# Lumping or Splitting: How Context Shapes Motor Sequence Representations

**Mohan W. Gupta (mg9965@princeton.edu)[1] & Jordan A. Taylor (jordanat@princeton.edu)[1,2]**
**Department of Psychology[1], Princeton Neuroscience Institute[2], Princeton University, Princeton, NJ 08544**

## Abstract

Motor skills are often assumed to rely on a shared representation that generalizes across superficial changes in framing. Yet identical actions are practiced under distinct contexts that may gate what is retrieved and expressed. We tested whether performance–inconsequential context pushes learning toward generalization ("lumping") or context–specific codes ("splitting"). Simple recurrent networks (SRNs) learned next–step prediction for a fixed eight–item sequence in no–context, single–context, or alternating–context conditions, with weight scale manipulated to bias the learned solution while holding the SRN architecture fixed. Humans trained on the same sequence under matched contexts. All groups improved with practice, but alternating contexts produced a sustained acquisition cost and a marked transition cost when switching between contexts. Under identical training, Low–Scale SRNs showed cross–context generalization, whereas High–Scale SRNs reproduced the context–dependent costs observed in human behavior. White–box analyses revealed separable context–specific representations and greater weighting of context inputs in High–Scale SRNs, consistent with representational splitting.

**Keywords:** motor sequence learning; context–dependent memory; similarity; neural networks

## Introduction

Did you know that *Twinkle, Twinkle Little Star* and the *Alphabet Song* have the same melody? If not, you are in good company. Prior work has found that the representation of songs (melody and lyrics) is bound together (Serafine et al., 1984, 1986; Crowder et al., 1990). In this example, the melody is shared between the two songs, but the lyrics provide enough context that practice in one does not transfer to the other. An open question is whether this contextual binding extends beyond melody and lyrics to include motor memories for singing or playing the song. Would a budding pianist, who has already mastered *Twinkle, Twinkle Little Star*, generalize their learning when attempting to learn the *Alphabet Song* or would they have to start *de novo*?

Motor sequence learning is often framed as the acquisition of a relatively unitary, task–specific skill representation that becomes more fluent with practice and is not strongly tied to incidental contextual cues, implying substantial transfer across many surface–level changes in how the sequence is cued or framed (Keele et al., 1995; Hikosaka et al., 2002; Ungerleider, 2002; Willingham et al., 2000; Doyon et al., 2009; Dayan & Cohen, 2011; Abrahamse et al., 2013). Yet real–world skills often challenge this intuition: the same underlying action structure is frequently learned under distinct contextual frames that can shape what is retrieved and expressed and, in some cases, constrain transfer (Česonis & Franklin, 2022; Howard et al., 2013; Ruitenberg et al., 2012; Schmidt et al., 2012; Wright & Shea, 1991). This raises a basic question for motor learning: how are motor sequence representations shaped by context?

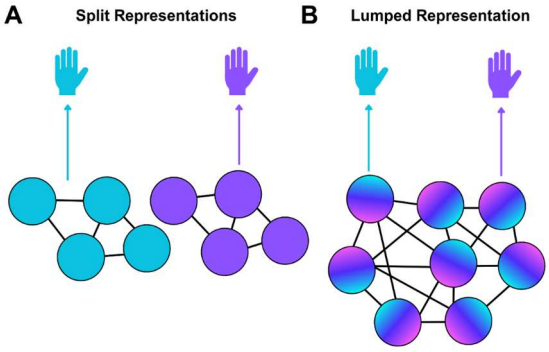


Figure 1. (A) Split representations: the network forms split representations for each context (teal vs. purple), such that the same motor mapping is supported by context–specific latent states. (B) Lumped representation: a shared, overlapping latent code (mixed colors) supports performance across contexts, promoting generalization.

Despite decades of work using motor sequence learning as a model paradigm, evidence for context–dependent sequence memory and transfer remains mixed and appears to depend on boundary conditions such as whether contextual features are incidental vs task–relevant and whether performance is stimulus–driven or memory–based (Wright & Shea, 1991; Bapi et al., 2000; Abrahamse & Verwey, 2008; Ruitenberg et al., 2012; Sobierajewicz et al., 2017). Recent advances in neural network modeling, in other domains of memory, have revealed that subtle factors such as task similarity and prior initialization can have profound effects on whether singular or multiple representations are formed. The advantage afforded by these networks is that the internal representations are both learned and directly observable. Indeed, sequence or statistical structure learning was one of the first applications of simple recurrent networks (SRNs; Elman networks; Servan–Schreiber et al., 1988; Elman, 1990; Cleeremans & McClelland, 1991).

Building on the recent advances in this tradition (Chizat et al., 2019; Woodworth et al., 2020; Dominé et al., 2025; Holton et al., 2026), we begin by adopting a minimal computational approach, leveraging SRNs to predict the next element of a sequence, like the next note in a melody. Crucially, the same task objective can be solved through multiple internal solutions by the network (Chizat et al., 2019; Woodworth et al., 2020; Dominé et al., 2025; Holton et al., 2026). One class of solutions "lumps" experience across contexts by constructing a shared representation that is employed throughout learning (Figure 1B). Another class "splits" hidden state representations into independent spaces (Figure 1A). The lumping solution affords generalization across similar contexts and/or task demands, while the



In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* 

splitting solution keeps the representations context–bound at the cost of generalization.

Returning to our budding pianist example, a consequence of splitting is that the same sequence of piano moves learned across two different song contexts effectively must be learned twice. This is true even though the efficient solution for this task is to build a single context–agnostic representation — only the motor sequence matters. Whether our pianist lumps or splits their learned representation is not a difference in the kinematic demands of the task but a difference in where training lands within the network solution space.

While the degree of context (or psychological) similarity is known to play a key role in representational generalization (Shepard, 1987; Tenenbaum & Griffiths, 2001; Marjieh et al., 2024), such generalization can be understood, mechanistically, as the emergence of a lumped representation in which multiple contexts share an overlapping latent code. In our framework, this overlap implies that training updates induced in one context transfer directly to the other, whereas separation implies context–specific learning. Another way to push neural networks to either lump or split is by manipulating the weight scale (i.e., the variance of initial weights), which changes the starting distribution of connection strengths while leaving architecture and task objective fixed (Chizat et al., 2019; Woodworth et al., 2020; Dominé et al., 2025; Holton et al., 2026). Specifically, weight scale can bias how strongly training reshapes internal representations. Under lower weight scales (less variance among initial weights), learning tends to remain closer to an initial feature basis, often yielding a more context–invariant, lumped representation. Higher scales increase the network's effective degrees of freedom for feature learning, making it easier for context cues to carve separable trajectories in hidden state—opening the possibility of context–indexed (split) representations (Morcos et al., 2018; Kornblith et al., 2019).

This framing motivated us to train SRNs on a motor sequence learning task to track how lumped or split representations emerge as a function of network initialization. Here, we simulated and probed the behavior of SRNs in a motor sequence–learning paradigm with three conditions: learning with no context (a context–absent baseline), learning under a single context, and learning while switching between contexts. At each time step, the context was given to the network as a one–hot vector. These paradigms enable us to ask how both context and weight initialization shape representations. We connect the SRN behavior with white–box analyses on the internal representations and network dynamics: (i) whether contexts occupy overlapping versus separable subspaces (e.g., similarity between the representations), and (ii) how the network uses context signals.

We tested whether these SRN predictions bore out by running the same task in humans under matched experimental conditions. Of specific interest was whether radically different training contexts could induce a split representation when learning the same motor sequence in humans. By aligning behavioral contrasts in humans with mechanistic diagnostics in SRNs, we provide a bridge between computational solution spaces and how context organizes motor sequence learning.

# Methods

## SRNs

SRNs were trained to predict the next sequence item from the current item (i.e., next–step prediction). The model is a one–layer tanh RNN with a 6–unit hidden state and a linear output layer over a four–item vocabulary. Sequence inputs were encoded as Gaussian–smoothed token vectors (width σ), imposing an inductive bias that adjacent keys have partially overlapping codes. Context was instantiated in the networks through a two–dimensional one–hot vector that was concatenated to the input at every time step (Yang et al., 2017; Low et al., 2023). Training minimized cross–entropy loss using the Adam optimizer (Kingma & Ba, 2017). To test how weight scale shaped the learned solution, we set the Low–Scale networks to 0.02 and the High–Scale networks to 0.5.

## Participants

One hundred and eleven (age = 27.41, F = 46.3%) participants were recruited from the online Prolific participant pool in exchange for payment. Recruitment was limited to individuals who were right–handed. Three were excluded because their data did not meet reasonable standards for quality (see below). Informed consent was given via button press. All procedures were approved by the relevant institutional review board.

## Human Task

Participants learned to repeatedly type an eight–item sequence using the 'A', 'S', 'D', and 'F' keys on their computer keyboard with their non–dominant left hand. The eight–item sequence was selected by exhaustive generate–and–filter to balance key frequency and transition statistics while excluding obvious regularities. Each block consisted of 10 sequences followed by a 10–second break. The sequence was explicitly displayed on the screen throughout the trial. Correct responses were indicated via visual feedback (a sprite/marker adjacent to the current item and/or a color change of the letter). Participants completed a short familiarization phase (two consecutive sequences) and were told to complete the sequences as quickly and accurately as possible. They proceeded to the main task only after achieving ≥ 90% accuracy and completing two correct sequences in under 4 s. The main experiment consisted of 10 blocks where each block had 10 sequences. Participants were randomly assigned to one of the four groups. For the no–context group, participants were tasked with simply pressing the computer key of the presented letter (Figure 2A). For the single–context groups, participants were assigned to either a "maze" context or a "melody" context. In the maze context, the sequence unfolds as a 2D side–scrolling platformer game where each key press allows an avatar to successfully

navigate obstacles (Maze group; Figure 2B). In contrast, in the melody context, the sequence appears as staff–like notation that scrolls across the screen and each key press is accompanied by an auditory note (Melody group; Figure 2C). Finally, for the Maze–Melody group, the participants alternated between the maze and melody contexts. Critically, all groups performed the same sequence each block with the only difference being the context in which it was executed.

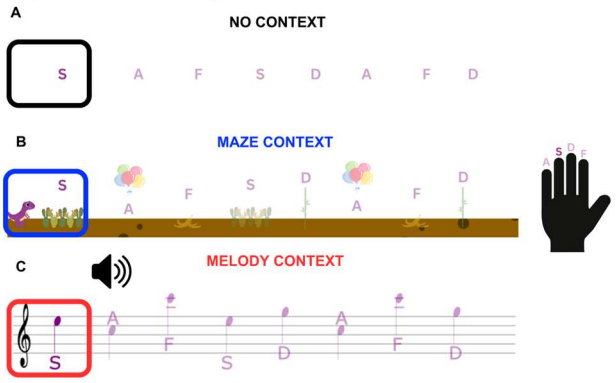


Figure 2. Participants produced a fixed eight–item sequence using four response keys (A, S, D, F) mapped to the four fingers of the left hand. (A) No–context condition: sequence cues were presented without any additional contextual scene. (B) Maze context: the same sequence was embedded in a gamified maze–like display, with each correct keypress advancing progress through the environment. (C) Melody context: the same keypress sequence was presented as musical notation and accompanied by an auditory note.

## Transition Cost

To quantify context–transition cost in the Maze–Melody condition, we first reduced trial–level data to block–level summaries for each SRN and participant/run. For each participant or SRN, we computed block–level mean performance (sequence RT for humans; loss for SRNs). Transition cost was defined as the absolute difference between consecutive blocks (|block n+1 − block n|) at context switches. To place humans (RT) and models (loss) on a common scale, the transition cost was z–scored within the SRN or participant.

## SRN White–Box Analyses

We used PCA to obtain a low–dimensional summary of each trajectory. Hidden states were mean–centered across time and projected onto the first two principal components (PC1–PC2), providing a 2D visualization of how the network state evolves as the sequence unfolds. Overlapping trajectories across contexts indicate a shared, context–invariant representation (lumping), whereas systematically separated trajectories indicate context–dependent representations (splitting).

To quantify context–dependent splitting, we computed the centroid of each trajectory in PC space (mean PC1/PC2 across steps) and measured context separation as the Euclidean distance between the Maze and Melody centroids (after Procrustes alignment). Specifically, we applied an orthogonal Procrustes transform (Goodall, 1991) to align one context's PC scores to a reference embedding, ensuring that differences reflect representational structure rather than arbitrary PCA axis orientation.

To quantify whether SRNs functionally relied on the context input, we measured how much of the learned input–to–hidden mapping was allocated to the context dimensions. Concretely, because the SRN input at each time step is a concatenation of the (Gaussian–smoothed) tokens and the 2–d one–hot context vector, the input→hidden weight matrix can be partitioned into token–related columns versus context–related columns. We then computed a context–weight ratio as the norm of the context–related columns relative to the total input weight norm.

## Human Data Analysis

Participant–level exclusions were applied to remove low–quality data, which is common practice for studies conducted using online platforms (Crump et al., 2013; Masís Obando et al., 2025). First, we computed error rates and excluded participants whose error rate exceeded the group mean + 2.5 SD. Second, we excluded unusually slow responders using a two–pass criterion on participant mean RTs computed from correct, non–warmup trials: participants above mean + 2.5 SD were removed, the distribution was recomputed, and the same cutoff was applied again to define the final inclusion set. Sequence RT was defined as the sum of keypress RTs within each correctly completed sequence. Incomplete sequences due to an error keypress were removed.

# Results

## Context Switching Worsens Performance in Humans

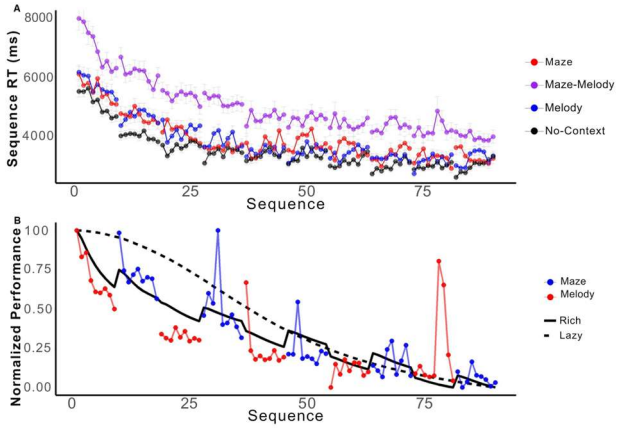


Figure 3. (A) Human motor sequence learning curves showing mean sequence RT across practice for single–context training (Maze, Melody), a switching condition (Maze–Melody), and a no–context control. RT decreases over sequences in all groups, with elevated RT in the Maze–Melody group relative to single–context and no–context groups. (B) SRN performance (normalized loss; lower is better) plotted as a function of cumulative sequence experience during training in the Maze–Melody group.

Across acquisition, participants' sequence execution times decreased with practice in all conditions, consistent with online skill improvement. However, a mixed–effects model predicting log RT from practice, group, and their interaction confirmed that the Maze–Melody group was slower overall than the pooled no–context and single–context groups, $F(1, 108) = 10.02$, $p = .002$, producing a sustained acquisition cost: performance was slower when contexts alternated, even though the motor sequence was identical (Figure 3A).

Notably, the Maze–only and Melody–only learning curves appear to be closely matched in both initial performance and learning rate, suggesting that the two contexts were well balanced in baseline difficulty. Again, it is worth noting that these contexts were inconsequential to successful execution of the motor sequence. With this in mind, the selective slowing in the Maze–Melody group is therefore consistent with a lack of generalization between these two task–inconsequential contexts, rather than intrinsic differences between Maze and Melody stimuli.

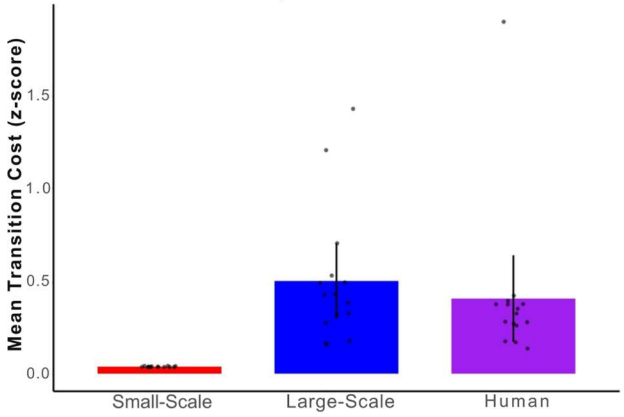


Figure 4. Bars show the mean transition cost (z–scored), defined as the absolute difference immediately following a context change relative to the preceding trials in the Maze–Melody group. Low–Scale SRNs exhibit near–zero transition costs, whereas High–Scale SRNs and humans show similar levels of transition costs.

When we trained SRNs on the same sequential structure with context inputs, we observed two qualitatively distinct behavioral signatures. Both High– and Low–Scale models learned the sequence structure (improving performance over practice), but they differed in whether they treated Maze and Melody as distinct "tasks." The High–Scale solution, which more strongly picked up contextual features, captured a context–dependent acquisition profile more consistent with the behavioral data showing a separation between contexts (Figure 3B). For ease of comparison to the SRN simulations, we visualize and analyze the subset of human participants in the Maze–Melody condition who experienced the same context order (i.e., a single counterbalance); results were qualitatively similar when collapsing across counterbalance order. In contrast, the Low–Scale model solution class exhibited substantially more cross–context sharing (i.e., more uniform performance across contexts; Figure 3B). This dissociation suggests that identical training can yield either a "context–gated" representation that partitions learning across contexts (splitting) or "shared" representation that supports generalization across contexts (lumping). This dissociation was revealed by examining the transition cost both in the human behavioral data and the SRN simulations when context changed, even though the motor sequence was unchanged.

Indeed, a one–way ANOVA on the mean transition cost revealed a main effect, $F(2, 42) = 8.61$, $p < .001$, $\eta^2 = 0.29$. Tukey–corrected post hoc tests indicated that both High–Scale SRNs and humans exhibited larger transition costs than Low–Scale SRNs (High–Low–Scale = 0.461, 95% CI [0.176, 0.746], $p < .0001$; Human–Low = 0.366, 95% CI [0.081, 0.651], $p < .0001$). The difference between Human and High–Scale SRNs did not approach significance (Human–High = −0.095, 95% CI [−0.380, 0.190], $p = 0.701$). Overall, these behavioral signatures suggest that context alternation selectively disrupts performance by limiting cross–context generalization, even when the motor mapping is unchanged.

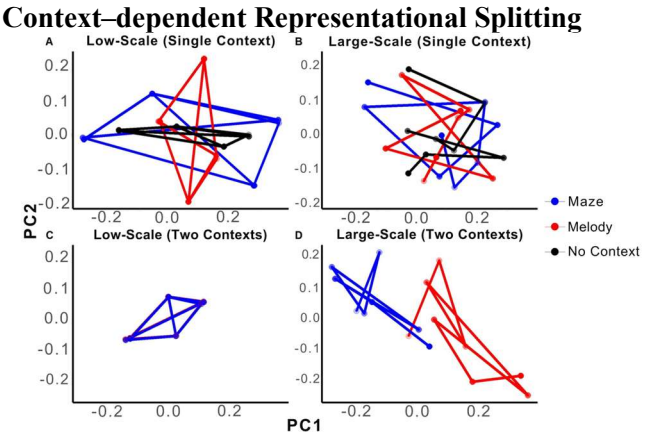


Figure 5. Hidden–state activity patterns were projected into the first two principal components (PC1–PC2) and plotted separately for Low– and High–Scale networks trained under single–context or no–context (A–B) or two–context (C–D) regimes. Colors denote the context condition used to generate the activity patterns (Maze = blue, Melody = red, No Context = black). Low–Scale networks show largely overlapping structure across contexts, whereas High–Scale networks—particularly in the two–context regime—exhibit clear context–dependent separation consistent with representational splitting.

Our modeling approach, leveraging SRNs and white–box analyses, affords us the opportunity to peek underneath the hood of learning. If this task–inconsequential context truly splits the learned representations, then it should be apparent by examining the hidden–state geometry of the SRNs during sequence execution. We first examined single–context networks (Figure 5A–B) to establish a baseline: when trained in only one context, both Low– and High–Scale SRNs expressed a single trajectory for that trained context, and differences between Maze and Melody reflect context–specific input statistics rather than learned partitioning across

contexts. The critical test is therefore the Maze–Melody condition, where the same network must accommodate both contexts. Here, High–Scale networks showed a clear separation between Maze and Melody trajectories in the first two principal components (Figure 5D), indicating that the internal state evolved in context–specific regions of state space (splitting). In contrast, Low–Scale networks showed strong cross–context overlap (Figure 5C), consistent with a shared trajectory that supports generalization (lumping).

To quantify differences in representational geometry across training regimes, we computed for each network and context the centroid of the PCA trajectory (mean PC1/PC2 across steps) and summarized its magnitude as Euclidean distance from the origin and submitted this to a 2 (Network Type) × 4 (Context) ANOVA on centroid magnitude. Because the PCA basis can differ across datasets when fit separately, we first aligned each context–specific space to a common reference space (the Maze–Melody group) using a Procrustes transform fit (Goodall, 1991) to the mean trajectory across steps. High–Scale networks differed from Low–Scale networks in overall centroid magnitude, $F(1, 170) = 9.11$, $p < .005$, partial $\eta^2 = 0.05$, and centroid magnitude varied across contexts, $F(3, 170) = 14.11$, $p < .0001$, partial $\eta^2 = 0.20$. Despite the apparent differences in representational geometry between the High– and Low–Scale networks according to context, the Model Type × Context interaction only trended in the expected direction, $F(3, 170) = 2.20$, $p = .09$, partial $\eta^2 = 0.04$. Follow–up contrasts suggested that context effects were more pronounced in High–Scale networks.

**Networks Use Context Signals Differently**

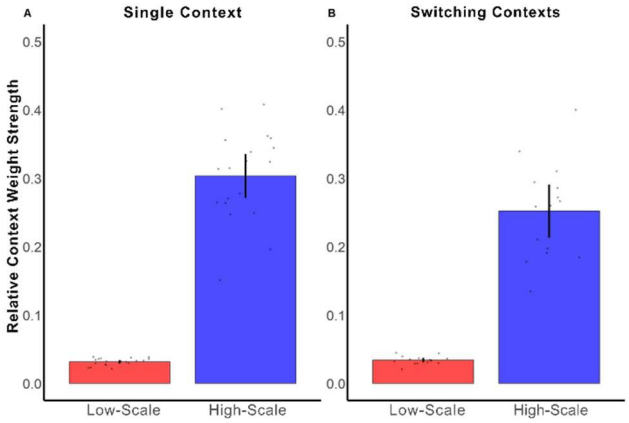


Figure 6. Bars show the mean context weight strength for Low– and High–Scale SRNs. (A) In the single–context condition, High–Scale networks exhibit substantially higher context–weight strength than Low–Scale networks, indicating greater use of the context input during sequence processing. (B) In the Maze–Melody group, High–Scale networks still show stronger context weights than Low–Scale networks.

Taken together, the PCA results distinguish lumped versus split representational geometries, but they do not identify whether context cues are driving those differences. If splitting reflects context–gated computation, High–Scale networks should allocate more input–to–hidden capacity to context features than Low–Scale networks. We tested this prediction by computing a context–to–input weight ratio from the input weight matrix (Figure 6). A 2 (Network Scale: Low vs. High) × 3 (Context: Maze, Melody, Maze–Melody) ANOVA revealed a dominant main effect of network scale: High–Scale networks showed substantially higher context weight ratios than Low–Scale networks, $F(1, 100) = 725.19$, $p < .0001$, partial $\eta^2 = 0.88$. In addition, there was a significant main effect of Context, $F(2, 100) = 4.07$, $p < .05$, partial $\eta^2 = 0.08$, qualified by a significant scale × context interaction, $F(2, 100) = 4.50$, $p < .05$, partial $\eta^2 = 0.08$, suggesting that the impact of context structure differed across network types.

## Discussion

The present work asked how context shapes motor sequence representations when the underlying action sequence is held constant. In humans, alternating between two performance–inconsequential contexts produced a sustained acquisition cost relative to single–context and no–context baselines, along with a reliable transition cost when switching contexts. In parallel, SRNs trained on the same sequential structure exhibited two qualitatively distinct representational solutions under identical performance objectives: some networks learned a shared, context–invariant representation that supported cross–context generalization (lumping), whereas others learned two context–linked representations that treated the same motor mapping as effectively context–specific (splitting). Crucially, these behavioral regimes were mirrored by mechanistic signatures in the SRNs: High–Scale networks showed separable context–specific representations and greater weighting of context inputs, whereas Low–Scale networks showed overlapping representations and minimal context reliance. Together, these results suggest that even when context is irrelevant to successful execution, context can steer learning toward splitting or lumping.

A central implication is that psychological distance, and therefore generalization, is not determined solely by kinematic similarity in motor sequence learning. Shepard's law of generalization formalizes the relationship between similarity and transfer, emphasizing that generalization falls off as experiences become more psychologically distant (Shepard, 1987; Marjieh et al., 2024). Our results highlight that contextual framing can itself create psychological distance: although the motor sequence was identical, alternating between Maze and Melody contexts impaired performance, consistent with reduced cross–context generalization. This perspective aligns with classic principles of context–dependent memory. Encoding specificity emphasizes that contextual features present during encoding can become bound to a trace and later guide retrieval (Tulving & Thomson, 1973; Reder et al., 1974; Godden & Baddeley, 1975). Formal memory models captured this influence by treating context as a slowly varying state that links temporally adjacent experiences (Estes, 1955) and later

as a high–dimensional evolving representation that organizes retrieval and similarity structure (Howard & Kahana, 2002; Polyn et al., 2009). Extending these ideas to motor sequence learning, our findings suggest that contextual cues, even when irrelevant to performance, shape underlying representations and expressed behavior.

Our results also connect naturally to motor–learning accounts in which context shapes memory formation, updating, and expression. A growing literature argues that contextual cue structure can determine whether learning updates an existing representation, protects it from interference, or creates a new representation altogether (Howard et al., 2013; Heald et al., 2018; Avraham et al., 2022; Heald et al., 2021). From this perspective, alternation between contexts can encourage the learner to treat each context as requiring a distinct latent state, producing context–dependent expression and reduced transfer. The split–lump framing provides a representational instantiation of this idea: splitting corresponds to learning context–linked representations that behave like separate internal models, whereas lumping corresponds to a shared representation that generalizes across contexts. This mapping aligns with modular accounts of motor learning in which contextual cues can act as a gating signal that selects among multiple internal models (Haruno et al., 2001; Wolpert & Kawato, 1998) and extends the logic to motor sequence learning by showing that gating can emerge even when the motor mapping itself is identical.

Although our manipulation is not a classic interleaved task design, the Maze–Melody learning curve mirrors switch and mixing costs, where task changes incur a transient penalty and maintaining readiness to switch can produce sustained slowing (Rogers & Monsell, 1995; Monsell, 2003). These patterns also echo the broader contextual interference tradition, which emphasizes that interleaving conditions can alter how representations are formed and expressed (Magill & Hall, 1990; Shea & Morgan, 1979). Importantly, in our paradigm, the transition cost cannot be attributed to differences in the motor demands of the sequence itself; instead, it reflects a cost introduced purely by contextual framing. This suggests that contextual alternation can recruit control– or gating–like processes that shape performance, even when such processes are not strictly necessary for accurate execution.

The SRN results sharpen this interpretation by showing that identical training can yield distinct internal solutions with different implications for generalization. Building on prior work, we used weight scale as a minimal way to bias learning dynamics while holding architecture, training data, and objective fixed (Chizat et al., 2019; Woodworth et al., 2020; Dominé et al., 2025; Holton et al., 2026). In our task, Low–Scale SRNs formed shared, cross–context trajectories that supported generalization, whereas High–Scale SRNs formed context–separated trajectories and weighted context inputs more strongly. Thus, weight scale shifted the network between lumped and split solutions. Although both regimes learned the sequence behaviorally, white–box analyses revealed a clear representational dissociation: High–Scale networks showed context–separated subspaces (splitting) and substantially greater reliance on context input weights, while Low–Scale networks showed overlapping subspaces (lumping) and minimal context reliance. Multiple possible solutions may help explain why context effects can appear mixed across paradigms—context dependence may reflect task structure, inductive biases, and learning dynamics that determine whether a representation is lumped or split. In our data, the split solution most closely matched the Maze–Melody group.

Several limitations point to clear extensions. First, splitting versus lumping is inferred from a set of behavioral and mechanistic diagnostics, but it may exist on a continuum. A natural next step is to formalize a continuous "degree of splitting" metric such as centroid distance between subspaces and relate this to the transition cost across weight scales. Second, our contexts were intentionally salient and conceptually distinct; future work can parametrize contextual similarity more finely (e.g., graded cue similarity or shared task goals) to test whether the probability of splitting varies smoothly with psychological distance, as suggested by Shepard's Law (Shepard, 1987; Marjieh et al., 2024). Third, the human analogue of weight scale remains underspecified. We use scale as a lever, not as a literal biological parameter to show that the same task can be solved by shared or context–specific representations. In humans, similar biases may arise from prior experience, attentional weighting of contextual cues, inferred task boundaries, or context–gating policies that determine whether learning is lumped or split. Relatedly, other factors, such as hidden–unit number or learning rate, may also push networks toward splitting or lumping. Finally, human and SRN performance were indexed with different measures: accurate–trial RTs for humans and cross–entropy loss for SRNs. These measures were chosen to capture fluent sequence execution and continuous prediction performance, respectively, but they are not formally equivalent.

Overall, these findings show that context can shape motor sequence representations even when it is irrelevant to motor performance. The SRN results suggest that context dependence is not a fixed property of the task, but an emergent solution dependent on network initialization, which can yield lumped or split representations. Mirroring this distinction, humans exhibited robust costs of contextual switching. By linking these behavioral signatures to white–box measures of hidden–state geometry and context reliance in SRNs, our work connects a computational split–lump solution space to how context organizes motor sequence learning. Ultimately, whether a budding pianist's practice transfers smoothly to a new song with the same underlying fingering may depend not only on the contexts they train in, but also on learner–specific biases that determine whether experience is split or lumped.

## Acknowledgments

We thank Eleanor Holton for her helpful comments on this project. This project was supported by the Innovation Fund for New Ideas in the Natural Sciences from the Office of the Dean of Research at Princeton University. MWG and JAT were also supported by awards from the National Institute of Neurological Disorders and Stroke of the National Institutes of Health R01NS131552 and R01NS134754.

## References

Abrahamse, E. L., Ruitenberg, M. F. L., de Kleine, E., & Verwey, W. B. (2013). Control of automated behavior: Insights from the discrete sequence production task. *Frontiers in Human Neuroscience*, *7*, 82. https://doi.org/10.3389/fnhum.2013.00082

Abrahamse, E. L., & Verwey, W. B. (2008). Context dependent learning in the serial RT task. *Psychological Research*, *72*(4), 397–404. https://doi.org/10.1007/s00426–007–0123–5

Avraham, G., Taylor, J. A., Breska, A., Ivry, R. B., & McDougle, S. D. (2022). Contextual effects in sensorimotor adaptation adhere to associative learning rules. *eLife*, *11*. https://doi.org/10.7554/elife.75801

Bapi, R. S., Doya, K., & Harner, A. M. (2000). Evidence for effector independent and dependent representations and their differential time course of acquisition during motor sequence learning. *Experimental Brain Research*, *132*(2), 149–162. https://doi.org/10.1007/s002219900332

Česonis, J., & Franklin, D. W. (2022). Contextual cues are not unique for motor learning: Task–dependant switching of feedback controllers. *PLOS Computational Biology*, *18*(6), e1010192. https://doi.org/10.1371/journal.pcbi.1010192

Chizat, L., Oyallon, E., & Bach, F. (2019). On Lazy Training in Differentiable Programming. *Advances in Neural Information Processing Systems*, *32*. https://proceedings.neurips.cc/paper/2019/hash/ae614c557843b1df326cb29c57225459–Abstract.html

Cleeremans, A., & McClelland, J. (1991). *Learning The Structure of Event Sequences*.

Crowder, R. G., Serafine, M. L., & Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & Cognition*, *18*(5), 469–476. https://doi.org/10.3758/bf03198480

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Dayan, E., & Cohen, L. G. (2011). Neuroplasticity Subserving Motor Skill Learning. *Neuron*, *72*(3), 443–454. https://doi.org/10.1016/j.neuron.2011.10.008

Dominé, C. C. J., Anguita, N., Proca, A. M., Braun, L., Kunin, D., Mediano, P. A. M., & Saxe, A. M. (2025). *From Lazy to Rich: Exact Learning Dynamics in Deep Linear Networks* (arXiv:2409.14623). arXiv. https://doi.org/10.48550/arXiv.2409.14623

Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., Lehéricy, S., & Benali, H. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behavioural Brain Research*, *199*(1), 61–75. https://doi.org/10.1016/j.bbr.2008.11.012

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*(5), 369–377. (1956–04110–001). https://doi.org/10.1037/h0046888

Godden, D. R., & Baddeley, A. D. (1975). Context–Dependent Memory in Two Natural Environments: On Land and Underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044–8295.1975.tb01468.x

Goodall, C. (1991). Procrustes Methods in the Statistical Analysis of Shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, *53*(2), 285–339.

Haruno, M., Wolpert, D. M., & Kawato, M. (2001). MOSAIC Model for Sensorimotor Learning and Control. *Neural Computation*, *13*(10), 2201–2220. https://doi.org/10.1162/089976601750541778

Heald, J. B., Ingram, J. N., Flanagan, J. R., & Wolpert, D. M. (2018). Multiple motor memories are learned to control different points on a tool. *Nature Human Behaviour*, *2*(4), 300–311. https://doi.org/10.1038/s41562–018–0324–5

Heald, J. B., Lengyel, M., & Wolpert, D. M. (2021). Contextual inference underlies the learning of sensorimotor repertoires. *Nature*, *600*(7889), 489–493. https://doi.org/10.1038/s41586–021–04129–3

Hikosaka, O., Nakamura, K., Sakai, K., & Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*, *12*(2), 217–222. https://doi.org/10.1016/S0959–4388(02)00307–0

Holton, E., Braun, L., Thompson, J. A., Grohn, J., & Summerfield, C. (2026). Humans and neural networks show similar patterns of transfer and interference during continual learning. *Nature Human Behaviour*, *10*(1), 111–125. https://doi.org/10.1038/s41562–025–02318–y

Howard, I. S., Wolpert, D. M., & Franklin, D. W. (2013). The effect of contextual cues on the encoding of motor memories. *Journal of Neurophysiology*, *109*(10), 2632–2644. https://doi.org/10.1152/jn.00773.2012

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Keele, S. W., Jennings, P., Jones, S., Caulton, D., & Cohen, A. (1995). On the Modularity of Sequence Representation. *Journal of Motor Behavior*, *27*(1), 17–30. https://doi.org/10.1080/00222895.1995.9941696

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (arXiv:1412.6980). arXiv. https://doi.org/10.48550/arXiv.1412.6980

Kornblith, S., Norouzi, M., Lee, H., & Hinton, G. (2019). *Similarity of Neural Network Representations Revisited*.

Low, I. I., Giocomo, L. M., & Williams, A. H. (2023). Remapping in a recurrent neural network model of navigation and context inference. *eLife*, *12*, RP86943. https://doi.org/10.7554/eLife.86943

Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science*, *9*(3), 241–289. https://doi.org/10.1016/0167–9457(90)90005–X

Marjieh, R., Jacoby, N., Peterson, J. C., & Griffiths, T. L. (2024). The universal law of generalization holds for naturalistic stimuli. *Journal of Experimental Psychology: General*, *153*(3), 573–589. https://doi.org/10.1037/xge0001533

Masís Obando, J. A., Musslick, S., & Cohen, J. D. (2025). Learning expectations shape cognitive control allocation. *Proceedings of the National Academy of Sciences*, *122*(44), e2416720122. https://doi.org/10.1073/pnas.2416720122

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364–6613(03)00028–7

Morcos, A. S., Raghu, M., & Bengio, S. (2018). *Insights on representational similarity in neural networks with canonical correlation* (arXiv:1806.05759). arXiv. https://doi.org/10.48550/arXiv.1806.05759

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Reder, L. M., Anderson, J. R., & Bjork, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, *102*(4), 648–656. https://doi.org/10.1037/h0036115

Rogers, R., & Monsell, S. (1995). Costs of a Predictable Switch Between Simple Cognitive Tasks. *Journal of Experimental Psychology: General*, *124*, 207–231. https://doi.org/10.1037/0096–3445.124.2.207

Ruitenberg, M. F. L., Abrahamse, E. L., De Kleine, E., & Verwey, W. B. (2012). Context–dependent motor skill: Perceptual processing in memory–based sequence production. *Experimental Brain Research*, *222*(1), 31–40. https://doi.org/10.1007/s00221–012–3193–6

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back–propagating errors*.

Schmidt, L., Lebreton, M., Cléry–Melin, M.–L., Daunizeau, J., & Pessiglione, M. (2012). Neural Mechanisms Underlying Motivation of Mental Versus Physical Effort. *PLoS Biology*, *10*(2), e1001266. https://doi.org/10.1371/journal.pbio.1001266

Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition*, *16*(3), 285–303. https://doi.org/10.1016/0010–0277(84)90031–3

Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody–text integration in memory for songs. *Journal of Memory and Language*, *25*(2), 123–135. https://doi.org/10.1016/0749–596X(86)90025–2

Servan–Schreiber, D., Cleeremans, A., & McClelland, J. (1988). Learning Sequential Structure in Simple Recurrent Networks. *Advances in Neural Information Processing Systems*, *1*. https://papers.neurips.cc/paper_files/paper/1988/hash/9dcb88e0137649590b755372b040afad–Abstract.html

Shea, J., & Morgan, R. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179. https://doi.org/10.1037/0278–7393.5.2.179

Shepard, R. N. (1987). Toward a Universal Law of Generalization for Psychological Science. *Science*, *237*(4820), 1317–1323.

Sobierajewicz, J., Przekoracka–Krawczyk, A., Jaśkowski, W., & van der Lubbe, R. H. J. (2017). How effector–specific is the effect of sequence learning by motor execution and motor imagery? *Experimental Brain Research*, *235*(12), 3757–3769. https://doi.org/10.1007/s00221–017–5096–z

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640. https://doi.org/10.1017/S0140525X01000061

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Ungerleider, L. (2002). Imaging Brain Plasticity during Motor Skill Learning. *Neurobiology of Learning and Memory*, *78*(3), 553–564. https://doi.org/10.1006/nlme.2002.4091

Willingham, D. B., Wells, L. A., Farrell, J. M., & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory & Cognition*, *28*(3), 366–375. https://doi.org/10.3758/BF03198552

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7–8), 1317–1329. https://doi.org/10.1016/S0893–6080(98)00066–5

Woodworth, B., Gunasekar, S., Lee, J. D., Moroshko, E., Savarese, P., Golan, I., Soudry, D., & Srebro, N. (2020). *Kernel and Rich Regimes in Overparametrized Models* (arXiv:2002.09277). arXiv. https://doi.org/10.48550/arXiv.2002.09277

Wright, D. L., & Shea, C. H. (1991). Contextual dependencies in motor skills. *Memory & Cognition*, *19*(4), 361–370. https://doi.org/10.3758/BF03197140

Yang, G. R., Song, H. F., Newsome, W. T., & Wang, X.–J. (2017). *Clustering and compositionality of task representations in a neural network trained to perform many cognitive tasks*. https://doi.org/10.1101/183632