-world deployment logs. We treat the thermal-comfort setting as a *testbed*, not as the paper's subject matter; every claim we make is intended to generalize to other proxy-to-preference settings (e.g., simulator-based reward proxies for human preference, rule-based risk scores standing in for expert judgment, or heuristic engagement metrics standing in for genuine user satisfaction) wherever the same three structural ingredients hold.

Our real deployment data make the mismatch concrete before any model is trained. In a 279-hour real trajectory from a single building, TSV spans six discrete levels from -2 to $+3$, including 24 warm responses at $+2/+3$. PMV in the same trajectory never exceeds 0.98 and remains centered on the cool side (mean PMV -0.9706 vs. mean TSV 0.2079; rounded-PMV exact agreement with TSV is only 0.147). The mismatch is *directional* — PMV systematically under-represents warm subjective states — and *regime-dependent*: under (window= 1, door= 1, ac= 0), mean PMV is -2.197 while mean TSV is 0.795, a gap of 2.992. This is the profile of a structured proxy, not a noisy one, and it is what makes substitution the wrong default strategy.

The paper's central question is:

> When abundant proxy supervision and scarce human-preference supervision are correlated but systematically mismatched, what specifically happens when a model pretrained on the proxy is adapted to the human target with a small, non-i.i.d. adaptation set — and is the resulting behavior something we can name and predict, rather than only observe as noise?

Our answer, developed empirically in Section 6 and stated as a general hypothesis in Section 3.2, is that naive adaptation under these conditions tends to produce **tail-class collapse**: the adapted model's output distribution contracts toward the classes that dominate the small adaptation set, and coverage of underrepresented target classes is lost — sometimes completely — even as aggregate exact accuracy improves. In our main 128/32/119 chronological split, the fine-tuned model never predicts TSV -2 or -1 on the test set (0/119 predictions), even though those two classes make up 38% of ground truth. Its regression outputs are almost entirely non-negative, the opposite bias direction from PMV's own cool skew. This is a specific, checkable failure mode, not a vague statement that "small data is hard."

This paper makes four contributions:

1. It formulates **proxy-to-preference transfer under structured label mismatch** as a general multimodal learning problem, and states **tail-class collapse under adaptation** as a specific, falsifiable failure mode within it (Section 3).
2. It instantiates the problem with a synthetic-to-real setup combining 141,330 proxy-labeled source examples and a scarce, chronologically-ordered real target set with human TSV annotations (Section 4).
3. It shows that strong zero-shot proxy transfer coexists with unstable few-shot adaptation, and — going beyond the aggregate observation — that the instability is explained by class-level collapse rather than uniform degradation: disaggregating by true label class shows the adapted model's error is concentrated almost entirely in classes absent or rare in the adaptation window (Section 6).
4. It argues, and gives supporting evidence, that language and operational context help characterize *where* physically grounded proxy labels diverge from human preference, and that per-class / tail-coverage metrics — not aggregate MAE or exact accuracy alone — are necessary to detect this failure mode in future proxy-to-preference benchmarks (Sections 6–7).

2 Related Framing

This work sits at the intersection of four lines of research, and we use each to sharpen a specific claim rather than as background color.

Multimodal learning from heterogeneous evidence. Our setting combines objective sensor observations with subjective language, following the general representation/fusion/alignment taxonomy of multimodal learning (Baltrušaitis et al., 2019). Unlike large-scale vision-language contrastive pretraining such as CLIP (Radford et al., 2021), our text-sensor fusion is routed through a differentiable physical layer rather than a similarity objective — the physics is part of the computation graph, which is what makes the proxy label PMV a *structured*, not merely correlated, proxy for the target.

Synthetic-to-real transfer. The source domain here provides abundant labels through simulation rather than annotation (Nikolenko, 2019); the scientific question is not whether pretraining on synthetic data helps in the abstract — a large synthetic-to-real literature already establishes that it often does for covariate shift — but whether the *label space itself*, not just the input distribution, differs between source and target, which is a harder and less studied variant of the problem than the feature-level domain adaptation setting studied by, e.g., Ganin & Lempitsky (2015) or Tzeng et al. (2017).

Learning under proxy-target mismatch. In Zhou's taxonomy of weak supervision (Zhou, 2018), our setting is closest to *inexact supervision*: the proxy label is well-defined and reliably computed, but it targets a related-but-distinct construct rather than the true label. This differs from the more commonly studied *inaccurate* supervision (label noise) or *incomplete* supervision (missing labels), and from pseudo-labeling approaches where the proxy is model-generated (Ruder, 2018) or rule-generated at scale (Ratner et al., 2017). The distinguishing feature of our setting is that the proxy-target gap is *structured and directional* rather than approximately unbiased noise — which is precisely the condition under which naive fine-tuning is most likely to overcorrect, as we show empirically.

Instability of small-sample adaptation. That fine-tuning a pretrained model on a small dataset is unstable is documented for language models (Mosbach et al., 2021) and connects to the broader catastrophic-forgetting literature (Kirkpatrick et al., 2017): gradient updates on a small, non-representative batch can overwrite structure learned from a much larger pretraining distribution. Our contribution relative to this literature is not to reintroduce instability as a concept, but to show a *specific, class-conditional signature* of that instability in a discrete, imbalanced target label space — the model does not merely become noisier, it systematically stops predicting an entire contiguous region of the label space. To our knowledge this class-collapse signature has not been previously connected to the proxy-to-preference transfer setting specifically, where it interacts with (and can be mistaken for) the proxy's own directional bias.

The thermal-comfort domain (van Hoof, 2008) is the concrete testbed through which we study this more general transfer problem: using abundant, structured, but imperfect supervision to predict scarce human preference labels, and understanding precisely how naive adaptation fails when it does.

3 Problem Definition

3.1 Formal Setting

We consider a source domain of synthetic multimodal samples $(x_s^{\text{env}}, x_s^{\text{text}}, y_s^{\text{proxy}})$ and a target domain of real multimodal samples $(x_t^{\text{env}}, x_t^{\text{text}}, y_t^{\text{human}})$. In our instantiation, y_s^{proxy} is PMV (continuous) and y_t^{human} is TSV, a value in a small ordered discrete label space $L = \{-2, -1, 0, 1, 2, 3\}$.

We take as given (and verify empirically in Section 4) that y_s^{proxy} and y_t^{human} are correlated but not identical: there exists a latent human-environment interaction that both partially reflect, but the map between them, $g : y^{\text{proxy}} \to y^{\text{human}}$, is not the identity and is not even approximately unbiased — it has a direction (PMV under-predicts warm states) and depends on context (the gap varies by operation regime). We call this **structured label mismatch** to distinguish it from the i.i.d.-noise mismatch typically assumed when a proxy is used as a noisy substitute for a target.

Adaptation proceeds by fitting a small target-labeled set $D_{\text{adapt}} = \{(x_{t,i}^{\text{env}}, x_{t,i}^{\text{text}}, y_{t,i}^{\text{human}})\}_{i=1}^{n}$, with n small (here $n \in \{64, 128\}$), drawn as a **chronological prefix** of the target trajectory rather than an i.i.d. sample — again mirroring realistic deployment logging, where labeled feedback accumulates over time rather than being drawn uniformly from the eventual deployment distribution. The empirical label distribution of D_{adapt},

call it $P_{\text{adapt}}(y)$, can therefore differ substantially from the population label distribution $P_{\text{target}}(y)$ realized over the full deployment trajectory.

3.2 Tail-Class Collapse: A Falsifiable Hypothesis

We hypothesize that when (i) $|D_{\text{adapt}}|$ is small relative to $|L|$ and to the model's parameter count, (ii) $P_{\text{adapt}}(y)$ is a skewed or truncated version of $P_{\text{target}}(y)$ (as chronological, non-i.i.d. sampling tends to produce), and (iii) L is discrete and ordinal, gradient-based adaptation will preferentially fit the modes of $P_{\text{adapt}}(y)$ and will not merely add uniform noise to predictions for classes underrepresented or absent in D_{adapt} — it will systematically stop predicting them. Formally, writing f_{adapt} for the adapted model and f_{zero} for the zero-shot (proxy-only) model, we predict:

$$\text{range}(f_{\text{adapt}}(X_{\text{test}})) \subseteq [\min(\text{supp}(P_{\text{adapt}})), \max(\text{supp}(P_{\text{adapt}}))] + \epsilon,$$

i.e. the adapted model's predictive range contracts toward the support of the adaptation label distribution, even when the true test-label support is wider — and this contraction, not a uniform increase in per-example error, is what drives aggregate MAE degradation.

This is a specific enough claim to be falsified: it predicts *which* examples the adapted model will do worst on (those whose true label lies outside P_{adapt}'s support) and *why* aggregate exact accuracy can rise even as aggregate MAE falls (the model becomes very accurate on the classes it now exclusively predicts, at the cost of catastrophic error on the classes it no longer predicts at all). Section 6.6 tests this hypothesis directly against saved model predictions.

The transfer problem is hard for three reasons, restated with this hypothesis in mind:

1. **Label mismatch**: PMV is a structured physical approximation, while TSV is a subjective human response — the mismatch has direction and structure, not just variance.
2. **Distribution shift**: the source domain is synthetic and multi-building; the target domain is a real, single-trajectory, chronologically-ordered log, so D_{adapt} is not an i.i.d. sample of the deployment distribution.
3. **Modal mismatch**: objective sensors and subjective language carry overlapping but non-equivalent information about the latent human state.

The goal is to learn a function that benefits from abundant source structure while remaining faithful to the target human label space *in its entirety*, including classes the small adaptation window happens not to cover.

4 Data

4.1 Synthetic Source Domain

The synthetic dataset contains 141,330 multimodal records derived from EnergyPlus simulations and text rendering, covering 47,111 base physical states, 5,999 buildings, and four building prototypes: office, residential, retail, and school. Each sample contains environmental and personal variables including t_a, t_r, rh, vel, clo, and met, together with PMV, PPD, and a text description of the same state. Its scientific value is not just size — it provides a large, controllable source domain with paired physical state and language, consistent proxy supervision, and cross-building diversity, enabling representation learning under structured supervision before the model sees any real human TSV label.

4.2 Real Target Domain

The real dataset contains 279 chronological records from a single building trajectory: environmental variables, human TSV annotations, outdoor temperature and humidity, and operation states (door, window, air-conditioning). This is small, but it is the actual human label of interest, and — critically for Section 3.2

Table 1: Dataset summary.

Dataset	Records	Label	Domain	Distinctive value
Synthetic source	141,330	PMV / PPD	Multi-building simulation	Scale, diversity, structured proxy supervision
Real target	279	TSV	Single real trajectory	Human labels, chronology, operation context

— its chronological ordering means that any fixed-size training prefix realizes a specific, non-representative slice of the label distribution rather than a random sample of it (Section 6.6 quantifies exactly how non-representative).

The original human feedback was in Chinese; the English text used here is translated and partially humanized. We treat the English text as a controlled semantic rendering of the original feedback, not as a native free-form English corpus, and flag this explicitly rather than presenting it as naturalistic language data (Section 8).

4.3 Motivation Statistics

- PMV range: $[-2.42, 0.98]$; TSV range: $[-2, 3]$
- mean PMV: -0.9706; mean TSV: 0.2079
- direct PMV-to-TSV MAE: 1.4064; Pearson correlation: 0.2058
- rounded PMV exact agreement with TSV: 0.147
- under (window= 1, door= 1, ac= 0): mean PMV -2.197 vs. mean TSV 0.795 (gap 2.992)

These statistics establish that the mismatch is structured and directional before any model is trained — the paper's transfer experiments (Section 6) test what a model does with that structure, not whether it exists.

5 Method

5.1 Overview

Our model follows a multimodal neuro-symbolic design: a text encoder processes tokenized comfort descriptions, a sensor encoder processes low-dimensional environmental features, a fusion module maps the combined representation into a latent physical state, and a differentiable physics layer predicts PMV and PPD from that latent state. Transfer to TSV attaches a small target head on top of the pretrained backbone. This architecture is not the paper's contribution — it is a minimal instrument for producing a model whose zero-shot behavior is genuinely proxy-grounded (via the differentiable physics layer, not just correlated with PMV through shared inputs), so that the adaptation phenomenon studied in Section 6 is attributable to the label-mismatch mechanism in Section 3.2 rather than to an idiosyncratic architecture choice.

5.2 Backbone Encoder

A compact Transformer-based model of roughly five million parameters: (1) a SentencePiece tokenizer converts text to subword tokens; (2) a four-layer bidirectional Transformer text encoder with sinusoidal positional encoding and `[CLS]` pooling; (3) a Fourier-feature sensor encoder maps continuous sensor values into a higher-dimensional representation; (4) a fusion Transformer combines the sensor and text streams and predicts the mean and log-variance of a six-dimensional latent physical state (temperature, humidity, air speed, clothing, metabolic rate, and one auxiliary dimension).

5.3 Differentiable Physics Layer

The latent physical state passes through a differentiable Fanger model to produce PMV and PPD, turning physics into part of the computation graph rather than a post hoc baseline. This is what makes PMV a *structured* proxy in this pipeline rather than an arbitrary correlated feature.

5.4 Two-Stage Synthetic Pretraining

Phase 1 trains the encoder to recover latent physical parameters under a Gaussian negative log-likelihood objective. Phase 2 refines the model toward PMV prediction while retaining the latent physical objective and adding a weak monotonicity penalty encouraging physically sensible gradients with respect to temperature and humidity. This two-stage design is the current instantiation used to study proxy-to-preference transfer, not a claimed optimum for proxy-supervised pretraining in general.

5.5 Transfer to Human TSV

A small prediction head is attached to the pretrained backbone, taking the concatenation of the latent mean representation and the backbone's PMV estimate as input. We evaluate regression and classification variants; the main story is driven by the regression setting. The transfer procedure supports freezing/unfreezing the backbone, letting us test whether direct adaptation helps, whether freezing preserves source structure, and how much instability is introduced when small target sets are used for full fine-tuning — the mechanism formalized in Section 3.2 and tested in Section 6.6.

5.6 Splits and Protocol

Synthetic pretraining uses grouped splitting by (building_id, hour) rather than record-level random splitting, reducing leakage across alternative views of the same physical state. Real-data evaluation uses **chronological** splits to preserve trajectory structure — a deliberate choice consistent with realistic deployment (labels accumulate over time), and the choice responsible for the non-representative $P_{\text{adapt}}(y)$ discussed in Section 3.2.

Main real-data experiments use train size 64 / val 32 and train size 128 / val 32, with the remainder used as test data, giving splits 64/32/183 and 128/32/119. We report both because they probe different data-scarcity regimes.

5.7 Tasks

1. **Zero-shot proxy transfer**: use the pretrained PMV-producing backbone without real TSV adaptation, mapping PMV to TSV via nearest-label rounding.

2. **Few-shot regression transfer**: adapt a TSV regression head using small real training sets.

3. **Few-shot classification transfer**: adapt a TSV classification head over the discrete TSV label space.

5.8 Metrics

We report MAE, RMSE, exact accuracy, and within-one accuracy, and — following the finding in Section 6.6 — argue that **per-class** breakdowns of these metrics should be treated as a required reporting field, not an optional appendix, whenever the adaptation set is small and non-i.i.d.

5.9 Reproducibility Details

Tokenizer and text processing: SentencePiece, vocabulary 4096, max sequence length 64, shared tokenizer across synthetic and real data.

Backbone architecture: text encoder 4 layers / hidden 256 / 4 heads / feed-forward 512; fusion encoder 2 layers / hidden 256 / 4 heads / feed-forward 512; latent physical state dimension 6; ∼5M parameters.

Synthetic pretraining: phase 1 — 10 epochs, batch 256; phase 2 — 10 epochs, batch 128, lr 5×10^{-5}; optimizer AdamW; synthetic split grouped by (building_id, hour); sensor mode for reported transfer runs: `real_context`.

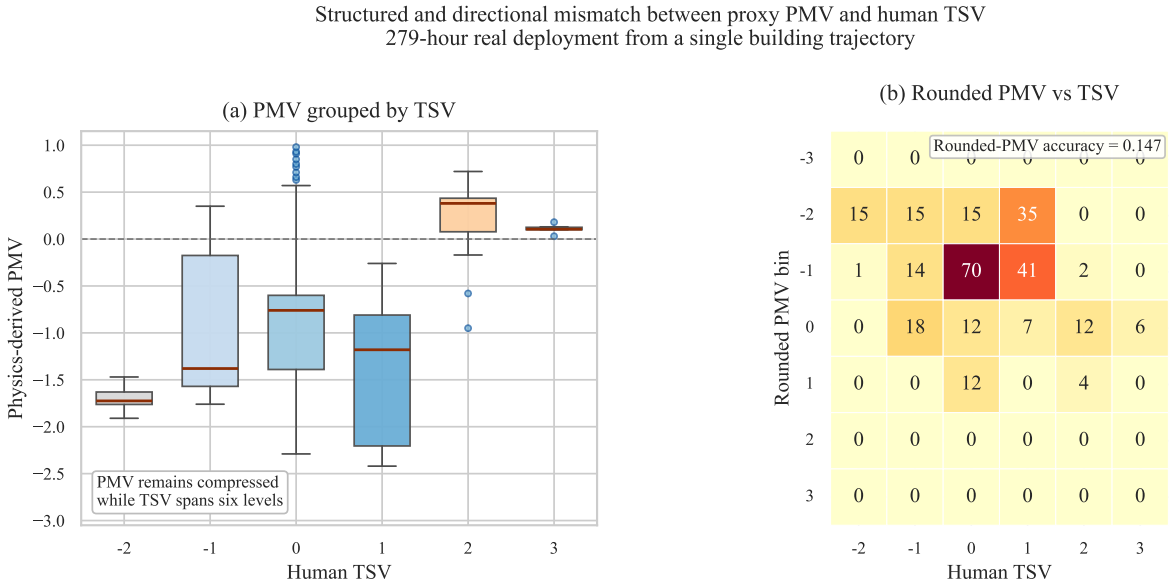

Figure 1: Structured and directional mismatch between physics-derived proxy labels and human preference labels in a 279-hour real deployment. Left: grouped by human TSV, the PMV distribution remains compressed and shifts weakly across preference levels. Right: after PMV discretization, agreement with TSV is low (0.147 exact accuracy), with error mass falling systematically off-diagonal. This motivates transfer rather than substitution — but not yet the specific failure mode of Section 6.6.

Real-data adaptation: 20 epochs, batch 16, optimizer AdamW, default transfer lr 2×10^{-4}, validation model selection by validation MAE; freeze-then-unfreeze schedule with `freeze_backbone_epochs=10` for the 128-sample baseline.

Randomness and implementation: seed 42; device CUDA in reported runs; chronological splits deterministic once seed and train/validation sizes are fixed.

5.10 Release and Benchmark Plan

We plan to publicly release the dataset package and benchmark code, containing: the synthetic source corpus, the real target trajectory, tokenizer/preprocessing scripts, reference training/evaluation scripts, fixed chronological split definitions (64/32/183, 128/32/119), and metric/reference-table utilities, framed around three tasks — **Task A** (zero-shot proxy transfer), **Task B** (few-shot preference adaptation), **Task C** (modality ablation). We recommend benchmark submissions report at minimum MAE, RMSE, exact accuracy, within-one accuracy, **per-class MAE and exact accuracy** (new, motivated by Section 6.6), train/validation/test sizes, whether pretraining was used, and whether text was translated/humanized/normalized.

6 Results

6.1 Structured Mismatch Before Training

Figure 1 is the main motivation figure.

Table 2: Transfer results across the five main experiments.

Experiment	Mode	Train	Zero MAE	Tuned MAE	Zero Exact	Tuned Exact	Zero W1	Tuned W1
`transfer_64`	Regression	64	1.0437	1.6263	0.3005	0.1858	0.6995	0.5956
`transfer_64_frozen`	Regression	64	1.0437	2.0701	0.3005	0.1421	0.6995	0.4809
`transfer_64_cls`	Classification	64	1.0437	2.1639	0.3005	0.0874	0.6995	0.2459
`transfer_128_reg`	Regression	128	0.9916	1.5388	0.3109	0.3361	0.6975	0.6050
`transfer_128_ctx`	Regression	128	1.1092	1.6018	0.2185	0.3866	0.6723	0.5798

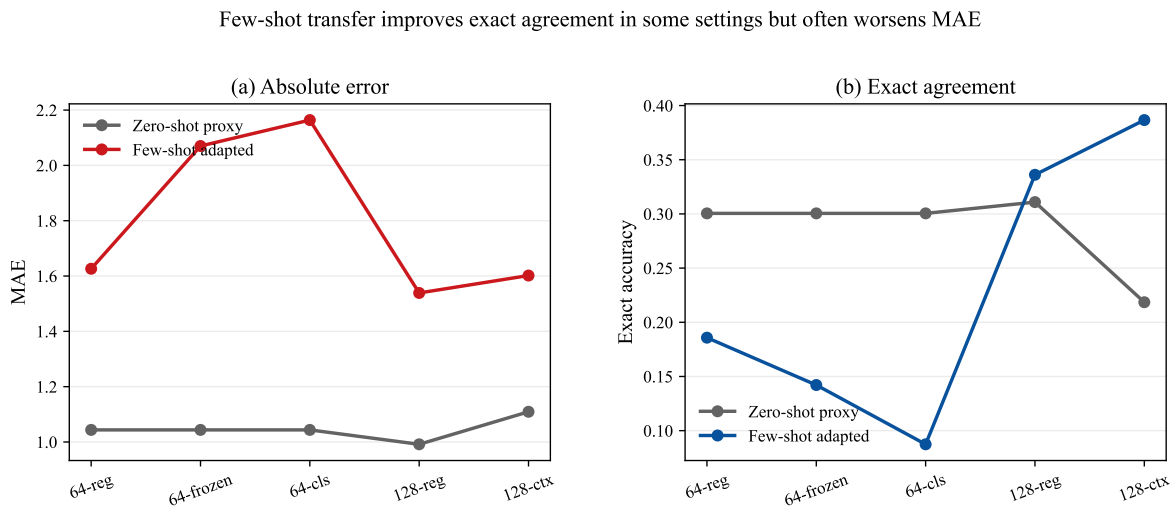

Figure 2: Few-shot transfer improves exact agreement in some settings but often worsens MAE, and adapted models consistently underperform the zero-shot baseline in MAE across current experiments even as exact agreement improves, especially in `128-reg` and `128-ctx`. This divergence is the aggregate symptom; Section 6.6 identifies its mechanism.

6.2 Transfer Results Table

6.3 Main Empirical Pattern (Aggregate View)

The dominant pattern is not "fine-tuning consistently improves performance" — often the opposite is true for error-based metrics. Zero-shot proxy transfer is already strong (MAE 0.99–1.11, within-one accuracy 0.67–0.70). Few-shot adaptation can improve exact agreement — most clearly `transfer_128_ctx`, where exact accuracy rises from 0.2185 to 0.3866 — but the same adaptation often worsens MAE (`transfer_128_reg`: 0.9916 → 1.5388; `transfer_128_ctx`: 1.1092 → 1.6018). Small-data settings with freezing or classification are particularly unstable.

6.4 Baseline Comparison on the 128/32 Split

A sensor-only model is the weakest learned baseline and does not outperform the PMV proxy. A multimodal model trained from scratch beats the proxy in MAE only marginally. The pretrained multimodal transfer model achieves the lowest MAE (0.8278) with strong exact agreement (0.4622). The text-only baseline is unexpectedly strong, slightly exceeding the pretrained multimodal model in exact accuracy (0.4706 vs 0.4622) — evidence that the current real-data text channel carries a high fraction of target-relevant signal, plausibly related to its translated/regularized style (Section 8).

Table 3: Baseline comparison on the chronological 128/32/119 split.

Method	MAE	RMSE	Exact	Within-1
PMV proxy zero-shot	1.1092	1.3284	0.2185	0.6723
Sensor-only scratch	1.2078	1.4930	0.2185	0.6555
Text-only scratch	0.8782	1.2453	0.4706	0.8655
Multimodal scratch	1.0538	1.3131	0.3025	0.6723
Multimodal pretrained	0.8278	1.1188	0.4622	0.8655

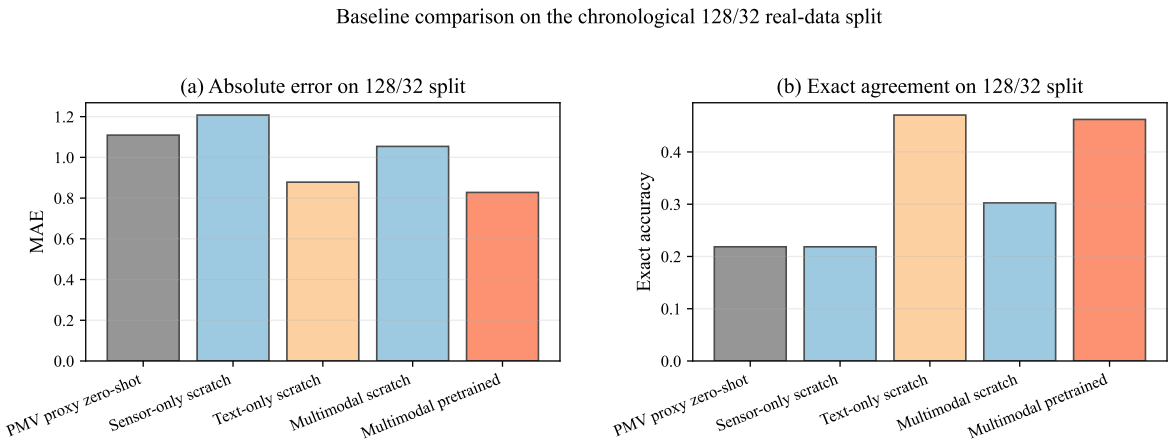

Figure 3: Baseline comparison on the 128/32 split. The pretrained multimodal transfer model achieves the lowest MAE; the text-only baseline outperforms both the PMV proxy and the sensor-only model, underscoring that language carries highly target-aligned information here.

6.5 Supplemental Baseline Comparison on the 64/32 Split

Under a smaller adaptation budget the ranking changes: the pretrained multimodal model is no longer best and degrades substantially, while the text-only baseline remains strongest among learned models. This reinforces that transfer under label mismatch is not monotonically improved by source-domain pretraining — when real supervision is very scarce, preserving or adapting proxy-induced structure becomes unstable, consistent with the mechanism identified next.

6.6 Tail-Class Collapse: Testing the Hypothesis

Sections 6.1–6.6 establish *that* adaptation is unstable in aggregate. This subsection tests the specific mechanism hypothesized in Section 3.2 directly against the saved model predictions for the 128/32/119 split (regression head; the context-augmented head shows the same pattern). No retraining was performed; this is a re-analysis of existing predictions.

This is not a uniform degradation: the rounded fine-tuned predictions are drawn only from $\{0, 1, 2, 3\}$ (counts 27 / 58 / 30 / 4 across the 119-record test set) — **zero** predictions land on -2 or -1, even though those classes are 45/119 (38%) of ground truth, and raw regression outputs stay almost entirely non-negative (minimum ≈ -0.38). This confirms the Section 3.2 prediction: the adapted model's predictive range contracts toward the support of $P_{\text{adapt}}(y)$ (the 128-record chronological training prefix, which under-samples the cold tail relative to the full trajectory), not toward uniformly noisier predictions.

The direction is the striking part: PMV itself is *cool-biased* (Section 4), but the adapted model is *warm-biased* — small-sample adaptation does not correct the proxy's bias, it overshoots past the target distribution in the opposite direction, driven by which classes happen to appear in the adaptation window rather than by the true target distribution. Consistent with this account, splitting the test set by operation regime shows adaptation

Table 4: Baseline comparison on the chronological 64/32/183 split.

Method	MAE	RMSE	Exact	Within-1
PMV proxy zero-shot	1.0437	1.3570	0.3005	0.6995
Sensor-only scratch	1.1481	1.3898	0.2404	0.7432
Text-only scratch	0.9546	1.1897	0.4208	0.7814
Multimodal scratch	1.1155	1.3558	0.2350	0.7377
Multimodal pretrained	1.4415	1.8461	0.2022	0.6503

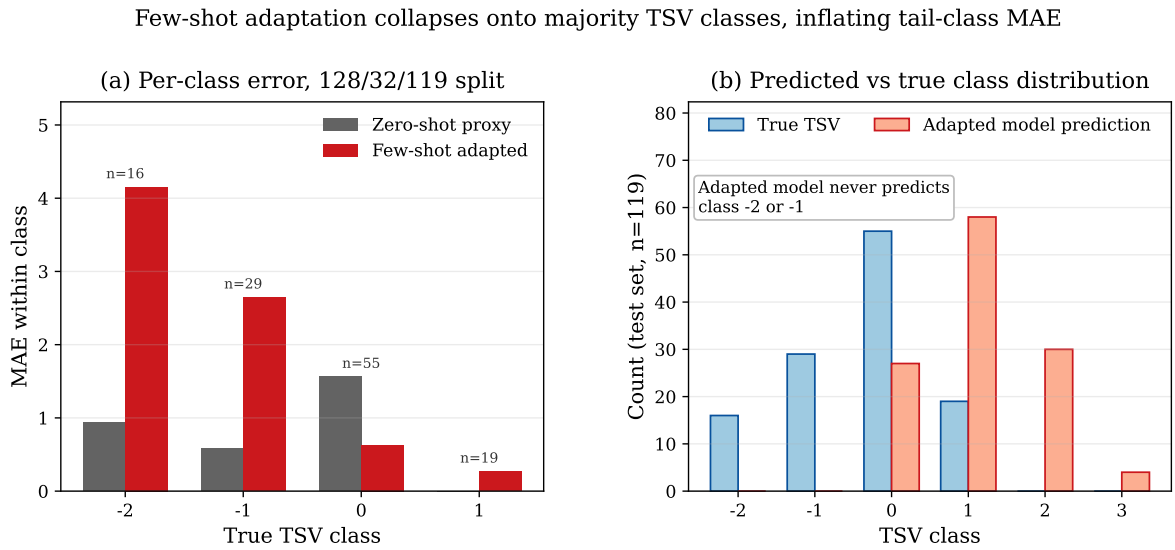

Figure 4: (a) Per-class MAE, zero-shot proxy vs. few-shot adapted, on the 128/32/119 test split. Adaptation *reduces* error on classes 0 and 1 — the classes that dominate the small 128-record adaptation window — while *increasing* it by 3–7× on classes −2 and −1. (b) Predicted-class distribution (adapted model) vs. true-class distribution on the same 119 test records: the adapted model's predictions never fall below TSV 0.

helps most in the small AC-on regime (window= 0, door= 0, ac= 1, $n = 21$, concentrated in classes 0/1: MAE drops from 1.19 to 0.25) and hurts most in the dominant window-open regime (window= 1, door= 0, ac= 0, $n = 95$, containing most of the cold-tail records: MAE rises from 0.98 to 1.86) — and splitting by $|\text{TSV} - \text{raw PMV}|$ disagreement magnitude shows adaptation *helps* on the "hard" 24-record slice where the proxy is most wrong (MAE 0.58 → 0.44) and *hurts* on the "easy" 95-record slice (MAE 1.10 → 1.82), the reverse of the intuitive expectation. Both breakdowns are explained by the same underlying mechanism — regime and disagreement-magnitude are confounded with TSV class in this single chronological trajectory — rather than being three independent phenomena.

6.7 Interpreting the Tension

Synthetic pretraining does learn transferable structure — if it did not, zero-shot proxy performance would be weak across the board, and it is not. What Section 6.6 adds is that the instability introduced by adaptation is not generic overfitting noise; it has a specific, predictable shape determined by the label distribution of the adaptation window relative to the deployment distribution. This reframes the paper's central claim from a domain-adaptation statement ("the target differs in distribution") to a proxy-to-preference-specific one: **the target differs in what the label means, and naive adaptation on a small, non-representative label sample will silently narrow that meaning to whatever the adaptation window happened to contain.**

Table 5: Per-class MAE and exact accuracy, 128/32/119 split, regression head.

TSV class	n (test)	Zero-shot MAE	Tuned MAE	Zero-shot Exact	Tuned Exact
−2	16	0.938	**4.153**	0.312	0.000
−1	29	0.586	**2.648**	0.414	0.000
0	55	1.564	0.632	0.018	0.436
1	19	0.000	0.270	1.000	0.842

7 Discussion

7.1 Why the 141k Synthetic Dataset Matters

The 141,330-example synthetic corpus is valuable because it provides a large, structured source domain for learning under proxy supervision, anchoring the representation to a physical geometry of comfort-relevant states. Without it, the 279-record real dataset would be too small to support meaningful multimodal representation learning at all — but Section 6.6 shows that this pretrained structure is exactly what naive adaptation is at risk of overwriting for the classes the adaptation window under-samples.

7.2 What the Results Suggest About Proxy Labels, Restated Generally

Physics-derived proxy labels are useful as structure, scale, and inductive bias — not as a target substitute. Transferred to a target domain with real human annotations, they support surprisingly strong zero-shot behavior. But the adaptation procedure that is supposed to correct the proxy's known bias can, under the three conditions of Section 3.2 (small n, unrepresentative $P_{\text{adapt}}(y)$, discrete imbalanced L), instead introduce a *new*, opposite-direction bias driven by which labels happened to appear early in the deployment log. This is a general risk for any pipeline that pretrains on a structured proxy and fine-tunes on a small, temporally- or otherwise-non-randomly-sampled slice of the true target — not specific to thermal comfort, PMV, or this architecture.

7.3 Operation Regime and Disagreement Magnitude Are Confounded with TSV Class

A natural next question is whether operation context (window/door/AC) or proxy-disagreement magnitude *independently* predicts where adaptation helps or hurts. Section 6.6 shows they correlate with the outcome, but in this single 279-hour trajectory they are confounded with TSV class itself. We flag this explicitly as a limitation of having one real trajectory (Section 8) rather than presenting regime or disagreement effects as independent findings.

7.4 Why Language May Be the Bridge

Language captures perceived experience in a way that may be closer to human TSV than raw physical variables alone (the text-only baseline results in Sections 7.5–7.7 are suggestive, if not yet conclusive given the translated/humanized text channel). The framing is that language may help mediate between structured physical proxies and subjective human labels precisely because it is not constrained by the same physical computation that produces the proxy label.

8 Limitations and Future Work

8.1 Limitations

1. The real target dataset is small and drawn from a single building trajectory; regime-level and disagreement-magnitude findings (Sections 6.6, 7) are confounded with TSV class within this one trajectory and should not be read as independent effects until replicated on a second real trajectory.
2. The English real text is translated and partially humanized, not native free-form occupant language.

3. The synthetic text is generated from structured simulated states and may encode stylistic regularities absent in natural feedback.
4. The benchmark is reproducible but narrow: one real trajectory is not enough to claim broad cross-building generalization of either the zero-shot results or the tail-class-collapse phenomenon, though the phenomenon's *mechanism* (Section 3.2) is stated generally enough to be tested on other proxy-to-preference datasets.
5. Some large PMV-TSV mismatch cases, especially under high air velocity, deserve further physical sanity-checking.
6. The real test suffix realizes only three operation-state combinations with sufficient support ($n \geq 3$), limiting the statistical power of the regime-level analysis specifically.

8.2 Future Work

1. Test the Section 3.2 hypothesis on a second real proxy-to-preference dataset outside thermal comfort, to check whether tail-class collapse is specific to this domain or general, as the mechanism argument predicts.
2. Evaluate adaptation strategies designed to counter tail-class collapse directly: class-balanced fine-tuning objectives, ordinal regression with label-distribution regularization, consistency regularization between the PMV head and the TSV head to bound how far adaptation can move predictions from the proxy-informed prior, and linear probing vs. full fine-tuning as a function of adaptation-set label coverage.
3. Report per-class metrics as a standard, not optional, benchmark field.
4. Expand the target evaluation to additional real TSV datasets/trajectories to break the regime/class confound identified in Section 7.
5. Test whether language improves robustness specifically in the mismatch-heavy, tail-class slices where PMV is most biased and where adaptation currently fails most.

9 Conclusion

This paper studies proxy-to-preference transfer — learning from abundant structured proxy labels to predict scarce human preference labels — using thermal sensation inference as a controlled testbed. We show the proxy-target mismatch (PMV vs. TSV) is structured, directional, and regime-dependent, and that a model pretrained on proxy supervision transfers surprisingly well zero-shot. The paper's central contribution is showing *why* naive few-shot adaptation is unstable in this setting: not generic small-sample noise, but **tail-class collapse** — the adapted model's predictive range contracts to the label support of a small, chronologically-sampled adaptation window, silently dropping coverage of classes that window under-samples, in a direction that can oppose the proxy's own known bias. We state this as a general, falsifiable hypothesis (Section 3.2), confirm it directly against model predictions (Section 6.6), and argue it is a structural risk for any proxy-to-preference pipeline with a small, non-representative adaptation set and a discrete, imbalanced target label space — motivating per-class, tail-coverage-aware evaluation as a requirement for this class of benchmark, not an optional diagnostic.

Broader Impact Statement

This work studies a general machine-learning failure mode using thermal-comfort data as a testbed; it does not itself deploy a comfort-control system. Practitioners building proxy-to-preference pipelines in higher-stakes domains (e.g., using simulator-derived or heuristic reward proxies with small amounts of human feedback) should be aware that naive fine-tuning on a small, non-representative feedback sample can silently narrow a model's effective label coverage in a way that aggregate metrics do not reveal, and should evaluate with tail-coverage-aware metrics before deployment.

References

ASHRAE. Ansi/ashrae standard 55-2020: Thermal environmental conditions for human occupancy. Technical report, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2020.

Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, 2019.

P. O. Fanger. *Thermal Comfort: Analysis and Applications in Environmental Engineering*. Danish Technical Press, Copenhagen, 1970.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning (ICML)*, 2015.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

Marius Mosbach, Maksym Andriushchenko, and Dietrich Klakow. On the stability of fine-tuning BERT: Misconceptions, explanations, and strong baselines. In *International Conference on Learning Representations (ICLR)*, 2021.

Sergey I. Nikolenko. Synthetic data for deep learning. *arXiv preprint arXiv:1909.11512*, 2019.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, 2021.

Alexander Ratner, Stephen H. Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. Snorkel: Rapid training data creation with weak supervision. In *Proceedings of the VLDB Endowment*, 2017.

Sebastian Ruder. An overview of proxy-label approaches for semi-supervised learning. `https://www.ruder.io/semi-supervised/`, 2018. Blog post.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Joost van Hoof. Forty years of Fanger's model of thermal comfort: comfort for all? *Indoor Air*, 18(3): 182–201, 2008.

Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 2018.

A Additional Notes

Per-class, per-regime, and disagreement-slice analysis code, the class-collapse figure generation script, and the full candidate reference list with confidence annotations are available in the accompanying code release.