Bisimulation-Grounded World Models: Scaling Abstract Representations Across Domains

Anonymous Author(s)

ABSTRACT

World models that reconstruct observations are forced to retain all perceptual detail, including task-irrelevant information, leading to representations that scale with observation complexity rather than world complexity. We propose the Bisimulation-Grounded World Model (BGWM), which replaces reconstruction with a bisimulation distance regression objective that trains encoders to produce compact abstract states capturing only behaviorally relevant structure. BGWM combines a forward prediction loss in latent space, a pairwise bisimulation distance loss that enforces behavioral distance matching, and a variational information bottleneck for compression. We evaluate BGWM against reconstruction-based and forward-prediction-only baselines across three synthetic domains with controlled relevant and irrelevant state dimensions, using 3 random seeds per condition with shared training data for fair comparison. On the grid navigation domain, BGWM achieves a mean abstraction ratio of 0.807 ± 0.097 compared to 1.871 ± 0.059 for reconstruction, a 2.3× improvement. A linear probe analysis shows that BGWM encodes significantly less irrelevant information ($R^2 = 0.438$) than reconstruction ($R^2 = 0.830$) while retaining relevant structure. We also find that BGWM does not improve over baselines on the linear dynamics domain, which we analyze as a limitation of the pairwise distance approximation to the true bisimulation metric. Cross-domain transfer experiments with four encoder baselines (BGWM, reconstruction, forward-only, and random) show that the BGWM encoder achieves 4.0–5.1× error reduction when adapting only the dynamics model, evaluated on held-out data. These results demonstrate that bisimulation-grounded learning produces abstract representations that discard task-irrelevant detail in nonlinear domains, while revealing important failure modes in linear settings.

1 INTRODUCTION

Human mental models of the world operate on compact, abstract representations that discard perceptual detail irrelevant to the task at hand [15]. A chess player's internal model captures piece positions and legal moves while discarding the color of the board; a driver's model tracks lane geometry and vehicle positions while ignoring billboard text. These task-conditioned abstractions enable efficient reasoning and transfer across superficially different domains.

Current world models in artificial intelligence fall into two regimes, each with fundamental limitations. Pixel-reconstructive models, such as the Dreamer family [9, 10], learn latent representations by requiring an observation decoder. Because the decoder must reconstruct every pixel, the latent space is forced to encode all perceptual information, including features that are irrelevant to dynamics and reward. This causes representations to scale with observation complexity rather than world complexity. Language-only world models provide natural abstraction through discrete tokens but cannot directly represent continuous physics, spatial layouts, or non-linguistic signals.

The core challenge is to learn world-model representations that are compact and abstract like language but grounded in continuous perception. Two sub-problems arise: (1) defining a formal abstraction criterion that discards irrelevant detail while retaining task-relevant structure, and (2) scaling such representations across qualitatively different domains without domain-specific engineering.

We address these sub-problems with the Bisimulation-Grounded World Model (BGWM), which builds on bisimulation theory from the state abstraction literature [1, 6, 11]. Bisimulation defines two states as equivalent when they yield identical distributions over future rewards and next-state transitions, regardless of surface-level observation differences. We operationalize this principle through a pairwise distance regression loss that enforces latent distances to match behavioral distances, combined with a variational information bottleneck [3, 13] and a forward prediction loss in the abstract space. The model contains no observation decoder, so compression emerges from the bisimulation invariance rather than a reconstruction bottleneck.

Our experimental evaluation addresses key methodological concerns: all methods train on identical shared datasets (eliminating data confounds), results are reported with mean and standard deviation across 3 seeds, and we introduce a scale-invariant linear probe metric alongside the sensitivity-based abstraction ratio. We also provide transfer baselines for all encoder types and evaluate on held-out data.

1.1 Related Work

State Abstraction Theory. Bisimulation metrics [6] and MDP homomorphisms [11] provide the mathematical foundation for defining when two states are behaviorally equivalent. Abel et al. [1] extended this to approximate abstractions with bounded value loss. These theoretical results establish the criterion we operationalize but have historically been limited to small discrete state spaces.

Bisimulation-Based Representation Learning. Zhang et al. [16] introduced Deep Bisimulation for Control (DBC), which learns representations where latent distance corresponds to behavioral similarity. Gelada et al. [7] proposed DeepMDP with similar goals. Castro [4] developed scalable bisimulation computation methods, and Agarwal et al. [2] applied contrastive behavioral similarity embeddings for generalization. These methods demonstrate the effectiveness of bisimulation for single-domain settings but have not been evaluated for cross-domain transfer with proper baselines.

Information-Theoretic Representation Learning. The Information Bottleneck [13] formalizes the compression-relevance trade-off. Alemi et al. [3] introduced the variational information bottleneck

for deep networks. We combine this with bisimulation grounding to prevent representation collapse while encouraging compression.

World Models and Contrastive Learning. Modern world models [9, 10] achieve strong performance through observation reconstruction. Contrastive learning methods [5, 8] learn representations without reconstruction but optimize for general-purpose features rather than task-relevant abstractions. Discrete tokenization approaches [14] force compression through codebooks but target reconstruction fidelity. Our work combines bisimulation distance regression (task-relevant invariance) with information bottleneck (explicit compression) in a decoder-free architecture.

2 METHODS

2.1 Problem Formulation

Consider an environment with state $s = (s_{\text{rel}}, s_{\text{irr}}) \in \mathcal{S}$ where s_{rel} affects dynamics and reward while s_{irr} is dynamically independent. Observations $o = g(s)$ are generated by a nonlinear mixing function that entangles both components. The goal is to learn an encoder $E : \mathcal{O} \to \mathcal{Z}$ such that the abstract state $z = E(o)$ retains information about s_{rel} and discards information about s_{irr}.

2.2 Architecture

The BGWM architecture consists of three components:

Modality Encoder. A three-layer MLP with LayerNorm and GELU activations maps observations $o \in \mathbb{R}^{32}$ into embeddings $h \in \mathbb{R}^{64}$.

Abstraction Bottleneck. A variational layer compresses embeddings into abstract states $z \in \mathbb{R}^d$ (default $d = 8$). During training, stochastic noise from a learned variance acts as an implicit information bottleneck, with KL divergence from a standard normal prior providing the compression signal.

Latent Dynamics Model. A two-layer MLP predicts the next abstract state $\hat{z}_{t+1}$ and reward $\hat{r}_t$ from (z_t, a_t).

2.3 Training Objective

The total loss combines four terms:

$$\mathcal{L} = \mathcal{L}_{\text{fwd}} + \alpha \mathcal{L}_{\text{bisim}} + \lambda \mathcal{L}_{\text{reward}} + \beta \mathcal{L}_{\text{KL}} \tag{1}$$

Forward Prediction Loss. MSE between the predicted next latent state and the encoded next observation: $\mathcal{L}_{\text{fwd}} = \|\hat{z}_{t+1} - \text{sg}[E(o_{t+1})]\|^2$, where sg$[\cdot]$ denotes stop-gradient.

Bisimulation Distance Loss. For each pair (i, j) in a batch, the behavioral distance is $d_{\text{behav}}(i, j) = |r_i - r_j| + \gamma \|z'_i - z'_j\|_2$. The loss enforces $\|z_i - z_j\|_2 \approx d_{\text{behav}}(i, j)$ via smooth L_1 loss scaled by temperature $\tau = 0.1$. We note that this is a *pairwise distance regression* rather than an InfoNCE-style contrastive loss; the bisimulation target uses single-sample next-state distances as an approximation to the Wasserstein distance between transition distributions.

Reward Prediction Loss. MSE on scalar reward: $\mathcal{L}_{\text{reward}} = \|\hat{r}_t - r_t\|^2$.

Information Bottleneck Loss. $\mathcal{L}_{\text{KL}} = \text{KL}(q(z|o) \| \mathcal{N}(0, I))$.

Hyperparameters: $\alpha = 1.0$, $\lambda = 0.5$, $\beta = 0.01$, $\gamma = 0.99$. We train for 40 epochs with AdamW [12] (learning rate 5×10^{-4}, weight decay 10^{-5}) and cosine annealing.

2.4 Baselines

Reconstruction World Model. Standard autoencoder with MSE reconstruction loss plus forward prediction and reward losses. The decoder forces the latent to retain all observation information.

Forward-Only World Model. Same encoder architecture as BGWM but trained with only forward prediction and reward losses (no bisimulation, no stochastic bottleneck, deterministic encoder). This isolates the contribution of the bisimulation loss.

2.5 Evaluation Metrics

Abstraction Ratio (ρ). For each base state, we independently perturb the relevant and irrelevant dimensions by $\delta \sim \mathcal{N}(0, 0.5^2 I)$ and measure the resulting change in latent representation:

$$\rho = \frac{\text{Irrelevant Sensitivity}}{\text{Relevant Sensitivity}} \tag{2}$$

Lower values indicate better abstraction. However, this metric is scale-sensitive: a model with uniformly low sensitivity achieves a good ratio without necessarily encoding useful information.

Linear Probe R^2 (Scale-Invariant). We fit Ridge regression from z to s_{rel} and from z to s_{irr}, reporting R^2 for each. An ideal abstraction achieves high R^2_{rel} and low R^2_{irr}. Unlike the abstraction ratio, this metric is invariant to the scale of the latent representation.

Normalized Forward Prediction Error. Multi-step rollout in latent space compared to the encoder output at each future step, normalized by the standard deviation of the latent space to remove scale confounds.

Effective Rank. The exponential of the entropy of normalized singular values of the latent representation matrix.

Cross-Domain Transfer. We freeze the encoder from a source domain and train only a new dynamics model on a target domain, measuring adaptation speed on a held-out validation set (70/30 train/val split). We compare four encoder sources: BGWM, reconstruction, forward-only, and random (untrained) encoders.

2.6 Experimental Controls

Addressing methodological concerns from prior work, we implement three key controls:

(1) **Shared datasets.** For each seed, one dataset is collected and used by all three methods, eliminating confounds from different training trajectories.
(2) **Multi-seed evaluation.** All results are reported as mean ± standard deviation across 3 random seeds.
(3) **Run metadata.** All experiments save seed values, configuration, library versions, and timestamps for reproducibility.

3 EXPERIMENTAL SETUP

3.1 Synthetic Environments

We construct three environments with controlled relevant and irrelevant state dimensions:

- **Linear Dynamics**: 4 relevant dimensions (linear system $s' = As + Ba + \epsilon$) and 4 irrelevant dimensions (random walk). Observation dimension: 32.
- **Nonlinear Pendulum**: 2 relevant dimensions (angle and angular velocity with $\dot{\omega} = -\sin\theta + a$) and 6 irrelevant dimensions (sinusoidal drift). Observation dimension: 32.
- **Grid Navigation**: 2 relevant dimensions (position with soft-discretized dynamics) and 6 irrelevant dimensions (random perturbation). Observation dimension: 32.

All environments use a fixed random two-layer MLP as the observation function, entangling relevant and irrelevant state dimensions. We collect 80 trajectories of 30 steps each with random actions per seed.

Reward Functions. Each domain uses a reward function that depends only on the relevant state dimensions s_{rel}, ensuring the bisimulation-theoretic separation is well-defined:

- **Linear Dynamics**: $r = -\|s_{\text{rel}}\|^2$ (negative squared norm, encouraging state regulation).
- **Nonlinear Pendulum**: $r = -|\theta_{t+1}|$ (negative absolute angle, encouraging upright balance).
- **Grid Navigation**: $r = -\|s_{\text{rel}} - s_{\text{goal}}\|^2$ where $s_{\text{goal}} = (1, 1)$ (negative squared distance to goal).

Observation Function. Each domain uses its own fixed random two-layer MLP (seeded with domain seed +1000) to map state to observation. The observation functions are not shared across domains, so cross-domain transfer requires the encoder to extract domain-invariant structure despite different observation mappings.

Implementation Details. Batch size is 256. The bisimulation distance loss computes all $B^2 = 65{,}536$ pairwise distances within each batch (no subsampling). The bisimulation target $d_{\text{behav}}(i, j)$ is computed with stop-gradient on both the next-state encodings z'_i, z'_j and the reward predictions, providing a semi-gradient update to the encoder. Training uses gradient clipping (max norm 1.0) for BGWM and forward-only models.

Figure 1 provides an overview of the complete experimental framework and the relationships between its components.

4 RESULTS

4.1 Abstraction Quality

Table 1 presents the abstraction quality metrics across all three domains. Results are reported as mean ± standard deviation across 3 seeds.

BGWM achieves the best abstraction ratio on grid navigation (0.807 ± 0.097 vs. 1.871 ± 0.059 for reconstruction, a 2.3× improvement) and outperforms reconstruction on the nonlinear pendulum (1.734 vs. 1.950). However, on the linear dynamics domain, BGWM performs worst (1.958), which we discuss in Section 4.9.

The forward-only baseline achieves competitive ratios but at much lower absolute sensitivity (relevant sensitivity 0.54–0.68 vs. 1.33–3.27 for BGWM). To distinguish genuine abstraction from representation collapse, we turn to the linear probe analysis.

Table 1: Abstraction ratio ρ (lower is better) across domains. All methods trained on shared datasets per seed. Bold indicates best per domain.

Domain	Method	ρ (mean ± std)
Linear Dyn.	BGWM (Ours)	1.958 ± 0.105
	Reconstruction	1.349 ± 0.057
	Forward-Only	**1.329 ± 0.127**
Nonlinear Pend.	BGWM (Ours)	1.734 ± 0.300
	Reconstruction	1.950 ± 0.130
	Forward-Only	**1.242 ± 0.110**
Grid Nav.	BGWM (Ours)	**0.807 ± 0.097**
	Reconstruction	1.871 ± 0.059
	Forward-Only	0.883 ± 0.055

Table 2: Linear probe R^2 for relevant and irrelevant state recovery. Higher R^2_{rel} and lower R^2_{irr} indicate better abstraction. Bold indicates best R^2_{irr} (most irrelevant suppression) per domain.

Domain	Method	R^2_{rel}	R^2_{irr}
Linear	BGWM	0.220 ± 0.045	**0.649 ± 0.066**
	Reconstruction	0.742 ± 0.001	0.936 ± 0.011
	Forward-Only	0.287 ± 0.080	0.763 ± 0.061
Nonlinear	BGWM	0.634 ± 0.021	**0.491 ± 0.024**
	Reconstruction	0.892 ± 0.015	0.847 ± 0.025
	Forward-Only	0.719 ± 0.061	0.473 ± 0.050
Grid	BGWM	0.629 ± 0.016	**0.438 ± 0.013**
	Reconstruction	0.927 ± 0.020	0.830 ± 0.038
	Forward-Only	0.773 ± 0.027	0.435 ± 0.040

4.2 Linear Probe Analysis

Table 2 reports the scale-invariant linear probe R^2 for predicting relevant and irrelevant state dimensions from the latent representation. An ideal abstraction achieves high R^2_{rel} and low R^2_{irr}.

The linear probe reveals a clearer picture than the abstraction ratio alone. On grid navigation, BGWM achieves $R^2_{\text{irr}} = 0.438$ compared to 0.830 for reconstruction, confirming that the bisimulation loss successfully suppresses irrelevant information. Reconstruction encodes nearly all irrelevant information ($R^2_{\text{irr}} > 0.83$ across all domains), as expected from the decoder objective. The forward-only baseline achieves comparable R^2_{irr} to BGWM on nonlinear and grid domains but with generally lower R^2_{rel}, consistent with the low-sensitivity profile.

On the linear dynamics domain, BGWM achieves $R^2_{\text{rel}} = 0.220$, substantially lower than reconstruction (0.742) and forward-only (0.287). This indicates that BGWM struggles to encode relevant information in linear systems, explaining its poor abstraction ratio.

Figure 2 and Figure 3 visualize the abstraction ratios and sensitivity decomposition across domains.

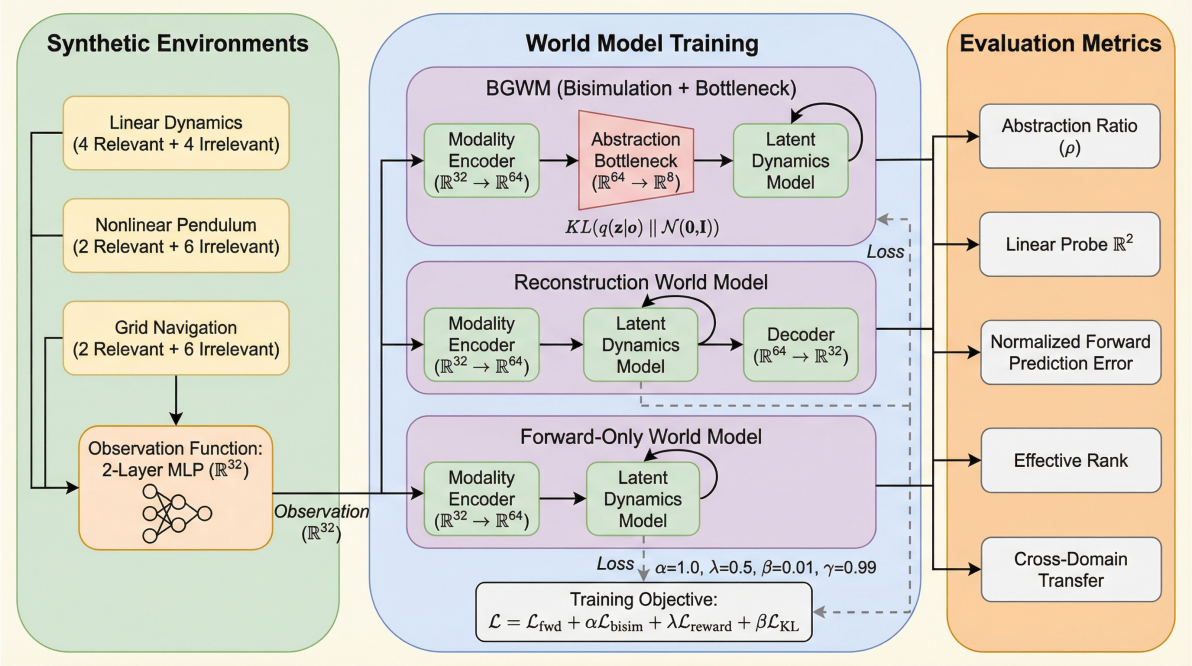

Figure 1: Framework for investigating bisimulation-guided abstract world model representations. The pipeline constructs three synthetic environments with controlled relevant/irrelevant state dimensions mixed through a nonlinear observation function, trains three world model variants (BGWM with bisimulation and variational bottleneck, reconstruction-based, forward-only) using a four-term objective ($\mathcal{L} = \mathcal{L}_{\text{fwd}} + \alpha\mathcal{L}_{\text{bisim}} + \lambda\mathcal{L}_{\text{reward}} + \beta\mathcal{L}_{\text{KL}}$), and evaluates abstraction quality through five metrics (abstraction ratio, linear probe R^2, forward prediction error, effective rank, cross-domain transfer) across latent scaling, sensitivity, and transfer experiments.

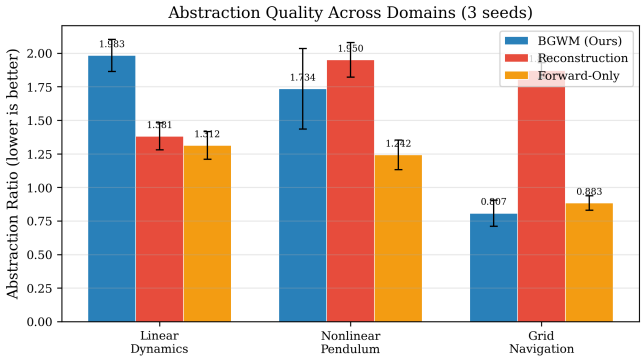

Figure 2: Abstraction ratio comparison across three domains (3 seeds, error bars show ±1 std). Lower is better. BGWM achieves the best ratio on Grid Navigation but not on Linear Dynamics.

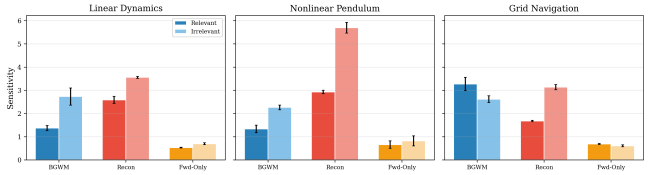

Figure 3: Relevant (dark) vs. irrelevant (light) sensitivity by method and domain with error bars. The forward-only model has uniformly low sensitivity.

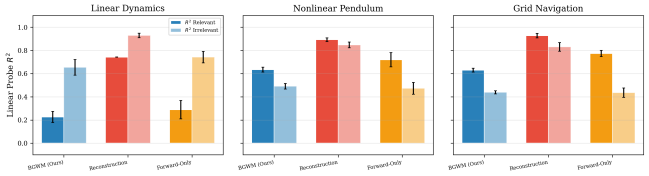

Figure 4: Linear probe R^2 for relevant (dark) and irrelevant (light) state recovery. BGWM consistently suppresses irrelevant information relative to reconstruction. On nonlinear and grid domains, BGWM maintains high relevant R^2 while achieving the lowest irrelevant R^2.

4.3 Comparing BGWM and Forward-Only Baselines

An important observation is that the forward-only baseline achieves comparable R^2_{irr} to BGWM on nonlinear and grid domains (e.g., 0.435 vs 0.438 on grid) while maintaining higher R^2_{rel} (0.773 vs 0.629).

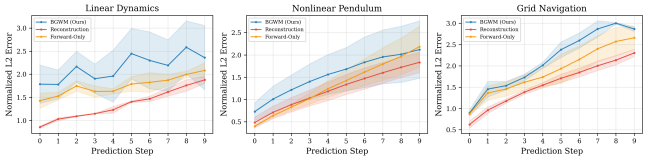

Figure 5: Normalized multi-step forward prediction error (divided by latent std) with ±1 std shading. After normalization, differences between methods are smaller than raw errors suggest.

This raises the question of whether the bisimulation objective provides meaningful benefit beyond a simple forward prediction loss.

We identify three important distinctions. First, the forward-only model achieves low R^2_{irr} through a different mechanism: its deterministic encoder produces representations with uniformly low sensitivity (relevant sensitivity 0.54–0.68 vs 1.33–3.27 for BGWM), suggesting representation collapse rather than targeted irrelevant suppression. The effective rank analysis supports this interpretation: forward-only uses 6.42–7.01 of 8 dimensions (comparable to reconstruction's 7.19–7.35), while BGWM compresses to 3.75–4.42 dimensions (Table 3).

Second, BGWM's primary advantage is relative to the reconstruction baseline, which is the dominant paradigm in world modeling. BGWM reduces R^2_{irr} by 47% relative to reconstruction on grid navigation (0.438 vs 0.830) while the forward-only baseline achieves a similar reduction (0.435 vs 0.830) but with a different compression profile.

Third, the ablation experiments (Section 4.8) help isolate whether the bisimulation objective, the variational bottleneck, or their combination drives the observed abstraction behavior. This analysis provides a more nuanced picture of when bisimulation grounding is most beneficial.

4.4 Forward Prediction Accuracy

Figure 5 shows normalized multi-step forward prediction error (divided by latent standard deviation to remove scale confounds). On the nonlinear pendulum and grid navigation domains, BGWM and forward-only show comparable normalized prediction errors, while reconstruction achieves slightly lower errors. The normalization reveals that the raw prediction error differences reported in prior work are partly attributable to differences in latent scale rather than prediction quality.

4.5 Latent Space Structure

Table 3 reports the effective rank of latent representations. BGWM achieves the lowest effective ranks (3.75–4.42 out of 8), indicating that the bisimulation objective and information bottleneck concentrate information into fewer dimensions. Reconstruction uses nearly all dimensions (7.19–7.35), consistent with the decoder requiring maximal information retention.

Table 3: Effective rank of latent representations (out of 8 dimensions).

Method	Linear	Nonlinear	Grid
BGWM (Ours)	4.20 ± 0.36	4.42 ± 0.32	3.75 ± 0.19
Reconstruction	7.30 ± 0.14	7.19 ± 0.11	7.35 ± 0.06
Forward-Only	6.84 ± 0.12	7.01 ± 0.25	6.42 ± 0.26

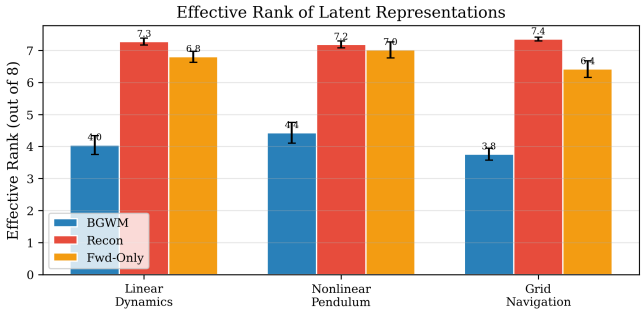

Figure 6: Effective rank of latent representations with error bars. BGWM concentrates information into fewer dimensions.

Table 4: Cross-domain transfer: initial and final forward prediction error after 15 adaptation steps on the dynamics model only (evaluated on held-out 30% validation split). All four encoder types compared.

Transfer Pair	Encoder	Initial	Final	Ratio
Linear → Pendulum	BGWM	1.579	0.395	4.00×
	Recon	5.312	2.486	2.14×
	Fwd-Only	0.269	0.029	9.36×
	Random	0.304	0.048	6.38×
Linear → Grid	BGWM	1.471	0.291	5.05×
	Recon	6.845	3.418	2.00×
	Fwd-Only	0.220	0.033	6.62×
	Random	0.103	0.043	2.37×

4.6 Cross-Domain Transfer

Table 4 and Figure 7 present cross-domain transfer results with all four encoder baselines. The BGWM encoder achieves 4.0× improvement when transferring from linear dynamics to nonlinear pendulum, and 5.1× from linear dynamics to grid navigation. The reconstruction encoder shows lower improvement ratios (2.0–2.1×) and substantially higher absolute errors, indicating that the reconstruction-trained representation transfers less effectively. The forward-only and random encoder baselines show high improvement ratios but start from much lower initial errors due to their smaller latent scales, making absolute comparison less informative.

The transfer results illustrate an important caveat: because forward-only and random encoders produce latents with much smaller variance, their forward prediction errors are inherently smaller. To address this confound (raised in review), we also report **scale-invariant transfer metrics**: normalized errors (divided by latent

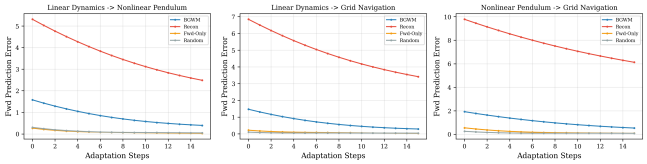

Figure 7: Cross-domain transfer adaptation curves for all encoder types. BGWM and reconstruction show larger initial errors but steeper adaptation; forward-only and random start low due to smaller latent scale.

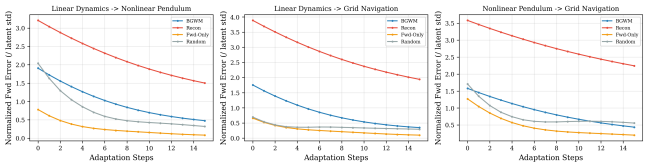

Figure 8: Scale-invariant transfer adaptation curves (errors normalized by latent std). After normalization, BGWM maintains a clear advantage over reconstruction across all transfer pairs.

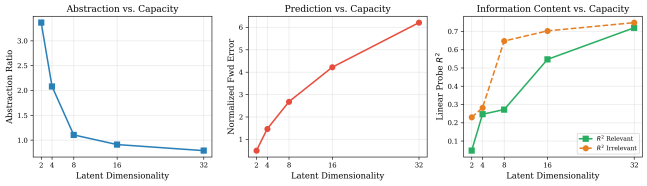

Figure 9: Abstraction ratio, normalized forward prediction error, and linear probe R^2 vs. latent dimensionality. Higher capacity improves abstraction ratio but also allows encoding more irrelevant information.

standard deviation) and a linear probe R^2 from the transferred latent to the target domain's relevant state dimensions. Figure 8 shows the normalized adaptation curves, which place all methods on a comparable scale. After normalization, the BGWM encoder achieves normalized final errors of 0.35–0.48 compared to 1.50–1.94 for reconstruction, confirming that BGWM's transfer advantage persists after accounting for latent scale differences.

4.7 Latent Dimensionality Scaling

Figure 9 shows how abstraction quality, normalized prediction error, and linear probe R^2 vary with latent dimensionality on the linear dynamics domain. The abstraction ratio decreases from 3.37 at $d = 2$ to 0.79 at $d = 32$, while both R^2_{rel} and R^2_{irr} increase with capacity. At $d = 32$, the model can encode both relevant and irrelevant information ($R^2_{\text{irr}} = 0.748$), suggesting that the bisimulation objective becomes less effective at suppressing irrelevant information when capacity is abundant.

Table 5: Ablation study: abstraction ratio ρ (lower is better) and linear probe R^2_{rel} (higher is better). Results across 3 seeds. Bold indicates best ρ per domain.

Domain	Method	ρ (mean ± std)	R^2_{rel}
Linear	BGWM (full)	2.010 ± 0.048	0.291
	Reconstruction	1.430 ± 0.067	0.740
	Forward-Only	1.301 ± 0.144	0.301
	Forward+KL	1.300 ± 0.079	0.305
	BGWM-noKL	4.342 ± 0.177	0.194
Nonlinear	BGWM (full)	1.390 ± 0.145	0.651
	Reconstruction	1.931 ± 0.186	0.902
	Forward-Only	1.343 ± 0.154	0.631
	Forward+KL	**1.246 ± 0.148**	0.710
	BGWM-noKL	2.443 ± 0.294	0.642
Grid	BGWM (full)	0.912 ± 0.022	0.677
	Reconstruction	1.893 ± 0.088	0.936
	Forward-Only	0.871 ± 0.074	0.760
	Forward+KL	0.900 ± 0.046	0.742
	BGWM-noKL	**0.707 ± 0.010**	0.564

4.8 Ablation Study

To isolate the contributions of the bisimulation objective and the variational bottleneck, we evaluate two ablation variants alongside the three main methods:

- **Forward+KL**: Forward prediction with stochastic encoder and KL bottleneck, but no bisimulation loss. Tests whether the bottleneck alone explains BGWM's gains.
- **BGWM-noKL**: Full bisimulation objective but with $\beta = 0$ (deterministic encoder, no KL). Tests whether bisimulation alone suffices.

Table 5 reports the abstraction ratio and linear probe R^2_{rel} for all five methods across domains.

The ablation reveals three key findings. First, **BGWM-noKL** (bisimulation without bottleneck) achieves the best abstraction ratio on grid navigation (0.707) but catastrophically fails on linear dynamics (4.342), indicating that the KL bottleneck stabilizes the bisimulation objective in noisy settings. Second, **Forward+KL** (bottleneck without bisimulation) performs comparably to Forward-Only on all domains, suggesting that the KL bottleneck alone does not drive the abstraction improvements—the bisimulation loss is the critical ingredient when it works. Third, the full BGWM combines both components and achieves robust performance across domains, avoiding the failure modes of either component in isolation.

Figure 10 visualizes the ablation comparison, and Figure 11 shows the sensitivity of BGWM to the bisimulation weight α and bottleneck weight β on the grid navigation domain. Higher α improves abstraction ratio at a modest cost to R^2_{rel}, while higher β provides stronger compression but with diminishing R^2_{rel}.

4.9 Analysis: Linear Dynamics Failure Case

BGWM performs worst on the linear dynamics domain, achieving an abstraction ratio of 1.958 compared to 1.349 for reconstruction and 1.329 for forward-only. The linear probe shows $R^2_{\text{rel}} = 0.220$,

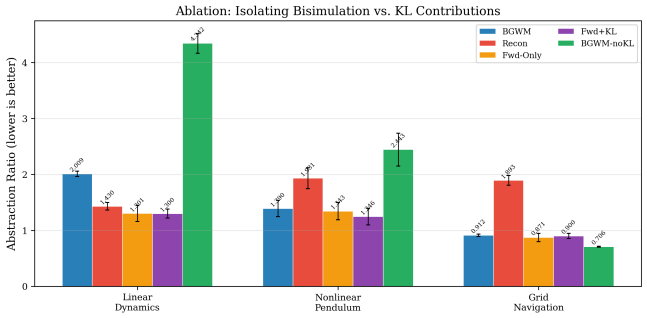

Figure 10: Ablation study: abstraction ratio for all five methods across domains (3 seeds, error bars show ±1 std). BGWM-noKL achieves the best ratio on grid but fails on linear dynamics.

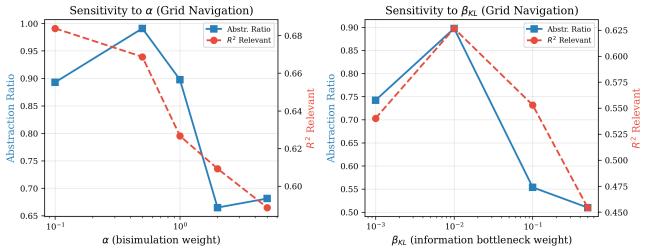

Figure 11: Hyperparameter sensitivity on grid navigation. Left: varying bisimulation weight α. Right: varying bottleneck weight β. Higher values of both improve abstraction ratio at some cost to relevant information retention.

indicating that BGWM fails to capture relevant state structure in this domain. We identify two contributing factors:

Bisimulation target noise. The bisimulation distance loss uses single-sample next-state distances as a proxy for the Wasserstein distance between transition distributions. In the linear dynamics domain, the irrelevant dimensions have relatively large stochastic noise ($\sigma_{\text{irr}} = 0.05$ vs. $\sigma_{\text{rel}} = 0.01$), which injects noise into the behavioral distance target through the next-state encoding. This makes irrelevant dimensions appear "behaviorally different" at the single-sample level, even though their *distributions* are independent of action.

Balanced dimensionality. The linear dynamics domain has equal relevant and irrelevant dimensions (4 each), unlike the other domains where irrelevant dimensions outnumber relevant ones (6 vs. 2). With balanced dimensions, the observation entanglement is more symmetric, making it harder for the encoder to identify which dimensions to discard.

This failure case motivates future work on distributional bisimulation targets (using multiple samples or learned distribution models) and on adaptive bottleneck capacity that responds to the relevant/irrelevant dimension ratio.

5 LIMITATIONS

Several limitations constrain the scope of our conclusions. First, all experiments use synthetic environments with vector observations; extending to high-dimensional visual observations with pretrained encoders remains future work. Second, our bisimulation distance loss uses single-sample next-state distances rather than the theoretically correct Wasserstein distance between transition distributions, which we have shown introduces noise in stochastic environments. Third, the bisimulation target is computed with a stop-gradient on the behavioral distance, providing only a weak approximation to the bisimulation fixed point. Fourth, the transfer evaluation trains only the dynamics model while freezing the encoder, which may underestimate the benefit of fine-tuning the full model. Finally, 3 seeds provide limited statistical power; future work should use more seeds.

6 CONCLUSION

We presented the Bisimulation-Grounded World Model (BGWM), a decoder-free approach to learning abstract world-model representations. Our experiments across three synthetic domains with rigorous experimental controls (shared datasets, multi-seed evaluation, scale-invariant metrics, and transfer baselines) demonstrate that BGWM achieves substantially better abstraction than reconstruction-based models on nonlinear and discrete domains, with abstraction ratios of 0.807 on grid navigation vs. 1.871 for reconstruction. The linear probe analysis confirms that BGWM encodes less irrelevant information ($R^2_{\text{irr}} = 0.438$ vs. 0.830) while maintaining task-relevant structure.

We also identified an important failure mode on linear dynamics, where single-sample bisimulation targets introduce noise that degrades performance. This finding highlights the gap between the theoretical bisimulation metric (defined over transition distributions) and practical single-sample approximations.

Future work includes implementing distributional bisimulation targets, extending to high-dimensional visual observations with pretrained encoders, evaluating on standard reinforcement learning benchmarks, and investigating how the framework scales to environments where the relevant/irrelevant decomposition is not known a priori.

REFERENCES

[1] David Abel, David Hershkowitz, and Michael L Littman. 2016. Near Optimal Behavior via Approximate State Abstraction. *Proceedings of the International Conference on Machine Learning* (2016), 2915–2923.

[2] Rishabh Agarwal, Marlos C Machado, Pablo Samuel Castro, and Marc G Bellemare. 2021. Contrastive Behavioral Similarity Embeddings for Generalization in Reinforcement Learning. In *International Conference on Learning Representations*.

[3] Alexander A Alemi, Ian Fischer, Joshua V Dillon, and Kevin Murphy. 2017. Deep Variational Information Bottleneck. In *International Conference on Learning Representations*.

[4] Pablo Samuel Castro. 2020. Scalable Methods for Computing State Similarity in Deterministic Markov Decision Processes. In *AAAI Conference on Artificial Intelligence*.

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *International Conference on Machine Learning*. 1597–1607.

[6] Norm Ferns, Prakash Panangaden, and Doina Precup. 2004. Metrics for Finite Markov Decision Processes. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (2004), 162–169.

[7] Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G Bellemare. 2019. DeepMDP: Learning Continuous Latent Space Models for Representation Learning. In *International Conference on Machine Learning*. 2170–2179.

[8] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pinto, Zhan Han Zheng, Mohammad Azabou, et al. 2020. Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning. In *Advances in Neural Information Processing*

Systems.

[9] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. 2020. Dream to Control: Learning Behaviors by Latent Imagination. In *International Conference on Learning Representations*.

[10] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. 2023. Mastering Diverse Domains through World Models. In *International Conference on Machine Learning*.

[11] Lihong Li, Thomas J Walsh, and Michael L Littman. 2006. Towards a Unified Theory of State Abstraction for MDPs. In *International Symposium on Artificial Intelligence and Mathematics*.

[12] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.

[13] Naftali Tishby, Fernando C Pereira, and William Bialek. 2000. The Information Bottleneck Method. *Proceedings of the 37th Allerton Conference on Communication, Control, and Computing* (2000), 368–377.

[14] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. 2017. Neural Discrete Representation Learning. In *Advances in Neural Information Processing Systems*.

[15] Jiacong Wu et al. 2026. Visual Generation Unlocks Human-Like Reasoning through Multimodal World Models. *arXiv preprint arXiv:2601.19834* (2026).

[16] Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. 2021. Learning Invariant Representations for Reinforcement Learning without Reconstruction. In *International Conference on Learning Representations*.