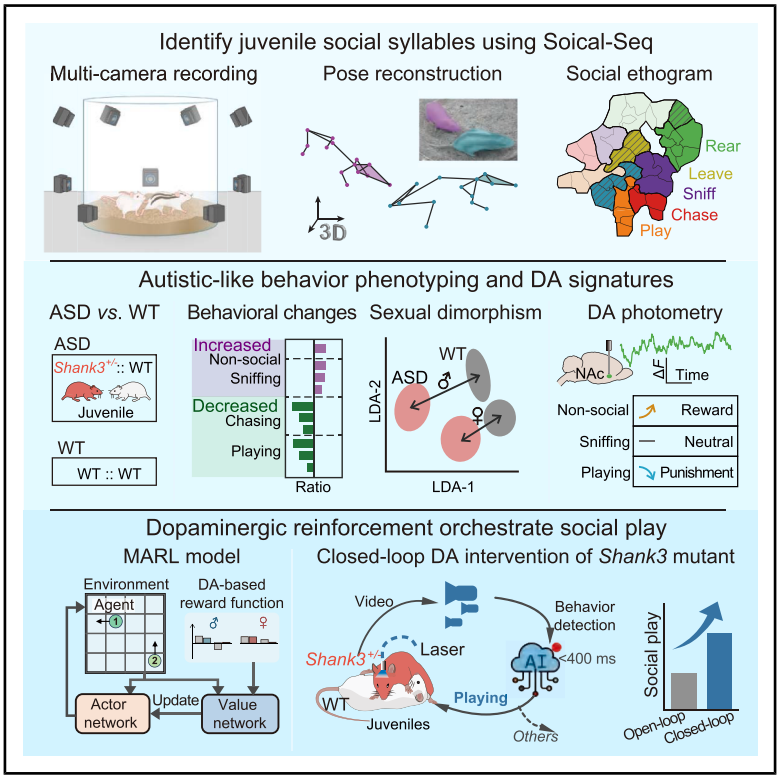
Cell

Sub-second dopamine reinforcement shapes juvenile social play

Graphical abstract



Highlights

- Dopamine reinforces juvenile social play, driving sex-divergent social strategies
- *Shank3* deficiency blunts dopamine in play and enhances it during non-social behaviors
- Reinforcement learning model shows altered reward value drives mutant social deficits
- Behavior-contingent dopamine stimulation restores reciprocal play in *Shank3*$^{+/-}$ rats

Authors

Xinfeng Chen (陈昕枫), Xianming Tao (陶现明), Zhenchao Zhong (钟振超), ..., Wei Xiong (熊巍), Ni Ji (吉妮), Ying Li (李莹)

Correspondence

niji@cibr.ac.cn (N.J.),
liying@cibr.ac.cn (Y.L.)

In brief

Dopamine acts moment by moment to reinforce the specific actions that make up juvenile social play. Disrupted reinforcement underlies social deficits in a *Shank3*-deficient autism model, and correcting it in real time restores normal play.

Chen et al., 2027, Cell *190*, 1–17
January 21, 2027 © 2026 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2026.10.040

CellPress

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

Cell

CellPress
OPEN ACCESS

Article

Sub-second dopamine reinforcement shapes juvenile social play

Xinfeng Chen (陈昕枫),[1,2,3,8,9] Xianming Tao (陶现明),[1,2,3,9] Zhenchao Zhong (钟振超),[2,4] Fanyu Zhu (朱璠宇),[2,3] Yuanqing Zhang (张远青),[2,3,5] Yangzhen Wang (王仰真),[2,3,6] Miao Wang (王淼),[2,3] Yixuan Li (李怡萱),[1,2,3] Le Sun (孙乐),[2,3] Ye Ouyang (欧阳烨),[2,3] Zhaoyi Ding (丁肇一),[2,3] Min An (安敏),[2,3] Yiyi Zhao (赵一一),[7] Jingfeng Zhou (周景峰),[2,3,5] Rong Zhang (张嵘),[7] Wei Xiong (熊巍),[2,3] Ni Ji (吉妮),[2,3,*] and Ying Li (李莹)[2,3,10,*]

[1]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China
[2]Beijing Institute for Brain Research, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 102206, China
[3]Beijing Key Laboratory of Brain Science and Brain-Machines Interface, Chinese Institute for Brain Research, Beijing 102206, China
[4]School of Biomedical Engineering, Capital Medical University, Beijing 100069, China
[5]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University , Beijing, 100875, China
[6]Department of Automation, Tsinghua University, Beijing 100084, China
[7]Neuroscience Research Institute, Key Laboratory for Neuroscience, Ministry of Education of China; Key Laboratory for Neuroscience, National Committee of Health of China; and Department of Neurobiology, School of Basic Medical Sciences, Peking University, Beijing, 100191, China
[8]Present address: Bio-Med Informatics & Clinical Research Center, Xinqiao Hospital, Army Medical University, Chongqing 400037, China
[9]These authors contributed equally
[10]Lead contact
*Correspondence: niji@cibr.ac.cn (N.J.), liying@cibr.ac.cn (Y.L.)
https://doi.org/10.1016/j.cell.2026.10.040

SUMMARY

Juvenile social play is critical for neurodevelopment, yet the neural mechanisms organizing moment-to-moment interactions remain elusive. Using social behavior sequencing (Social-seq)—an occlusion-robust 3D pose estimation and unsupervised clustering pipeline—we found that nucleus accumbens (NAc) dopamine was tightly coupled to sub-second social syllables during naturalistic interactions. In wild-type juveniles, males exhibited a sharper dopaminergic contrast between proactive play and forced submission than females did, consistent with sex-divergent play strategies. In *Shank3*-deficient autism models, this process was fundamentally corrupted: proactive play-evoked dopamine responses were abolished in males and significantly sign-inverted in females, with male-specific enhancements during non-social behaviors, driving social play deficits. A multi-agent reinforcement learning model parameterized by disrupted, play-related dopaminergic reinforcement values recapitulated mutant phenotypes. Closed-loop optogenetic stimulation of ventral tegmental area (VTA)-NAc dopamine projections, time-locked to proactive play, restored reciprocal play in *Shank3*-deficient rats. Sub-second dopaminergic reinforcement thus orchestrates naturalistic social behavior, providing a behaviorally contingent interventional framework for neurodevelopmental disorders.

INTRODUCTION

The juvenile period constitutes a critical development window for neuroplasticity, enabling the rapid acquisition of complex skills through learning.[1] This process is evolutionarily conserved, guiding skill acquisition across domains ranging from the precise correction of vocal syllables in songbirds[2,3] to the refinement of motor proficiency in mammals.[4,5] Mirroring these domains, social competence is similarly consolidated during this phase. Social interactions, particularly juvenile peer play, carry high incentive salience, potentially serving as an essential training ground where experience-dependent plasticity sculpts lifelong social circuits.[6–8] However, the specific neural mechanisms that shape the moment-to-moment organization of discrete social actions during juvenile peer play remain poorly understood.

The mesolimbic dopamine (DA) system, particularly projections from the ventral tegmental area (VTA) to the nucleus accumbens (NAc), is central to social learning. Extensive literature establishes that this circuit dynamically encodes key features of social interactions, with DA activities varying across distinct social contexts and behavioral strategies.[9–13] Crucially, VTA-NAc circuits are sexually dimorphic, exhibiting sex-specific NAc-DA release and neuronal representations.[11,14,15] Despite these insights, whether and how DA signals provide action-contingent reinforcement that biases the acquisition of specific social

This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

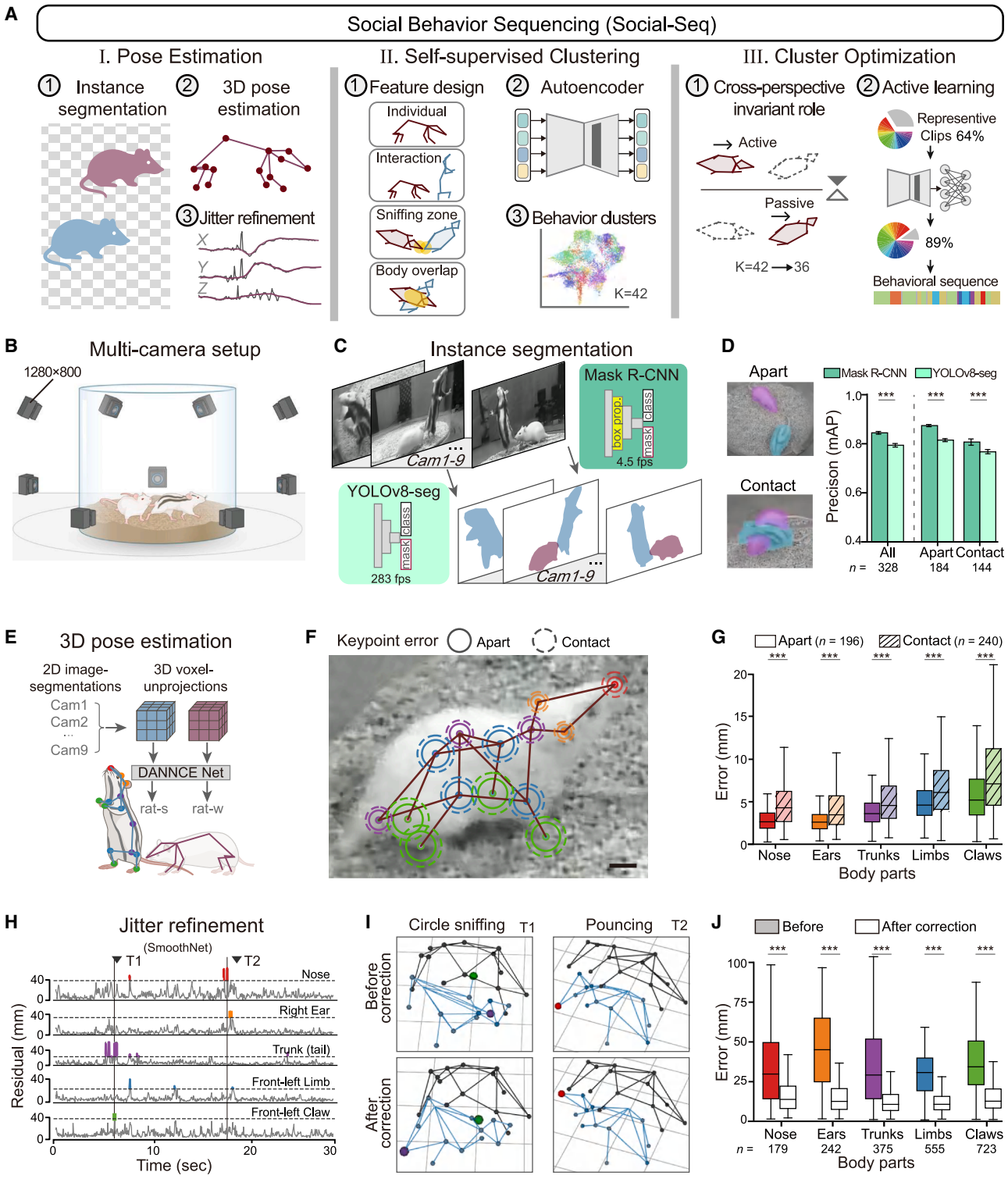


Figure 1. High-resolution 3D pose estimation in freely interacting juvenile rats

(A) Social-seq pipeline.

(B) Experimental setup. A nine-camera array (1,280 × 800 pixels, 120 fps) recorded dyadic interactions in an arena with corncob bedding. For individual identification, one rat received fur stripe markings.

(C) Instance segmentation. YOLOv8-seg (bottom left) enables high-speed processing, whereas Mask R-CNN (top right) yields higher-accuracy masks.

(legend continued on next page)

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

actions—and how this process diverges between sexes—remains largely unexplored.

To causally test these reinforcement rules, we utilized the *Shank3*-deficient rat as a mechanistic perturbation model. Rather than viewing social interaction impairments solely as a static deficit, compelling theoretical frameworks posit that they may arise from disruptions in developmental learning trajectories.[16,17] Notably, *Shank3* deficiency has been linked to impaired postnatal maturation of VTA-NAc circuits, leading to blunted dopaminergic synaptic transmission and altered social engagement.[18–21] However, it remains unclear whether disrupted dopaminergic signaling directly disorganizes the juvenile social repertoire to drive maladaptive strategies in *Shank3*-deficient animals and whether temporally precise DA reinforcement can normalize these interactions.

To link dopamine dynamics to the microstructure of naturalistic peer interactions under heavy occlusion, we developed social behavior sequencing (Social-seq; Figure 1A)—an integrated pipeline combining occlusion-resistant 3D pose estimation, self-supervised learning, and active learning to decompose naturalistic social behavior into sub-second behavioral syllables, while simultaneously monitoring NAc-DA dynamics. Applying this approach to both wild-type (WT) and *Shank3*-deficient juvenile rats,[22] we identified sexually dimorphic social interaction strategies and corresponding DA signatures. In *Shank3*-deficient juveniles, DA responses were action-specifically blunted during proactive social play, with males additionally exhibiting heightened signals during solitary rearing. Incorporating these empirical DA values into a multi-agent reinforcement learning (MARL) model computationally recapitulated the emergence of these distinct strategies. Crucially, closed-loop DA activation precisely timed to proactive play promoted long-lasting amelioration of social play deficits in *Shank3*-deficient males. Together, these findings identify NAc dopamine as a sex-differentiated, action-contingent reinforcement signal shaping juvenile social behaviors, demonstrating that temporally precise DA interventions during development potentially bias social action selection toward more typical trajectories.

RESULTS

Pose estimation during naturalistic juvenile social interactions

Accurate tracking of social interactions relies on resolving precise body kinematics, a challenging task when juveniles become entangled during play. Traditional multi-animal pose estimation methods[23–28] often fail as overlapping bodies obscure anatomical features. To overcome this, we developed a multi-view segmentation-first strategy using a nine-camera volumetric system (Figures S1A–S1E; Videos S1 and S2; STAR Methods). By applying distinct fur-dyeing patterns and instance segmentation, we isolated individual contours before predicting keypoints, eliminating cross-animal interference during contact-rich behaviors (Figures 1B and 1C). Comparison of the social behavioral profiles of dyed and undyed juvenile rats revealed no significant differences (Figure S2A).

To balance computational demands, our framework incorporates complementary segmentation models: mask region-based convolutional neural network (Mask R-CNN)[29] for high-precision offline analysis (84% precision at 4 fps) and YOLOv8[30] for high-speed, closed-loop applications (79% precision at 283 fps) (Figure 1D). Following segmentation, we applied three-dimensional aligned neural network for computational ethology (DANNCE)[31] to track 14 body keypoints (Figure 1E). The model demonstrated high precision for isolated rats but degraded during close contact, particularly for occluded body parts such as claws and limbs (Figures 1F and 1G). These errors manifested as biomechanically implausible trajectories with abrupt frame-to-frame keypoint displacements. To address this, we applied SmoothNet,[32] a temporal refinement algorithm that integrates multi-frame kinematic relationships to detect and eliminate artifactual motion jitters (Figures 1H–1J, S1F, and S1G). The corrected keypoints showed significantly greater agreement with manual annotations (Figure 1J).

Together, these components converged into an offline and a real-time pipeline while maintaining occlusion robustness (Figure S1H; Video S3). System accuracy scaled with camera count and remained high during intense interactions (Figures S1I–S1K). Fur pattern-specific training extended this to four simultaneously interacting animals (Figures S1L and S1M; Video S4), establishing a versatile and generalizable platform for studying naturalistic social behaviors.

Feature-guided self-supervised clustering reveals juvenile social ethogram

We next asked whether unsupervised analysis of kinematic signatures could identify both known and novel social play motifs in juveniles. From the 14 tracked keypoints, we derived 32 behavioral features spanning individual kinematics, inter-animal spatial relations, sniffing engagement, and body-occlusion dynamics

(D) Segmentation performance. Left: representative contours. Right: accuracy quantification for separated (>5 cm apart) vs. contacting rats (paired Student's *t* test).

(E) 3D pose estimation. Multi-view 2D segmentations were integrated into a voxel volume using DANNCE to track 14 landmarks.

(F and G) Pose estimation accuracy. (F) Spatial error distribution showing median errors (solid circles: separated rats; dashed circles: contacting pairs). Scale bar, 10 mm. (G) Euclidean errors of Mask R-CNN- and DANNCE-derived landmarks vs. manual annotations across body parts (unpaired Student's *t* test, false discovery rate [FDR]-corrected).

(H) Jitter detection. Residuals between raw and SmoothNet-predicted trajectories. Traces >2–3 SD (colored) indicate jitter events. Triangles mark example time points T1/T2 analyzed in (I).

(I) Pose corrections. Original (top) and SmoothNet-corrected (bottom) skeletons.

(J) Error reduction following SmoothNet correction (paired Student's *t* test, FDR-corrected).

Data are represented as mean ± SEM. n indicates the number of images/events per panel. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. See Table S4 for comprehensive statistical summaries and effect sizes.

See also Figure S1.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

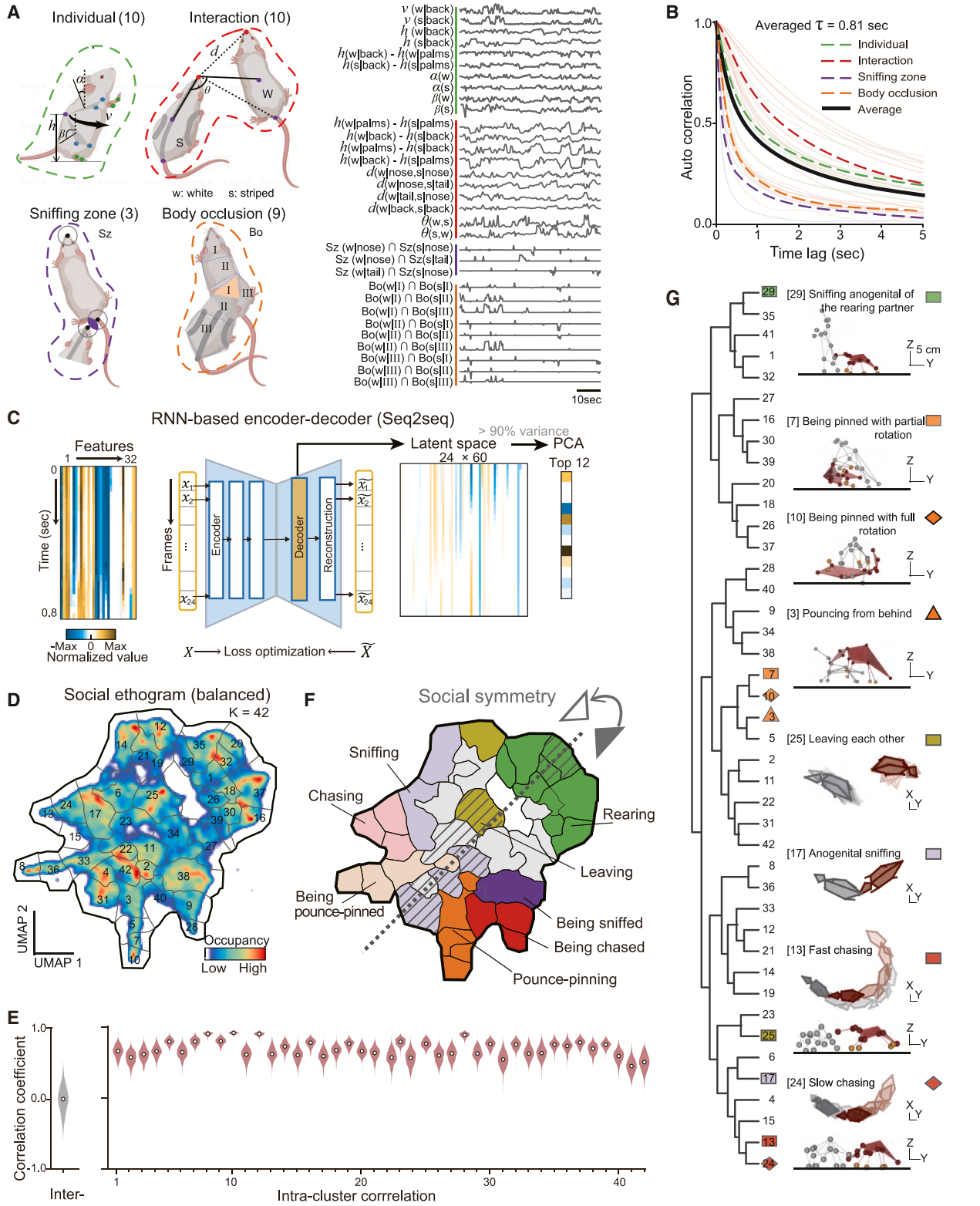


Figure 2. Self-supervised learning of a juvenile social interaction ethogram

(A) Feature extraction of social postures. Parameters included individual kinematics (trunk velocity [v], back height [h], and upper/lower body plane angles [α/β]); interaction metrics (reciprocal nose-back-tail distances [d], angles [θ], and inter-animal height differences); sniffing zones (Sz): defined as nose-nose, nose-tail,

(legend continued on next page)

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

(Figure 2A). Temporal autocorrelation analysis[33] revealed that natural interactions organized at a characteristic timescale of ~800 ms (Figure 2B). We therefore segmented continuous, contact-rich interactions (inter-animal distance <1 body length) into 800-ms action clips, yielding 66,666 clips from 40 juvenile pairs (postnatal day [PND] 50–60; 20 pairs/sex).

To identify discrete interaction categories, we trained a sequence-to-sequence (Seq2Seq) recurrent neural network (RNN) to capture spatiotemporal dynamics and projected latent features into a lower-dimensional space (Figure 2C).[34] This unsupervised approach partitioned the high-dimensional ethogram into 42 behavior clusters, exhibiting significantly higher intra-cluster than inter-cluster correlations (Figures 2D and 2E).

Strikingly, these clusters self-organized into broad ethological categories—such as sniffing, chasing, pouncing, rearing, and leaving—establishing the core modules for our subsequent DA analyses (Figure 2F). Furthermore, the uniform manifold approximation and projection (UMAP) geometrically segregated reciprocal social roles, differentiating active execution from passive reception at symmetric positions (Figure S2C; Video S5). Hierarchical clustering further exposed micro-behavioral variants (Figures 2G and S2B; Table S1). For example, pinning diverged by the pinned rat's rotation angle (*partial* vs. *full*), and chasing split into *fast* vs. *slow* based on acceleration profiles. These granular distinctions, validated against ethologists' annotations,[35] demonstrate our framework's power to resolve subtle interaction mechanics—differences often lost in manual annotations.

Cluster optimization through perspective-invariant labeling

While unsupervised clustering successfully identified ethologically relevant behaviors, substantial ambiguity remained at category boundaries, in which individual video clips often diverged from cluster prototypes. Given the intrinsic reciprocity of social interactions, we enforced perspective-invariant labeling—requiring consistent classification from both interacting rats' viewpoints—to yield high-confidence behavioral motifs (Figures 3A and 3B).

By establishing a cross-perspective consistency matrix, we identified reliable clusters (>50% inter-perspective overlap) encompassing both "mutual behaviors" (e.g., head-to-head approach) and "reciprocal role pairs" (e.g., chasing vs. being chased) (Figures 3A–3C; Table S1). Within these, representative clips with consistent cross-perspective labels exhibited higher pairwise latent feature correlations, lower entropy, and tighter trajectories, underscoring their reliability (Figures 3D, 3E, and S3A–S3C).

To optimize the classification of the remaining ambiguous clips, we implemented an automated refinement procedure. We trained a fully connected classifier network (FCN) on the high-consistency representative clips and iteratively updated the latent space (Figure 3F). Within four iterations, the perspective-invariant classification of proximal interactions improved from 64.0% to 88.7% (Figure 3H). This refinement significantly improved cluster validity indices, increasing the Calinski-Harabasz index (CHI) by 247% and intra-cluster correlation by 10.9% (Figure 3G). The resulting optimized Seq2Seq fully connected network (Seq2Seq-FCN) ethogram, comprising 36 robust behavioral clusters that generalize to distal interactions (Figures S3D and S3E), establishes a powerful foundation for dissecting social behavioral profiles across sexes and contexts.

Sex-dependent dopaminergic encoding of naturalistic social syllables

We began by characterizing naturalistic social interactions in same-sex WT juvenile rats (PND 50–60), identifying marked sex differences in behavioral profiles. At the population level, male pairs exhibited a stronger bias toward frequent, contact-rich interactions dominated by approaching, investigatory sniffing followed by pouncing and pinning—a pattern characteristic of proactive social play. Conversely, female pairs showed higher relative frequencies of distal interactions and environmental exploration, characterized by frequent chasing/leaving and solitary rearing (Figures 4A–4C and S2D–S2J). These distinct behavioral architectures prompted us to investigate whether the dopaminergic correlates of specific social microstructures ("syllables") are sexually divergent.

To characterize NAc dopamine dynamics during naturalistic social interactions, we expressed the G-protein-coupled receptor activation-based dopamine sensor 3m (GRAB-DA3m)[36] in the NAc core and monitored fluctuations via fiber photometry (Figure 4D). Following validation using rewarding and aversive tastants (Figure S6A), we explored how NAc dopamine fluctuated in different social contexts. Combining photometry with our Seq2seq-FCN behavioral classification framework, we observed that DA responses tracked running velocity during locomotor-related episodes (e.g., chasing) and aligned with

and tail-nose overlaps (radii = average nose-ear/tail-claw distances); and body occlusions (Bo): calculated as XY-plane overlaps across head (I), upper (II), and lower trunk (III). Right: temporal dynamics of these features.

(B) Temporal autocorrelations. Auto-correlograms of posture-related features declined and were fit with an exponentially decaying function to estimate the time-constant (τ). Average τ = 0.81 s.

(C) Seq2seq architecture. A three-layer RNN-based encoder-decoder model processed 0.8-s clips (24 frames × 32 features) to generate 24 × 60 latent representations, subsequently reduced to 12 PCs capturing >90% of variance.

(D) Unsupervised behavioral clustering (K-means, K = 42) of 66,666 clips subsampled from 86,560 clips across 40 rat pairs (20 pairs/sex), visualized via UMAP with overlaid cluster boundaries.

(E) Cluster validation. Violin plots comparing pairwise correlations of latent vectors (12 PCs) for intra-cluster (red) vs. inter-cluster (gray) clip pairs (n = 200 pairs/distribution).

(F) UMAP projection revealing symmetric action patterns (dashed line) and mutual interactions (slash). Eight core behavioral modules used in subsequent photometry analysis are highlighted.

(G) Behavioral hierarchy. Left: dendrogram of cluster relationships based on pairwise Euclidean distances between latent centroids. Right: representative 3D skeletons.

See also Figure S2.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

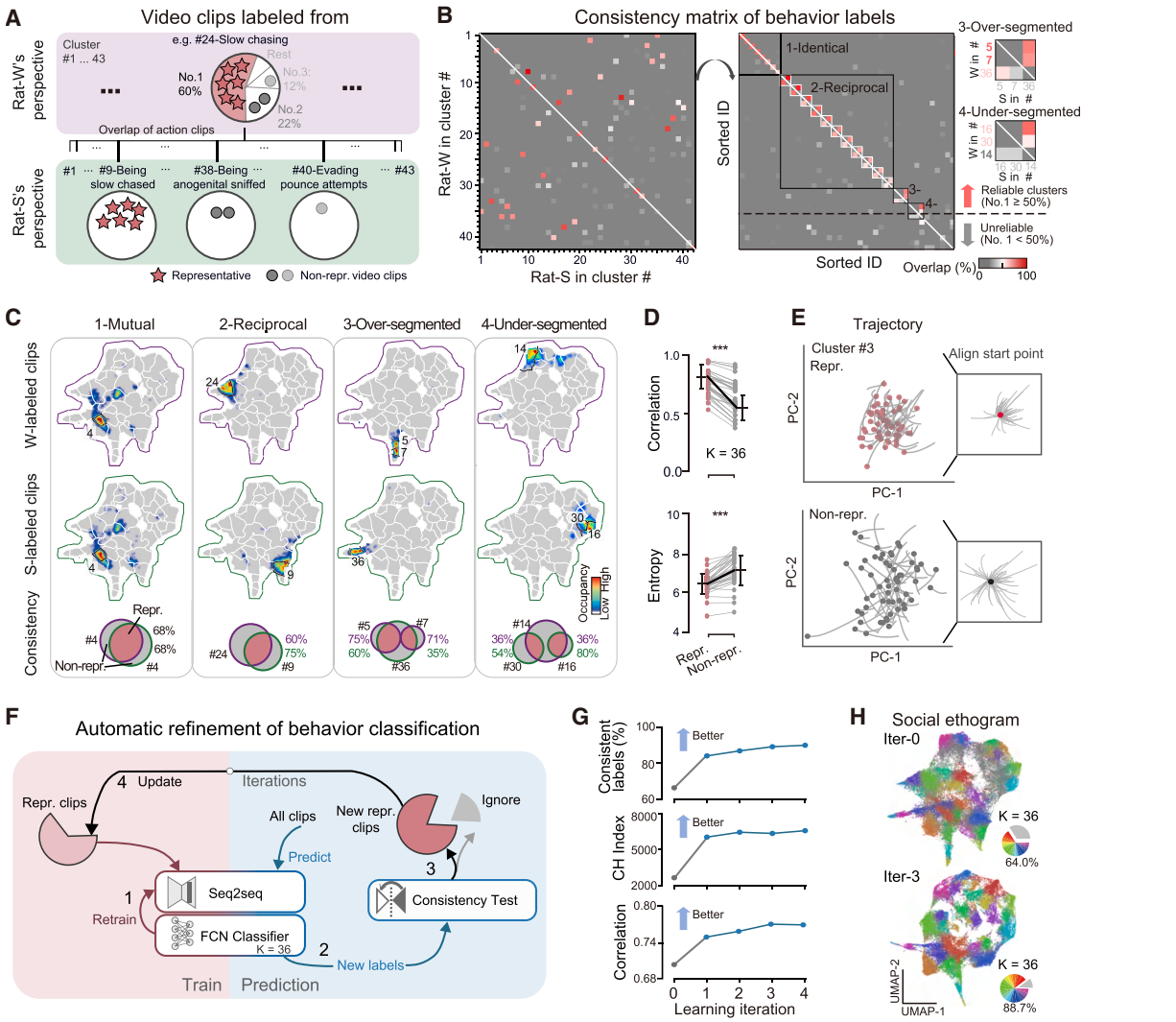


Figure 3. Active learning-based refinement of behavioral clusters

(A) Cross-perspective validation comparing video clips from white (w) and striped (s) rat perspectives. Clusters with >50% inter-perspective overlap were deemed reliable (representative).

(B) Cross-perspective consistency matrix of w- and s-labeled clips identifying 36 of 42 clusters as reliable. Four relationships emerged: mutual (8 clusters, high w-s overlap), reciprocal (12 clusters, matching reciprocal clusters), over-segmented (multiple [clusters #5 and #7] mapping to one [#36]), and under-segmented (one [cluster #14] mapping to multiple [#16 and #30]).

(C) UMAP projections of w- (top) and s-labeled (bottom) clips for representative cluster pairs from (B). Venn diagrams show inter-perspective overlap.

(D) Intra-cluster correlation (top) and entropy (bottom) for representative vs. non-representative clips across 36 clusters (paired Student's *t* test, ***$p < 0.001$). Data are represented as mean ± SEM.

(E) Decoder trajectories in PC space. Dots indicate starting positions; insets show aligned trajectories.

(F) Active learning pipeline. A Seq2seq-FCN framework iteratively improves labeling through: (1) supervised retraining of the Seq2seq-FCN on high-confidence clips; (2) prediction of new labels for all clips; (3) identification of reliable samples via cross-perspective consistency tests; (4) updating the high-confidence set with reliable samples for the next iteration.

(G) Performance metrics across iterations. Representative clip proportion (top), Calinski-Harabasz index (CHI; middle), and average intra-cluster correlation (bottom) plateaued after four iterations.

(H) UMAP visualization of clip labels before (iter-0, top) and after refinement (iter-3, bottom), with pie charts highlighting the increased proportion of representative clips (colored).

See also Figure S3.

body-occlusion levels during intense physical contact (e.g., pouncing; Figures 4E and S4A–S4D).

Feature-aligned analysis (STAR Methods; Figures S4E and S4F) revealed distinct sexual dimorphism in social syllable-related DA fluctuations (Figures 4F and 4G). In males, approach (ΔF/F = 0.26 ± 0.12, $p = 0.03$, peak latency [PL] = 0.80 ± 0.03 s) and proactive behaviors (sniffing, ΔF/F = 0.30 ± 0.07, $p = 0.0004$, PL = 0.94 ± 0.03 s; chasing, ΔF/F = 0.25 ± 0.11, $p = 0.04$, PL =

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

CellPress
OPEN ACCESS

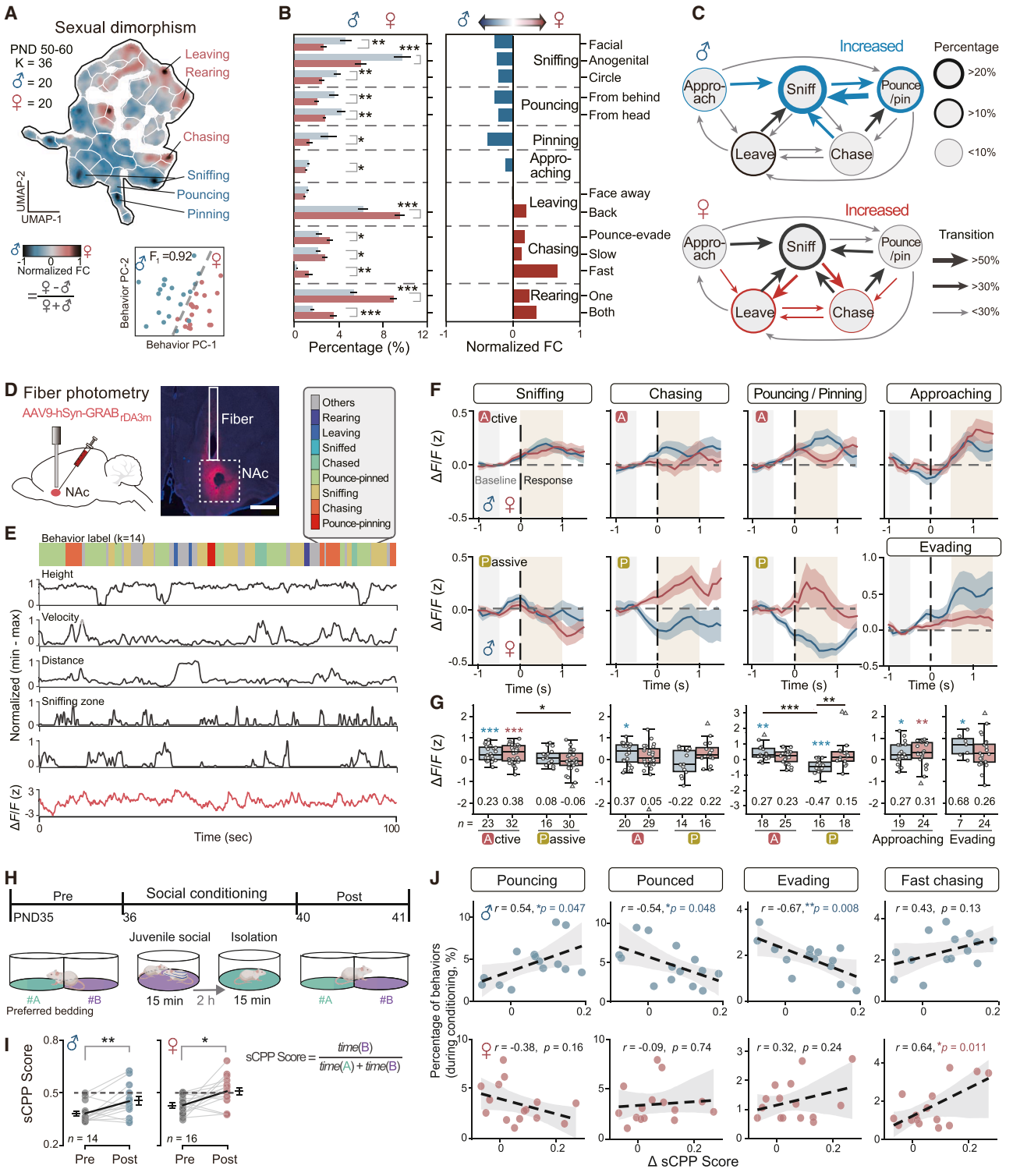


Figure 4. Sexually dimorphic behavioral and dopaminergic signatures in juvenile social interactions

(A) Sex differences in behavioral space. Top: smoothed fold differences in UMAP cluster occupancy between male and female rats (20 pairs/sex). Bottom: behavioral organization projected in low-dimensional space. SVM classification boundary (dashed line; F1 score via leave-one-out cross-validation) separating sex-specific profiles.

(legend continued on next page)

0.78 ± 0.04 s; pouncing, ΔF/F = 0.35 ± 0.09, $p = 0.002$, PL = 0.94 ± 0.05 s) elicited pronounced DA surges relative to baseline, comparable to sucrose-evoked signals (Figure S6A). By contrast, passive behaviors typically did not elicit significant DA fluctuations. Notably, forced submission (being firmly pounced) coincided with DA suppression (ΔF/F = −0.46 ± 0.11, $p = 0.0006$, PL = 0.88 ± 0.04 s), like quinine (Figure S6A). This suppression was state-dependent: while forced restraint coincided with a decrease in DA, successful evasion from a pounce or a chase was associated with a trend toward a signal rebound (ΔF/F = 0.63 ± 0.19, $p = 0.02$, PL = 0.70 ± 0.10 s).

In females, initial social approach (ΔF/F = 0.35 ± 0.10, $p = 0.002$, PL = 0.87 ± 0.04 s) and sniffing (ΔF/F = 0.28 ± 0.07, $p = 0.0006$, PL = 0.88 ± 0.03 s) elicited pronounced DA surges, whereas intense proactive behaviors (chasing and pouncing) evoked only blunted, non-significant responses (chasing: ΔF/F = 0.18 ± 0.09, $p = 0.054$, PL = 0.79 ± 0.03 s and pouncing: ΔF/F = 0.17 ± 0.09, $p = 0.074$, PL = 0.91 ± 0.04 s; Figures 4F and 4G). Notably, the most pronounced sexual dimorphism emerged during forced submission: whereas being firmly pounced evoked a robust aversive DA suppression in males, females exhibited no significant DA modulation during the identical passive event (male vs. female: −0.46 ± 0.11 vs. 0.07 ± 0.11, $p = 0.0065$, Cohen's d = −1.22; Figures 4F and 4G), consistent with a sex-divergent valuation of submissive outcomes. Instead, in females, heightened DA transients emerged at the initiation of active evasion (ΔF/F = 0.28 ± 0.13, $p = 0.05$, PL = 0.85 ± 0.06 s), leaving the partner (ΔF/F = 0.26 ± 0.08, $p = 0.002$, PL = 0.87 ± 0.04 s) and solitary rearing (ΔF/F = 0.21 ± 0.05, $p = 0.0004$, PL = 1.18 ± 0.02 s) (Figures 4F, 4G, 6E, and 6F). These observations further suggest that female dopaminergic responses preferentially time-lock to social avoidance rather than contact-heavy interactions.

To examine how these syllables relate to appetitive learning, we employed a social conditioned place preference (sCPP) paradigm (Figure 4H; STAR Methods). Following 5 days of conditioning, both sexes exhibited significant sCPP (Figure 4I). In males, preference magnitude positively correlated with proactive play (pouncing/pinning), whereas passive pouncing and evasion correlated with attenuated preference. In females, sCPP magnitude instead correlated with active fast chasing (Figure 4J), highlighting an appetitive, non-contact form of engagement. Collectively, these findings suggest that NAc dopamine exhibits sex-differentiated, action-aligned fluctuations consistent with a reinforcement signal shaping social action selection.

Social microstructural breakdowns in *Shank3*-deficient juveniles

Building on the premise that dopamine dynamically guides social sequences, we hypothesized that altered NAc dopamine signals contribute to microstructural social deficits in autism models. To explore this, we applied our platform to *Shank3*$^{+/-}$ rats (exons 11–21 deletion) to determine how interaction diverges across development (Figure 5A) and whether juvenile deficits correlate with dysregulated DA dynamics (Figure 6).

Across developmental stages, Shank3$^{+/-}$ rats interacting with WT peers showed age-dependent attenuation of social interaction deficits (Figures 5B–5G, S5A, and S5B). At the early juvenile stage (PND 35–45), both sexes exhibited a marked reduction in play-related syllables (e.g., chasing and pouncing decreased by 49.8%, 49.7% in males; 32.8%, 26.7% in females; Figure 5C), despite maintaining or exceeding WT levels of investigatory sniffing. By the late juvenile stage (PND 55–65), these deficits attenuated yet persisted (Figure 5E). In early adulthood (PND 75–85), although mild pouncing deficits persisted in both sexes, severe pinning impairment was retained exclusively by Shank3$^{+/-}$ males (reduced by 84.3%), while largely preserved in females (Figure 5G).

Behavioral space analysis demonstrated that WT and Shank3$^{+/-}$ divergence peaked during the juvenile stage, with males showing more profound deficits (Figures 5H and 5I). Transition analysis further revealed that Shank3$^{+/-}$ juveniles exhibited a strong backward progression from play/chasing into sniffing, whereas WT juveniles showed the opposite (Figures 5J and S5C). Strikingly, although WT partners adapted by increasing active play solicitations, Shank3$^{+/-}$ juveniles rarely sustained

(B and C) Behavioral dimorphism and state transitions. (B) Behavioral cluster proportion (left) and fold differences (right) for grouped behaviors. (C) Behavioral transition probabilities for male (top) and female (bottom) pairs. Significant sex differences ($p < 0.05$, Kruskal-Wallis, and post hoc Dunn's test, FDR-corrected) are color-coded.

(D) Viral expression and fiber placement. Left: sagittal adeno-associated virus (AAV) injection schematic. Right: representative coronal section confirming GRAB-DA3m expression (red) and NAc core fiber placement. Scale bar, 1 mm.

(E) Representative DA recording. Time-aligned behavioral labels (top), kinematic variables (black), and *Z* scored ΔF/F (red) from a male pair. Photometry signals were rigorously corrected for motion artifacts using the 405 nm isosbestic channel.

(F) Peri-event *Z* scored DA dynamics aligned to behavior onset (t = 0, dashed black line). Traces represent mean ± SEM (male: blue; female: red; ≥3 epochs/video). Shaded background blocks demarcate precise analysis windows: baseline (gray, −1 to −0.5 s) and post-onset response (yellow, approaching/evading, 0.5 to 1.5 s; other behaviors, 0 to 1 s).

(G) Peak *Z* scored ΔF/F. Boxplot represents the 25th and 75th percentiles (interquartile range [IQR]) and median value (black numbers), with whiskers at 1.5 × IQR. Outliers beyond whiskers (triangles) were excluded from statistical analyses. Colored and black asterisks indicate within-group (response vs. baseline, paired Student's *t* test) and between-group significance (two-way ANOVA), respectively.

(H) sCPP experimental timeline (top) and protocol (bottom). Animals were deliberately conditioned in their initially less-preferred chamber to rigorously assess social reward.

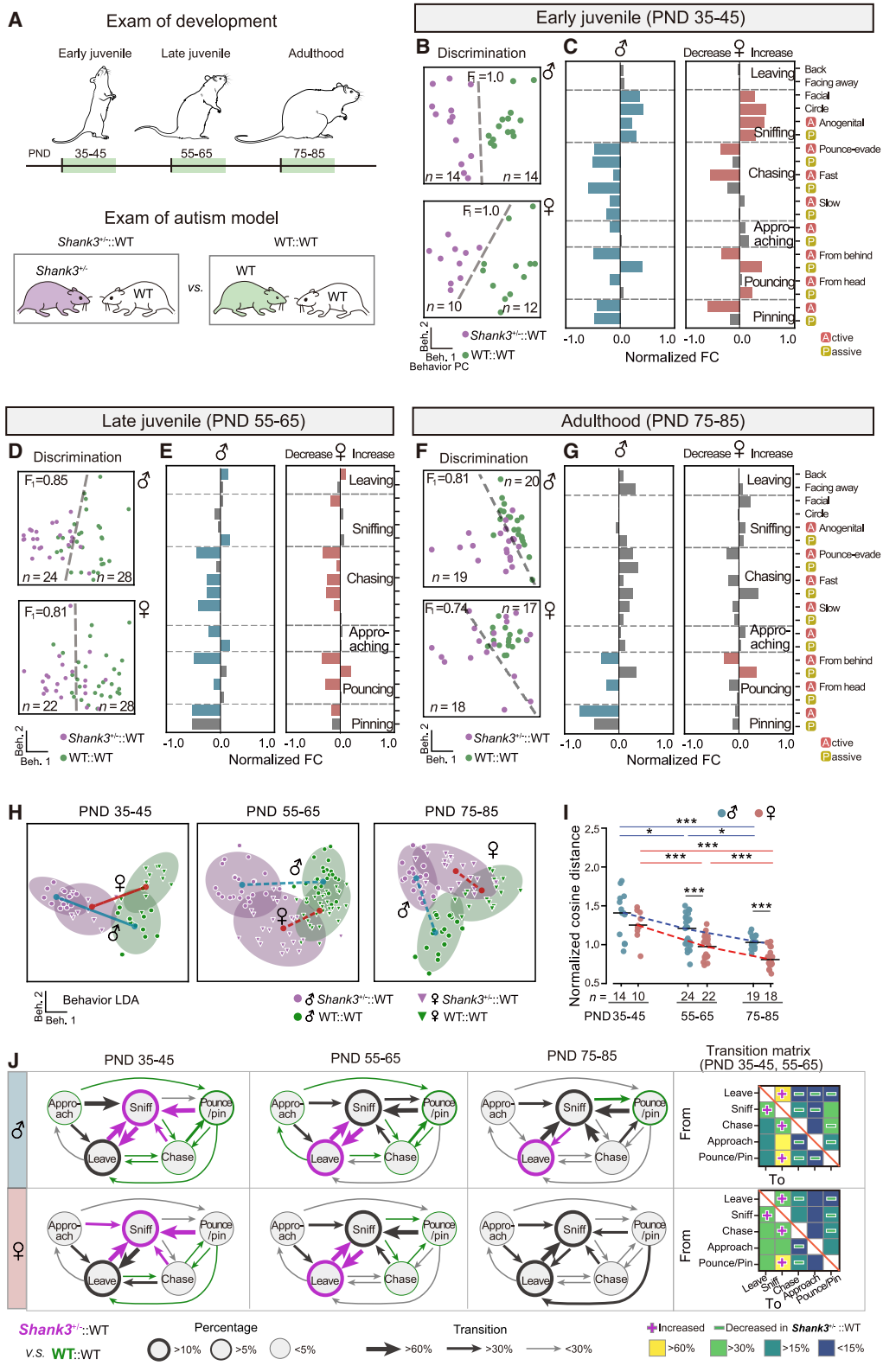
(I) Pre- vs. post-conditioning sCPP scores. The 0.5 dashed line indicates chance-level preference. The pre-test mean (<0.5) arises from initial pairing with the non-preferred chamber. Paired Student's *t* test.

(J) Pearson correlations (r and p values shown) between individual sCPP score changes and behavioral syllable time fractions (averaged across all conditioning sessions). Dashed lines show linear fits; shading indicates 95% confidence interval.

Data are represented as mean ± SEM. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. See Table S4 for comprehensive statistical details and effect sizes.
See also Figures S4 and S6.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

engagement, suggesting a deficit in reciprocating play sequences (Figures S5D and S5E).

To assess whether this microstructural breakdown is associated with an absence of contextual preference driven by peer interaction, we evaluated these dyads using the sCPP paradigm. We found that, particularly among males, neither Shank3$^{+/-}$ nor their WT partners developed a preference for the peer-associated context (Figures S5F and S5G). Together, these observations suggest that disrupted NAc dopaminergic valuation in Shank3$^{+/-}$ juveniles collapses reciprocal play sequences, thereby degrading the interaction reward for the WT partner.

Altered dopamine dynamics track social play deficits in *Shank3*$^{+/-}$ juveniles

To investigate the neural mechanisms underlying social deficits in *Shank3*$^{+/-}$, we recorded NAc-DA dynamics in *Shank3*$^{+/-}$ juveniles interacting with WT peers. In a non-social context, *Shank3*$^{+/-}$ juveniles maintained typical DA suppression to quinine but exhibited an abolished response to sucrose (Figure S6B), aligning with reports that *Shank3*-VTA deletion mice lack sucrose preference.[18] In social contexts, *Shank3*$^{+/-}$ animals exhibited normal DA responses (*Z* scored ΔF/F) during social approach (Figure 6A; males: 0.28 ± 0.17; females: 0.17 ± 0.12), but a more pronounced reduction during active sniffing compared with WT controls (Figure 6B; male WT vs. *Shank3*$^{+/-}$: 0.30 ± 0.07 vs. −0.08 ± 0.06, $p = 0.0007$, Cohen's d = −1.25; female WT vs. *Shank3*$^{+/-}$: 0.28 ± 0.07 vs. 0.07 ± 0.08, $p = 0.06$, Cohen's d = −0.52). Similarly, during active chasing, *Shank3*$^{+/-}$ males, but not females, exhibited a trend of DA reduction (Figure 6C).

The most striking divergence emerged during social play: instead of the typical play-induced DA release observed in WT animals, proactive play-evoked dopamine responses were abolished in males and significantly sign-inverted in females (Figure 6D; male WT vs. *Shank3*$^{+/-}$: 0.35 ± 0.09 vs. −0.17 ± 0.12, $p = 0.002$, Cohen's d = −1.36; female WT vs. *Shank3*$^{+/-}$: 0.17 ± 0.09 vs. −0.44 ± 0.09, $p < 0.0001$, Cohen's d = −1.44). However, during reactive behaviors (being pounced and evading in males; being chased in females), *Shank3*$^{+/-}$ animals lacked the characteristic DA fluctuations seen in WT controls (Figures S6C–S6F). Conversely, non-social behaviors, particularly solitary rearing, elicited enhanced DA responses (*Z* scored ΔF/F) selectively in *Shank3*$^{+/-}$ males (WT vs. *Shank3*$^{+/-}$: 0.01 ± 0.07 vs. 0.29 ± 0.11, $p = 0.04$, Cohen's d = 0.73) (Figures 6E and 6F), despite no significant differences in their corresponding kinematic features (Figure S6G).

We next evaluated whether DA magnitudes scaled with behavioral duration. For solitary behaviors (i.e., rearing), DA positively tracked duration (Figure S6H). However, this relationship was uncoupled for social behaviors (Figure S6I). Crucially, DA magnitudes during active pouncing positively correlated with the proportion of pounces preceded by social approach ($r = 0.32$, $p = 0.005$; Figure S6J). These results suggest that juvenile NAc DA potentially serves as a reinforcement signal driving the transition from approach to active play.

To further test whether these aberrant DA signatures are computationally sufficient to generate the observed phenotypes, we trained artificial agents using MARL in a simulated "social" environment.[37,38] Pairs of agents were designed to mimic juvenile behaviors (including sniffing, pouncing, chasing, approaching, and leaving) in a fully observable, compact grid world (Figures 6G and 6H). Both agents were trained to maximize their own rewards using proximal policy optimization (PPO)[39]—a robust and stable policy gradient method based on an actor-critic architecture.[40,41]

To model genotypic and sex-specific phenotypes, we evaluated four cohorts of agent pairs: male/female WT, male/female *Shank3*. To bridge empirical physiological data with the MARL model, we used a hybrid reward parameterization. The reinforcement parameters for the primary play-related syllables—pouncing and being pounced—were linearly scaled from the corresponding empirical DA responding amplitudes, whereas the reward parameters for the other six ancillary behaviors were calibrated to support stable learning and preserve the broader organization of the social repertoire (Figure 6I; STAR Methods). By analyzing the temporal evolution of their interactions throughout training (up to 60,000 steps), the biologically constrained MARL agents reproduced key features of the empirical phenotypes. For instance, compared with WT controls, *Shank3*$^{+/-}$-derived agents exhibited a faster progressive reduction in active pouncing and chasing (Figure S6K).

Evaluating across three independent seeds (STAR Methods), the simulated agents reliably reproduced divergent behavioral profiles across sex and genotype (Figure 6J). Specifically, agents driven by WT male parameters converged on contact-rich strategies characterized by frequent proactive play, whereas

Figure 5. Age- and sex-dependent social behavior deficits in *Shank3*-deficient rats

(A) Experimental timeline across early/late juvenile and adulthood stages.

(B, D, and F) Principal component analysis (PCA) of behavioral profiles showing distinct separation between *Shank3*$^{+/-}$:WT (purple) and WT::WT (green) dyads in males (top) and females (bottom). Dashed lines indicate SVM classifier boundaries.

(C, E, and G) Genotype-induced behavioral alterations across development. Color codes indicate significant changes in males (left) and females (right) ($p < 0.05$, Kruskal-Wallis and post hoc Dunn's test, FDR-corrected).

(H) Linear discriminant analysis (LDA) of behavioral profiles. Distinct clustering of *Shank3*$^{+/-}$:WT (purple) and WT::WT (green) groups in males (circles) and females (triangles). Ellipses represent 90% confidence intervals. Dashed lines indicate centroid distances.

(I) Behavioral profile divergence. Cosine distance between *Shank3*$^{+/-}$:WT and WT::WT groups, normalized to intra-group distances (males: blue; females: red; two-way ANOVA).

(J) Behavioral state transitions detailing the probability of progressing between proactive actions. Right: transition matrices capturing the disrupted progression from investigation to sustained play in *Shank3*$^{+/-}$ juveniles. Significant genotype differences (unpaired Student's *t* test, FDR-corrected, $p < 0.05$) are marked (*Shank3*$^{+/-}$ increase/decrease: purple [+]/green [−]).

Sample sizes (*n*) are indicated in the respective panels. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. See Table S4 for comprehensive statistical details and effect sizes. See also Figure S5.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

CellPress
OPEN ACCESS

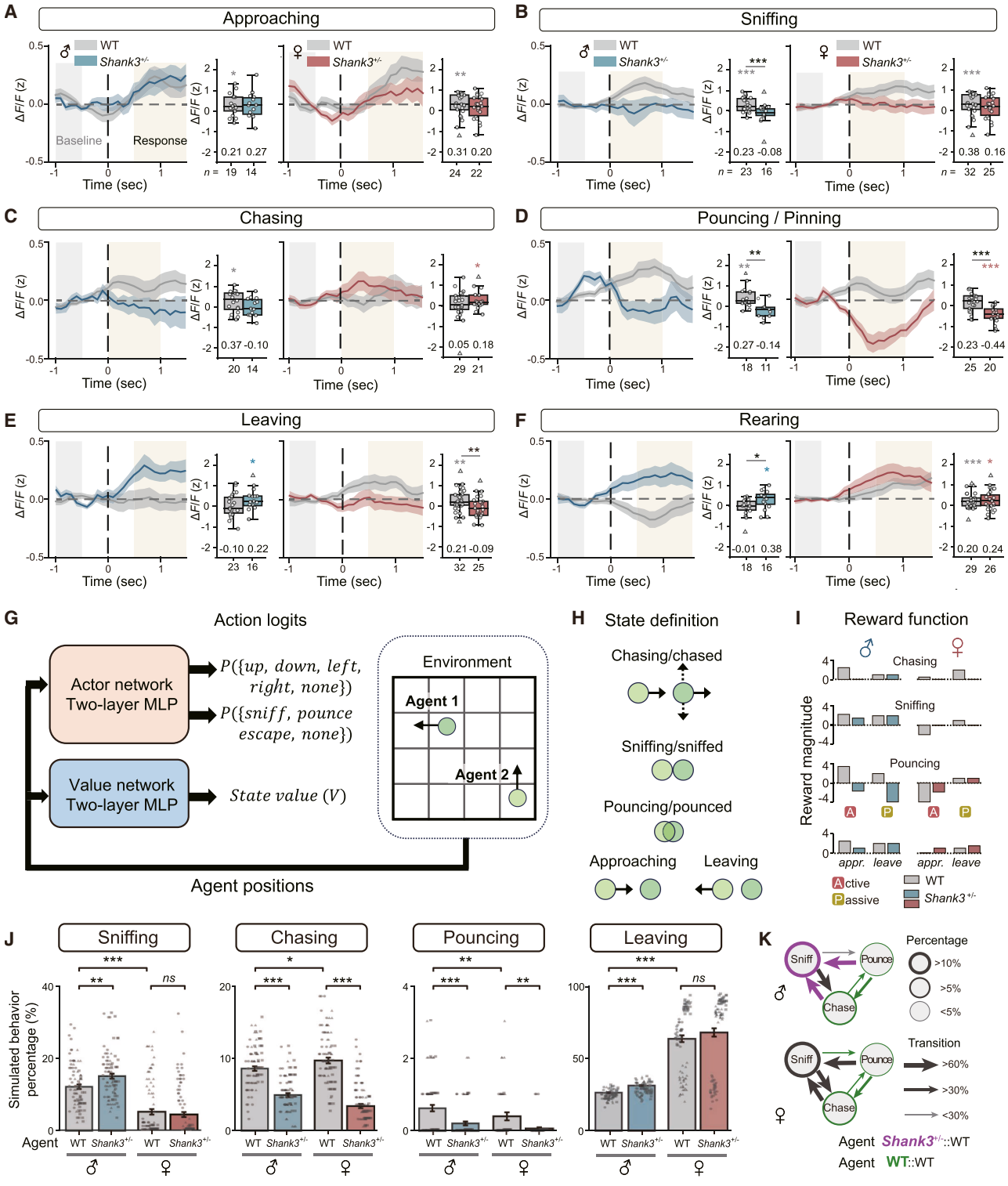


Figure 6. Dopaminergic responses across different behavioral categories in WT and *Shank3*$^{+/-}$ juvenile rats

(A–F) Time-resolved dopaminergic dynamics across different behavioral categories.

Peri-event DA dynamics aligned to behavior onset (t = 0, dashed black line). Traces represent mean ± SEM (WT: gray; Shank3$^{+/-}$ males/females: blue/red; ≥3 epochs/video). Shaded background blocks demarcate analysis windows: baseline (gray, −1 to −0.5 s) and post-onset response (yellow, approaching/leaving/rearing, 0.5 to 1.5 s; other behaviors, 0 to 1 s). Boxplots represent the 25th and 75th percentiles (IQR) and median values (black numbers), with whiskers at

(legend continued on next page)

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

WT female parameters shifted the agents toward non-contacting strategies favoring chasing and leaving the partner. When driven by *Shank3*-deficient DA values, agents robustly reproduced key pathological features—preserved sniffing alongside severely reduced proactive play. The model also captured the characteristic breakdown in transition probabilities from sniffing to pouncing in mutants (Figure 6K). Together, these results indicate that, under the model's assumptions, a dopamine-informed hybrid reinforcement framework can reproduce play-deficient, transition-disrupted social strategies and their sex-dependent expression, suggesting that disrupted DA dynamics may induce maladaptive dyadic policy learning in juvenile *Shank3*$^{+/-}$ rats.

Closed-loop activation of the VTA-NAc pathway biases juvenile play microstructure in *Shank3*$^{+/-}$ mutants

To test if temporally precise DA signaling modifies social behavioral microstructure, we implemented closed-loop optogenetics to drive VTA tyrosine hydroxylase (TH)-positive (TH^{+}) projections in the NAc core during naturalistic social interactions (Figures 7B–7D; STAR Methods). Male *Shank3*$^{+/-}$ juveniles interacted daily with WT partners for 15 min over 8 consecutive days (training), followed by a 10-day stimulation-free observation (post) (Figure 7A).

Our end-to-end real-time pipeline—comprising multi-camera streams, YOLOv8 segmentation, DANNCE pose estimation, feature extraction, and bidirectional long short-term memory FCN (BiLSTM-FCN) classification—achieved a processing latency of 266 ± 32 ms (mean ± SD) (Figures 7E, 7F, S7A, and S7B; Video S6). By evaluating a continuous 400-ms sliding history window, the classifier maintained high behavioral specificity with an average F1 score of 0.8 (Figures S7C and S7D). Consequently, laser stimulation was typically initiated 633 ± 369 ms (mean ± SD) after the physical onset of the target behavior (e.g., proactive pouncing) (Figure S7E). This sub-second precision ensured that closed-loop optogenetic reinforcement was delivered concurrently with the ongoing execution phase of the targeted action.

Using this system, we delivered 40-Hz stimulation specifically during deficient proactive play (pouncing/pinning), which comprised 12.1% ± 1.0% of total behavioral testing time (Figures 7G–7I). Open-loop controls received temporally unpatterned stimulation matched in duration (11.2% ± 0.3%; Figure S7F). Closed-loop stimulation delivered 19.6% ± 1.8% of its pulses during proactive play bouts, compared with only 0.7% ± 0.1% in open-loop controls, yielding 90.7% ± 1.0% true positives and 12.1% ± 0.8% false positives (Figures 7H and 7I), illustrating strict behavioral contingency.

Closed-loop activation resulted in persistent changes during both training and stimulation-free post-training periods (Figures 7J, 7K, and S7G). The most pronounced effects occurred in targeted proactive play (pouncing/pinning), which increased significantly in both relative and absolute durations across training sessions (Figures 7L–7N, S7H, and S7I). Crucially, these behavioral adaptations were highly action-specific. While targeted play increased, non-targeted investigatory behaviors (e.g., sniffing, chasing) remained unchanged or even slightly decreased post training (Figure 7N). Similarly, stimulated mutants exhibited significant reductions in non-social behaviors (e.g., leaving the partner, distal non-contacting behaviors) that occupy spatially distinct regions in UMAP space and are characterized by larger inter-animal distances (Figures 7M and 7N).

Intriguingly, WT partners interacting with stimulated mutants also exhibited altered behavioral profiles post training, with slightly increased pinning and decreased sniffing (Figure S7J). This suggests that resolving the micro-level DA valuation deficit in mutants may help restore the dyadic feedback loop, driving enhanced overall social engagement. Together, these results support the hypothesis that precisely timed DA release during play behaviors can bias juvenile dyadic interaction structure toward increased play engagement in *Shank3*$^{+/-}$ mutants.

DISCUSSION

By integrating ethologically relevant behaviors, dopaminergic dynamics, and temporally precise circuit manipulations, our findings support a role for NAc dopamine as an action-specific reinforcement signal shaping naturalistic social strategies during development.[42,43] Notably, we identify a specific disruption of DA-behavior coupling—specifically, blunted (male) and even sign-inverted (female) dopamine responses during proactive play—in *Shank3*-deficient models, mirroring atypical social valuation in autistic children.[44,45] Furthermore, computational modeling using a dopamine-informed MARL framework and closed-loop interventions reveal that the assignment of play-associated DA reinforcement value, under model assumptions, adaptively shapes macroscopic social interaction profiles. Together, these findings provide a quantitative link between sub-second dopaminergic reinforcement signals and dyadic social strategies, supporting the feasibility of temporally contingent DA-pathway interventions to ameliorate juvenile play deficits in *Shank3*-deficient rats.

Identification of juvenile-typical social ethogram

Decoding reciprocal social play requires resolving sub-second behavioral transitions. While recent breakthroughs like social-

1.5 × IQR. Outliers beyond whiskers were excluded from statistical analyses. Colored and black asterisks indicate within-group (response vs. baseline, paired Student's *t* test) and between-genotype significance (unpaired Student's *t* test), respectively. WT data are the same as in Figure 4G.

(G–I) Architecture of the MARL framework for social simulation. (G) Agents are modeled using an actor-critic network architecture optimized via the PPO algorithm to determine their movement direction and social interaction at each timestep. The simulated arena is configured as a fully observable 4 × 4 discrete grid world. (H) Different behavioral states are defined by the relative position and velocity between the two agents. (I) Reinforcement parameters for different behavioral syllables are scaled from the mean value of *in vivo* DA response magnitudes observed across sexes and genotypes (STAR Methods).

(J and K) Behavioral percentage (J) and transition probabilities (K) of simulated agents (averaged results from three seeds). Significant sex/genotype differences in (J) (two-way ANOVA) and (K) (FDR-adjusted unpaired Student's *t* test, color-coded) are shown.

Sample sizes (*n*) are indicated in the respective panels. Data are represented as mean ± SEM. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. See Table S4 for comprehensive statistical details and effect sizes.

See also Figure S6.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

CellPress
OPEN ACCESS

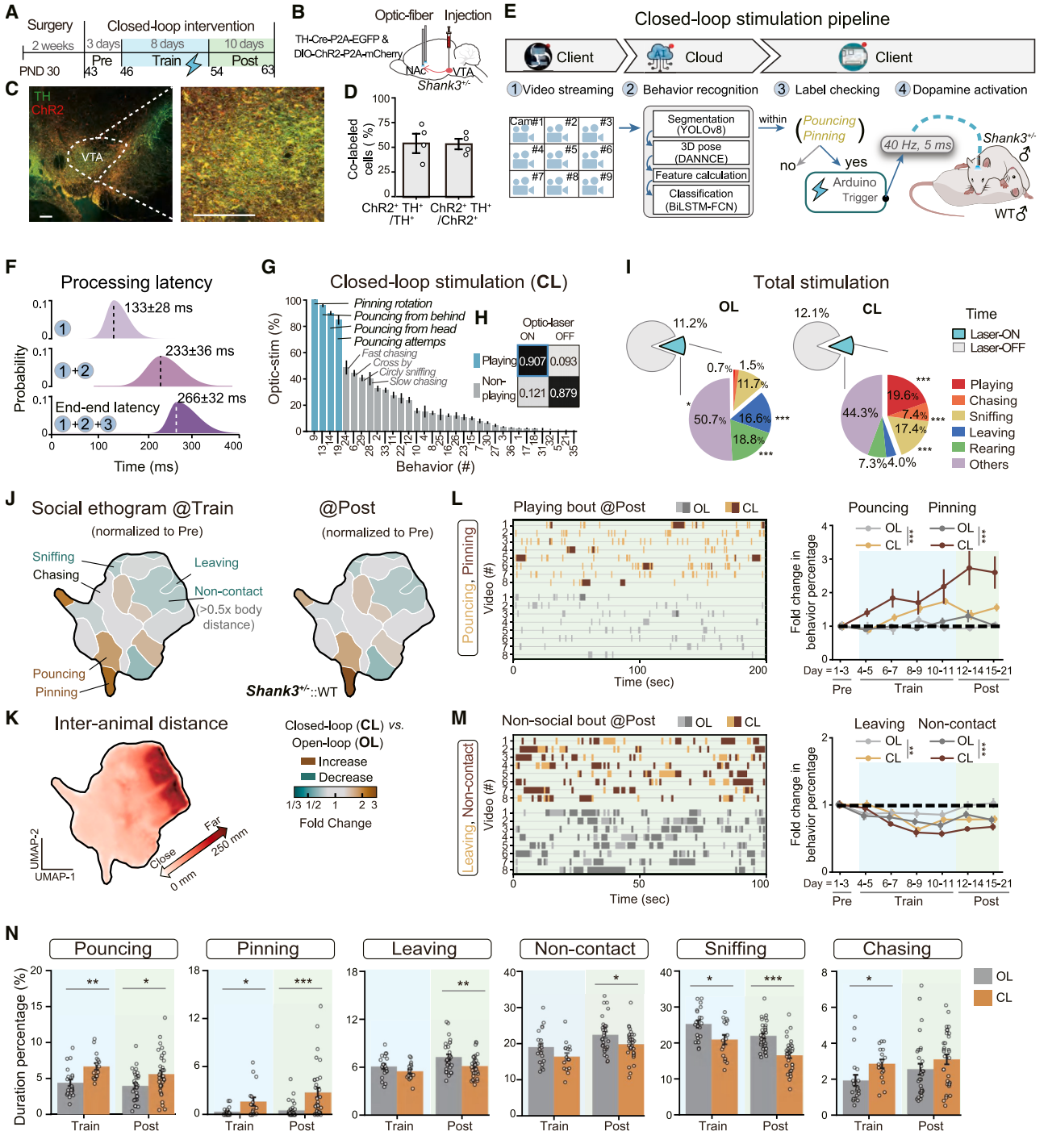


Figure 7. Closed-loop NAc-DA activation rescues social play deficits in *Shank3*$^{+/-}$ juvenile males

(A) Experimental timeline. The 8-day closed-loop (CL) optogenetic intervention and 10-day post-stimulation assessment were performed in *Shank3*-deficient males (closed-loop: *N* = 9; open-loop: *N* = 5; PND 43–63).

(B) Viral targeting and fiber optic stimulation. TH-Cre-driven channelrhodopsin-2 (ChR2)-mCherry expression in VTA dopamine neurons with optical fiber implantation in the NAc core.

(C) Representative histology. ChR2-mCherry (red) colocalization with TH$^+$ neurons (green) in the VTA. Scale bars, 200 μm.

(D) Quantification of ChR2$^+$ and TH$^+$ co-expression in VTA (*N* = 4 rats).

(legend continued on next page)

DANNCE enabled 3D pose estimation in adults,[28,31] our nine-camera system achieves the high spatiotemporal fidelity necessary for juvenile-specific, contact-heavy behaviors. The ultra-low latency (<400 ms) of our pipeline bridges the technical gap between passive observation and active intervention, permitting closed-loop manipulations during ongoing bouts previously impossible with offline methods.

Combining ethologically relevant feature extraction with active learning addressed a persistent challenge in unsupervised behavioral clustering: the trade-off between interpretability and granularity.[46] Our iterative refinement produced an ethogram that aligns with expert annotations while retaining computational objectivity. This framework establishes a standardized behavioral atlas for juvenile social interactions, allowing us to comprehensively profile naturalistic play dynamics across varied social contexts and map distinct social syllables to their underlying neural substrates.

Dopaminergic shaping of developmental social strategies

Our findings delineate a mechanistic framework for how NAc dopamine acts as a learning signal to shape naturalistic social strategies during development.[1] In WT animals, DA transients do not uniformly encode a generalized state of social motivation[10] but instead reflect action-contingent fluctuations that bias selection. For WT males, phasic DA surges during proactive play may serve as positive reinforcement. Intriguingly, state-dependent DA suppression during forced submission may function analogously to a negative teaching signal—like those observed in vocal learning models[2,3]—which drives the animal to adjust its behavioral strategy (e.g., initiating an evasion).

In *Shank3*-deficient models, this reinforcement process appears fundamentally dysregulated. The observed DA suppression during play initiation may represent an abolished teaching signal,[47] assigning a maladaptive negative value to reciprocal play. Interestingly, this mutant profile during active play closely resembles WT DA suppression during passive submission, suggesting the valuation system misprocesses proactive engagement as unrewarding. This framework reconciles increased sniffing despite blunted DA: impaired signaling fails to reinforce the transition to play, trapping the animal in an investigatory state. Intriguingly, the valuation breakdown extends to the mixed-genotype dyad. Because naturalistic play relies on reciprocal feedback,[48] the mutants' failure to initiate appropriate proactive responses deprives WT partners of social cues, rendering the interaction less rewarding for the WT animal. This disrupted reciprocity explains how an individual's reinforcement deficit degrades engagement for both partners.

Our MARL model conceptually supports this framework: when reward values for the core play syllables (pouncing and being pounced) were derived from empirical play-associated DA amplitudes, the resulting agents were biased toward the play-avoidant strategies characteristic of *Shank3* mutants (Figures 6J and 6K). Further empirical support stems from our closed-loop interventions (Figure 7). By delivering a concurrent, closed-loop DA surge precisely during the initiation of proactive play, we selectively reinforced targeted play behaviors, without driving a generalized increase in non-targeted actions such as sniffing. Notably, this robust behavioral escalation emerged across consecutive training days rather than acutely within single sessions, suggesting that these sparse, sub-second teaching signals require offline consolidation to drive durable modifications. This specificity supports action-specific reinforcement over generalized social motivation, highlighting a potential therapeutic principle: correcting micro-level reinforcement deficits may reinstate adaptive, macroscopic social strategies.

Limitations of the study

While our study positions NAc dopamine as a critical node, social behavior is orchestrated by distributed neural networks. Executing reciprocal strategies likely requires the medial prefrontal cortex for contextualizing rules[9,49,50] and the dorsal striatum for consolidating motor sequences.[5,51] Furthermore, dopaminergic signals interact with other neuromodulatory systems, notably serotonin and oxytocin, which may play complementary roles in social engagement and learning.[52–54] Future work using multiplexed biosensors will be important to disentangle these interactions.

Methodologically, analyzing unconstrained, naturalistic behavior presents inherent challenges. Dynamic internal states make it difficult to isolate strictly identical behavioral events across subjects. Future analytical frameworks incorporating latent-state

(E) Closed-loop stimulation pipeline. Multi-camera inputs were first streamed from the client to the cloud for YOLOv8-based segmentation, DANNCE-derived 3D pose estimation, feature calculation, and BiLSTM-FCN classification (covering 36 behavioral categories). Detected pouncing/pinning events triggered 40-Hz blue light pulses (5-ms pulse, 0.5-s duration) to NAc-projecting VTA terminals (<400 ms latency).
(F) Processing latencies. Frame-by-frame processing latencies for each step shown in (E) were calculated across a 1-min session (1,800 frames, mean ± SD).
(G) Percentage of closed-loop stimulation pulses allocated to each defined behavioral category. Target play-related behaviors are highlighted in blue.
(H) Contingency matrix displaying the proportion of play and non-play behaviors during laser-ON (blue) and laser-OFF (gray) conditions.
(I) Proportion of stimulation time allocated to different behavioral categories in CL and open-loop (OL) paradigms, respectively.
(J) Training (left) and post-training (right) behavioral dynamics. UMAP occupancy maps show cluster occupancy, normalized to the pre-training baseline, highlighting increases (brown) and decreases (cyan) in behaviors in CL relative to the OL group.
(K) Social proximity mapping. UMAP embedding overlaid with color-coded inter-animal distance.
(L) Play behavior analysis. Left: representative raster plots of pouncing (light color) and pinning (dark color) in OL (gray) and CL (brown) groups during the post-training period. Right: fold change in behavioral percentages across days (CL/OL: 32/30 videos; two-way ANOVA).
(M) Non-social behavior comparison. Raster plots and quantification of leaving the partner (light color) and non-contact (>0.5 body length, dark color) clusters, using the same conventions as in (L).
(N) Averaged behavioral percentages during training (day 5–8) and post-training. Training/post training, CL: 16/32 videos; OL: 20/30 videos. Two-way ANOVA. Data are represented as mean ± SEM. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. See Table S4 for comprehensive statistical details and effect sizes.
See also Figure S7.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

modeling[55,56] will be highly valuable for tracking hidden internal dynamics, providing a more precise mapping of how behavioral history influences moment-to-moment social valuation.

Finally, while the *Shank3*-deficient rat is a highly construct-valid model for Phelan-McDermid syndrome,[57] human clinical autism is a remarkably heterogeneous spectrum. It remains to be determined whether this play-related deficit in DA dynamics represents a convergent pathophysiological hub across other genetic or environmental models of autism[58–61] or a *Shank3*-specific signature. Addressing these limitations will be a crucial next step before exploring whether such behaviorally contingent neuromodulation concepts could eventually inform future, personalized intervention frameworks for neurodevelopmental disorders.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Ying Li (liying@cibr.ac.cn).

Materials availability

This study did not generate any new, unique reagents.

Data and code availability

- Experimental data and analytical results generated in this study are available on Mendeley Data: https://data.mendeley.com/datasets/8w3b9xybzg.
- All custom code used in this study is available on GitHub: https://lilab-cibr.github.io/Social_Seq/en/.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon reasonable request.

DECLARATION OF GENERATION AI AND AI-ASSISTED TECHNOLOGIES IN THE MANUSCRIPT PREPARATION PROCESS

During the preparation of this work, the authors used DeepSeek and ChatGPT solely to refine the language of the manuscript. The authors reviewed and edited the output as needed and take full responsibility for the content of the published article.

ACKNOWLEDGMENTS

We thank Q. Yang, X. Mi, CIBR LARC staff, and the CIBR Imaging, Instrumentation, and Vector Cores for technical support; Y.L. Li for the dopamine sensor; and M.M. Luo, X. Yu, M. Jiao, and the Li lab for insightful comments. Supported by the National Natural Science Foundation of China (32330045), the Brain Science and Brain-Like Intelligence Technology Major Project (2025ZD0217400 and 2021ZD0203903), the CAMS Innovation Fund (2024-I2M-ZD-013), the Beijing Natural Science Foundation (JQ24036), the HFSP Career Development Award (CDA00005/2019), the Beijing Key Laboratory of Brain Science, and the China Postdoctoral Science Foundation (2021M700454 to X.T.).

AUTHOR CONTRIBUTIONS

Conceptualization: Ying Li, X.C., and X.T.; methodology (Social-seq): Ying Li, X.C., X.T., Z.Z., M.W., and Y. Zhang; methodology (MARL model): F.Z., J.Z., and N.J.; methodology (ultrasonic vocalization analysis): Y.W. and W.X.; formal analysis: X.C., X.T., Y.O., and Z.D.; experiments: Z.Z., X.C., X.T., Y. Zhang, M.W., L.S., Yixuan Li, R.Z., Y. Zhao, and M.A.; writing—original draft: Ying Li, X.C., and X.T.; writing—review & editing: Ying Li, X.C., X.T., and N.J.; visualization: X.C., X.T., and Z.Z.; supervision: Ying Li; funding acquisition: Ying Li and X.T.

DECLARATION OF INTERESTS

Ying Li, X.T., and X.C. declare a patent application associated with this study (CNIPA application no. 202610216317.1, "The method and system for recognizing animal social behavior coupled with optogenetic stimulation").

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Rats
- METHOD DETAILS
  - *Shank3*$^{+/-}$ genotyping
  - Fur dyeing
  - Surgery for fiber photometry
  - Surgery for optogenetics
  - Histology
  - Social behavioral recordings
  - Behavioral setup
  - Social behavior clustering
  - Manual annotation of clusters
  - Closed-loop, behavior-triggered optogenetic stimulation
  - Fiber photometry for GRAB-DA recordings
  - Temporal auto-correlation analysis of behavioral features
  - Similarity analysis of behavioral clusters
  - Active-learning based refinement of behavioral clusters
  - Behavioral Metrics Analysis
  - Analyzing behavior-related DA fluctuations
  - Multi-agent reinforcement learning model
  - Software packages
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2026.10.040.

Received: August 11, 2025
Revised: August 7, 2026
Accepted: October 23, 2026

REFERENCES

1. Knudsen, E.I. (2004). Sensitive periods in the development of the brain and behavior. J. Cogn. Neurosci. *16*, 1412–1425. https://doi.org/10.1162/0898929042304796.
2. Gadagkar, V., Puzerey, P.A., Chen, R., Baird-Daniel, E., Farhang, A.R., and Goldberg, J.H. (2016). Dopamine neurons encode performance error in singing birds. Science *354*, 1278–1282. https://doi.org/10.1126/science.aah6837.
3. Kasdin, J., Duffy, A., Nadler, N., Raha, A., Fairhall, A.L., Stachenfeld, K.L., and Gadagkar, V. (2025). Natural behaviour is learned through dopamine-mediated reinforcement. Nature *641*, 699–706. https://doi.org/10.1038/s41586-025-08729-1.
4. Jin, X., and Costa, R.M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature *466*, 457–462. https://doi.org/10.1038/nature09263.
5. Markowitz, J.E., Gillis, W.F., Jay, M., Wood, J., Harris, R.W., Cieszkowski, R., Scott, R., Brann, D., Koveal, D., Kula, T., et al. (2023). Spontaneous

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

behaviour is structured by reinforcement without explicit reward. Nature *614*, 108–117. https://doi.org/10.1038/s41586-022-05611-2.

6. Pellis, S.M., Pellis, V.C., and Bell, H.C. (2010). The function of play in the development of the social brain. Am. J. Play *2*, 278–296.
7. Foulkes, L., and Blakemore, S.-J. (2016). Is there heightened sensitivity to social reward in adolescence? Curr. Opin. Neurobiol. *40*, 81–85. https://doi.org/10.1016/j.conb.2016.06.016.
8. Nardou, R., Lewis, E.M., Rothhaas, R., Xu, R., Yang, A., Boyden, E., and Dölen, G. (2019). Oxytocin-dependent reopening of a social reward learning critical period with MDMA. Nature *569*, 116–120. https://doi.org/10.1038/s41586-019-1075-9.
9. Gunaydin, L.A., Grosenick, L., Finkelstein, J.C., Kauvar, I.V., Fenno, L.E., Adhikari, A., Lammel, S., Mirzabekov, J.J., Airan, R.D., Zalocusky, K.A., et al. (2014). Natural neural projection dynamics underlying social behavior. Cell *157*, 1535–1551. https://doi.org/10.1016/j.cell.2014.05.017.
10. Manduca, A., Servadio, M., Damsteegt, R., Campolongo, P., Vanderschuren, L.J., and Trezza, V. (2016). Dopaminergic neurotransmission in the nucleus accumbens modulates social play behavior in rats. Neuropsychopharmacology *41*, 2215–2223. https://doi.org/10.1038/npp.2016.22.
11. Dai, B., Sun, F., Tong, X., Ding, Y., Kuang, A., Osakada, T., Li, Y., and Lin, D. (2022). Responses and functions of dopamine in nucleus accumbens core during social behaviors. Cell Rep. *40*, 111246. https://doi.org/10.1016/j.celrep.2022.111246.
12. Willmore, L., Cameron, C., Yang, J., Witten, I.B., and Falkner, A.L. (2022). Behavioural and dopaminergic signatures of resilience. Nature *611*, 124–132. https://doi.org/10.1038/s41586-022-05328-2.
13. Pierce, A.F., Protter, D.S.W., Watanabe, Y.L., Chapel, G.D., Cameron, R.T., and Donaldson, Z.R. (2024). Nucleus accumbens dopamine release reflects the selective nature of pair bonds. Curr. Biol. *34*, 519–530.e5. https://doi.org/10.1016/j.cub.2023.12.041.
14. Kopec, A.M., Smith, C.J., Ayre, N.R., Sweat, S.C., and Bilbo, S.D. (2018). Microglial dopamine receptor elimination defines sex-specific nucleus accumbens development and social behavior in adolescent rats. Nat. Commun. *9*, 3769. https://doi.org/10.1038/s41467-018-06118-z.
15. Wei, A., Zhao, A., Zheng, C., Dong, N., Cheng, X., Duan, X., Zhong, S., Liu, X., Jian, J., Qin, Y., et al. (2025). Sexually dimorphic dopaminergic circuits determine sex preference. Science *387*, eadq7001. https://doi.org/10.1126/science.adq7001.
16. Chevallier, C., Kohls, G., Troiani, V., Brodkin, E.S., and Schultz, R.T. (2012). The social motivation theory of autism. Trends Cogn. Sci. *16*, 231–239. https://doi.org/10.1016/j.tics.2012.02.007.
17. Jones, W., and Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. Nature *504*, 427–431. https://doi.org/10.1038/nature12715.
18. Bariselli, S., Tzanoulinou, S., Glangetas, C., Prévost-Solié, C., Pucci, L., Viguié, J., Bezzi, P., O'Connor, E.C., Georges, F., Lüscher, C., et al. (2016). SHANK3 controls maturation of social reward circuits in the VTA. Nat. Neurosci. *19*, 926–934. https://doi.org/10.1038/nn.4319.
19. Liu, J., Ye, J., Ji, C., Ren, W., He, Y., Xu, F., and Wang, F. (2024). Mapping the Behavioral Signatures of Shank3b Mice in Both Sexes. Neurosci. Bull. *40*, 1299–1314. https://doi.org/10.1007/s12264-024-01237-8.
20. Barbier, M., Rajamani, K.T., Netser, S., Wagner, S., and Harony-Nicolas, H. (2025). Altered neural activity in the Mesoaccumbens pathway underlies impaired social reward processing in Shank3-deficient rats. Adv. Sci. (Weinh) *12*, e2414813. https://doi.org/10.1002/advs.202414813.
21. Folkes, O.M., Donahue, M., Wang, S.E., Negrón-Moreno, P.N., Oo, N.X., Qiao, S.N., Wang, X., and Jiang, Y.H. (2025). Deficiency of Shank3 in the nucleus accumbens reveals a loss of social-specific motivation. Commun. Biol. *8*, 1569. https://doi.org/10.1038/s42003-025-08942-8.
22. Meng, Y., Xu, D., Zhang, W., Meng, W., Lan, X., Wang, X., Li, M., Zhang, X., Zhao, Y., Yang, H., et al. (2022). Effect of early swimming on the behavior and striatal transcriptome of the Shank3 knockout rat model of autism. Neuropsychiatr. Dis. Treat. *18*, 681–694. https://doi.org/10.2147/NDT.S357338.
23. Marshall, J., Klibaite, U., Gellis, A., Aldarondo, D., Olveczky, B., and Dunn, T.W. (2021). The PAIR-R24M dataset for multi-animal 3D pose estimation. In Advances in Neural Information Processing Systems. NeurIPS (Curran Associates, Inc). 34. https://openreview.net/forum?id=-wVVl_UPr8.
24. Lauer, J., Zhou, M., Ye, S., Menegas, W., Schneider, S., Nath, T., Rahman, M.M., Di Santo, V., Soberanes, D., Feng, G., et al. (2022). Multi-animal pose estimation, identification and tracking with DeepLabCut. Nat. Methods *19*, 496–504. https://doi.org/10.1038/s41592-022-01443-0.
25. Marks, M., Qiuhan, J., Sturman, O., von Ziegler, L., Kollmorgen, S., von der Behrens, W., Mante, V., Bohacek, J., and Yanik, M.F. (2022). Deep-learning based identification, tracking, pose estimation, and behavior classification of interacting primates and mice in complex environments. Nat. Mach. Intell. *4*, 331–340. https://doi.org/10.1038/s42256-022-00477-5.
26. Pereira, T.D., Tabris, N., Matsliah, A., Turner, D.M., Li, J., Ravindranath, S., Papadoyannis, E.S., Normand, E., Deutsch, D.S., Wang, Z.Y., et al. (2022). SLEAP: A deep learning system for multi-animal pose tracking. Nat. Methods *19*, 486–495. https://doi.org/10.1038/s41592-022-01426-1.
27. Han, Y., Chen, K., Wang, Y., Liu, W., Wang, Z., Wang, X., Han, C., Liao, J., Huang, K., Cai, S., et al. (2024). Multi-animal 3D social pose estimation, identification and behaviour embedding with a few-shot learning framework. Nat. Mach. Intell. *6*, 48–61. https://doi.org/10.1038/s42256-023-00776-5.
28. Klibaite, U., Li, T., Aldarondo, D., Akoad, J.F., Ölveczky, B.P., and Dunn, T.W. (2025). Mapping the landscape of social behavior. Cell *188*, 2249–2266.e23. https://doi.org/10.1016/j.cell.2025.01.044.
29. He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask R-CNN. In IEEE International Conference on Computer Vision, pp. 2980–2988.
30. Jocher, G., Chaurasia, A., and Qiu, J. (2023). Ultralytics YOLOv8. GitHub. https://github.com/ultralytics/ultralytics.
31. Dunn, T.W., Marshall, J.D., Severson, K.S., Aldarondo, D.E., Hildebrand, D.G.C., Chettih, S.N., Wang, W.L., Gellis, A.J., Carlson, D.E., Aronov, D., et al. (2021). Geometric deep learning enables 3D kinematic profiling across species and environments. Nat. Methods *18*, 564–573. https://doi.org/10.1038/s41592-021-01106-6.
32. Zeng, A., Yang, L., Ju, X., Li, J., Wang, J., and Xu, Q. (2022). SmoothNet: A plug-and-play network for refining human poses in videos. In Computer Vision—ECCV 2022, S. Avidan, G. Brostow, M. Cissé, G.M. Farinella, and T. Hassner, eds. (Springer Nature Switzerland), pp. 625–642. https://doi.org/10.1007/978-3-031-20065-6_36.
33. Wiltschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abraira, V.E., Adams, R.P., and Datta, S.R. (2015). Mapping sub-second structure in mouse behavior. Neuron *88*, 1121–1135. https://doi.org/10.1016/j.neuron.2015.11.031.
34. Su, K., and Shlizerman, E. (2020). Clustering and recognition of spatiotemporal features through interpretable embedding of sequence to sequence recurrent neural networks. Front. Artif. Intell. *3*, 70. https://doi.org/10.3389/frai.2020.00070.
35. Pellis, S.M., Pellis, V.C., Burke, C.J., Stark, R.A., Ham, J.R., Euston, D.R., and Achterberg, E.J.M. (2022). Measuring play fighting in rats: A multilayered approach. Curr. Protoc. *2*, e337. https://doi.org/10.1002/cpz1.337.
36. Zhuo, Y., Luo, B., Yi, X., Dong, H., Miao, X., Wan, J., Williams, J.T., Campbell, M.G., Cai, R., Qian, T., et al. (2024). Improved green and red GRAB sensors for monitoring dopaminergic activity in vivo. Nat. Methods *21*, 680–691. https://doi.org/10.1038/s41592-023-02100-w.
37. Littman, M.L. (1994). Markov Games as a Framework for Multi-agent Reinforcement Learning. In Machine Learning Proceedings 1994 (Morgan Kaufmann), pp. 157–163. https://doi.org/10.1016/B978-1-55860-335-6.50027-1.
38. Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, O.P., and Mordatch, I. (2017). Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. The 31st International Conference on NIPS, pp. 6382–6393.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

39. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal Policy Optimization Algorithms. Preprint at arXiv. https://doi.org/10.48550/arXiv.1707.06347.

40. Barto, A.G., Sutton, R.S., and Anderson, C.W. (2012). Neuronlike adaptive elements that can solve difficult learning control problems. IEEE Trans. Syst. Man Cybern. *5*, 834–846. https://doi.org/10.7551/mitpress/4943.001.0001.

41. Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In Proceedings of Machine Learning Research, pp. 1928–1937.

42. Berridge, K.C., and Robinson, T.E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res., Brain Res. Rev. *28*, 309–369. https://doi.org/10.1016/s0165-0173(98)00019-8.

43. Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. Physiol. Rev. *95*, 853–951. https://doi.org/10.1152/physrev.00023.2014.

44. Kim, Y.S., Fombonne, E., Koh, Y.J., Kim, S.J., Cheon, K.A., and Leventhal, B.L. (2014). A comparison of DSM-IV pervasive developmental disorder and DSM-5 autism spectrum disorder prevalence in an epidemiologic sample. J. Am. Acad. Child Adolesc. Psychiatry *53*, 500–508. https://doi.org/10.1016/j.jaac.2013.12.021.

45. Supekar, K., Kochalka, J., Schaer, M., Wakeman, H., Qin, S., Padmanabhan, A., and Menon, V. (2018). Deficits in mesolimbic reward pathway underlie social interaction impairments in children with autism. Brain *141*, 2795–2805. https://doi.org/10.1093/brain/awy191.

46. Jiao, L., Yang, H., Liu, Z., and Pan, Q. (2022). Interpretable fuzzy clustering using unsupervised fuzzy decision trees. Inf. Sci. *611*, 540–563. https://doi.org/10.1016/j.ins.2022.08.077.

47. Watabe-Uchida, M., Eshel, N., and Uchida, N. (2017). Neural Circuitry of Reward Prediction Error. Annu. Rev. Neurosci. *40*, 373–394. https://doi.org/10.1146/annurev-neuro-072116-031109.

48. Vanderschuren, L.J.M.J., Achterberg, E.J.M., and Trezza, V. (2016). The neurobiology of social play and its rewarding value in rats. Neurosci. Biobehav. Rev. *70*, 86–105. https://doi.org/10.1016/j.neubiorev.2016.07.025.

49. Bicks, L.K., Koike, H., Akbarian, S., and Morishita, H. (2015). Prefrontal cortex and social cognition in mouse and man. Front. Psychol. *6*, 1805. https://doi.org/10.3389/fpsyg.2015.01805.

50. Kingsbury, L., Huang, S., Wang, J., Gu, K., Golshani, P., Wu, Y.E., and Hong, W. (2019). Correlated neural activity and encoding of behavior across brains of socially interacting animals. Cell *178*, 429–446.e16. https://doi.org/10.1016/j.cell.2019.05.022.

51. Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. *7*, 464–476. https://doi.org/10.1038/nrn1919.

52. Dölen, G., Darvishzadeh, A., Huang, K.W., and Malenka, R.C. (2013). Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. Nature *501*, 179–184. https://doi.org/10.1038/nature12518.

53. Luo, J., Feng, Q., Wei, L., and Luo, M. (2017). Optogenetic activation of dorsal raphe neurons rescues the autistic-like social deficits in Shank3 knockout mice. Cell Res. *27*, 950–953. https://doi.org/10.1038/cr.2017.52.

54. Walsh, J.J., Christoffel, D.J., Heifets, B.D., Ben-Dor, G.A., Selimbeyoglu, A., Hung, L.W., Deisseroth, K., and Malenka, R.C. (2018). 5-HT release in nucleus accumbens rescues social deficits in mouse autism model. Nature *560*, 589–594. https://doi.org/10.1038/s41586-018-0416-4.

55. Padilla-Coreano, N., Batra, K., Patarino, M., Chen, Z., Rock, R.R., Zhang, R., Hausmann, S.B., Weddington, J.C., Patel, R., Zhang, Y.E., et al. (2022). Cortical ensembles orchestrate social competition through hypothalamic outputs. Nature *603*, 667–671. https://doi.org/10.1038/s41586-022-04507-5.

56. Weinreb, C., Kannan, L.T., Newman-Boulle, A., Sainburg, T., Gillis, W.F., Plotnikoff, A., Makowska, S., Pearl, J.E., Osman, M.A.M., Linderman, S.W., et al. (2026). Spontaneous behavior is a succession of self-directed tasks. Neuron *114*, 922–937.e12. https://doi.org/10.1016/j.neuron.2025.11.021.

57. Peça, J., Feliciano, C., Ting, J.T., Wang, W., Wells, M.F., Venkatraman, T.N., Lascola, C.D., Fu, Z., and Feng, G. (2011). Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. Nature *472*, 437–442. https://doi.org/10.1038/nature09965.

58. Dölen, G., Osterweil, E., Rao, B.S.S., Smith, G.B., Auerbach, B.D., Chattarji, S., and Bear, M.F. (2007). Correction of fragile X syndrome in mice. Neuron *56*, 955–962. https://doi.org/10.1016/j.neuron.2007.12.001.

59. Peñagarikano, O., Abrahams, B.S., Herman, E.I., Winden, K.D., Gdalyahu, A., Dong, H., Sonnenblick, L.I., Gruver, R., Almajano, J., Bragin, A., et al. (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell *147*, 235–246. https://doi.org/10.1016/j.cell.2011.08.040.

60. Choi, G.B., Yim, Y.S., Wong, H., Kim, S., Kim, H., Kim, S.V., Hoeffer, C.A., Littman, D.R., and Huh, J.R. (2016). The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. Science *351*, 933–939. https://doi.org/10.1126/science.aad0314.

61. Estes, M.L., and McAllister, A.K. (2016). Maternal immune activation: Implications for neuropsychiatric disorders. Science *353*, 772–777. https://doi.org/10.1126/science.aag3194.

62. Kim, C.K., Yang, S.J., Pichamoorthy, N., Young, N.P., Kauvar, I., Jennings, J.H., Lerner, T.N., Berndt, A., Lee, S.Y., Ramakrishnan, C., et al. (2016). Simultaneous fast measurement of circuit dynamics at multiple sites across the mammalian brain. Nat. Methods *13*, 325–328. https://doi.org/10.1038/nmeth.3770.

63. Yuan, L., Dou, Y.N., and Sun, Y.G. (2019). Topography of reward and aversion encoding in the mesolimbic dopaminergic system. J. Neurosci. *39*, 6472–6481. https://doi.org/10.1523/JNEUROSCI.0271-19.2019.

64. de Chaumont, F., Ey, E., Torquet, N., Lagache, T., Dallongeville, S., Imbert, A., Legou, T., Le Sourd, A.-M., Faure, P., Bourgeron, T., et al. (2019). Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning. Nat. Biomed. Eng. *3*, 930–942. https://doi.org/10.1038/s41551-019-0396-1.

65. Chen, X., and Li, H. (2017). ArControl: An Arduino-Based Comprehensive Behavioral Platform with Real-Time Performance. Front. Behav. Neurosci. *11*, 244. https://doi.org/10.3389/fnbeh.2017.00244.

66. Simpson, E.H., Akam, T., Patriarchi, T., Blanco-Pozo, M., Burgeno, L.M., Mohebi, A., Cragg, S.J., and Walton, M.E. (2024). Lights, fiber, action! A primer on in vivo fiber photometry. Neuron *112*, 718–739. https://doi.org/10.1016/j.neuron.2023.11.016.

67. Li, Y., Mathis, A., Grewe, B.F., Osterhout, J.A., Ahanonu, B., Schnitzer, M.J., Murthy, V.N., and Dulac, C. (2017). Neuronal representation of social information in the medial amygdala of awake behaving mice. Cell *171*, 1176–1190.e17. https://doi.org/10.1016/j.cell.2017.10.015.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit Anti-Tyrosine Hydroxylase Antibody (1:500)	Sigma-Aldrich	Cat# AB152, RRID:AB_390204
Guinea Pig **mCherry/tdTom pAb** (1:500)	Oasis Biofarm	Cat# OB-PGP004, RRID:AB_2934239
Goat Anti-Rabbit IgG H&L (Alexa Fluor® 647) (1:500)	Abcam	Cat# AB150079, RRID:AB_2722623
Cy3 AffiniPure® Donkey Anti-Guinea Pig IgG (H+L) (1:500)	Jackson ImmunoResearch Labs	Cat# 706-165-148, RRID:AB_2340460
Bacterial and virus strains		
AAV2/9-rTH_PI-Cre-pA	Shanghai Taitoo Bioscience	Cat# S0584-9
AAV2/9-hEf1a-DIO-hChR2(H134R)-mcherry-WPRE-PA	Shanghai Taitoo Bioscience	Cat# S0170-9
pAAV-hSyn-DA3m-WPRE-hGH-polyA	BrainVTA	Cat# PT-4720
Chemicals, peptides, and recombinant proteins		
Mounting Medium antifading (with DAPI)	Beijing Solarbio Science & Technology	S2110
Deposited data		
Data to reproduce figures in paper	Figshare	Mendeley Data: https://data.mendeley.com/datasets/8w3b9xybzg
Experimental models: Organisms/strains		
Rat: Sprague-Dawley wildtype	Charles River Laboratories	Strain code: 101
Rat: Sprague-Dawley *Shank3*$^{+/-}$	Meng et al.[22]	N/A
Software and algorithms		
Code to reproduce figures in paper	Zenodo	https://doi.org/10.5281/zenodo.23492583
Mask R-CNN	He et al.[29]	N/A
YOLOv8	Jocher et al.[30]	https://github.com/ultralytics/ultralytics
DANNCE	Dunn et al.[31]	N/A
SmoothNet	Zeng et al.[32]	N/A
LabelMe	Github	https://github.com/wkentaro/labelme
Label3D	Github	https://github.com/diegoaldarondo/Label3D
Python (3.8)	Python	https://www.python.org/
OBS Studio (29.0) for video recording	OBS Studio Contributors	https://obsproject.com/
Social-seq pipeline	This paper	https://chenxinfeng4.github.io/social-seq-live-docs
Multi-camera calibration pipeline	This paper	https://github.com/chenxinfeng4/multiview_ball_calib
ffmpegcv for fast video/stream processing	This paper	https://pypi.org/project/ffmpegcv
ImageJ	NIH	https://imagej.net/ij/
Other		
Fibre photometry system	Kim et al.[62]	N/A
Behavior camera	Kingcent Techology	Model No.: K1A293-OV9281

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Rats

All experimental procedures were performed in accordance with protocols approved by the Animal Care and Use Committee of the Chinese Institute for Brain Research, Beijing (CIBR). The study employed both wild-type Sprague-Dawley (SD) rats and

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

CellPress
OPEN ACCESS

Shank3$^{+/-}$ mutants (exons 11-21 deletion) aged 5-12 weeks (juvenile stage).[22] All animals were group-housed (3-5 per cage) in standard ventilated cages containing bedding material and environmental enrichment (nesting material, PVC tubes), and maintained under a 12-h light/dark cycle (22°C, 30-70% humidity) with ad libitum access to food and water.

METHOD DETAILS

Shank3$^{+/-}$ genotyping

Regular genotyping was performed using tail biopsy samples and PCR amplification with the following primers: Forward 5′-CTGTTGGCTGAGCCTGGCATAGAG-3′ and Reverse 5′-GCTGGAAAGAAACAACGAGAGCCAG-3′ for the WT allele (559 base pairs) and Forward 5′-TTGTGCACTGCCTATGTTGACCACT-3′ and Reverse 5′- TAGGCGAGAGAAGATGGTGTGATTTCC-3′ for the mutant allele (688 base pairs). Behavioral experiments were conducted during the dark phase of the circadian cycle to accommodate the rats' natural activity patterns. Both facility-bred animals and wild-type rats purchased from Taconic Biosciences were allowed at least one week of acclimation prior to experimental procedures.

Fur dyeing

At 5-6 weeks of age, rats were anesthetized with isoflurane (3-5% induction, 1-3% maintenance) and their fur was marked using non-toxic blue pet dye (Wuhu Tongming Trading Co., Ltd.). For two-rat social behavior analysis, animals received distinct dorsal stripe patterns, while additional dot markings were applied for multi-animal 3D tracking experiments. All dyed animals were monitored for 24 hours post-procedure to ensure complete recovery from anesthesia and absence of skin irritation. Crucially, as detailed in our results and supplementary validations (Figure S2A), this non-toxic fur marking procedure did not significantly alter any quantitative behavioral metrics (e.g., total interaction time, movement speed, or play bout frequency) compared to unmarked controls, confirming the non-invasive nature of our tracking strategy.

Surgery for fiber photometry

Postnatal day (PND) 25–27 SD rats of both sexes (male/female) and genotypes (WT/*Shank3*$^{+/-}$) were anaesthetized using 1.5-3.0% isoflurane in oxygen. pAAV-hSyn-DA3m-WPRE-hGH-polyA (BrainVTA Co., Ltd, titer: 2.0×1012 μg/ml) was injected into the nucleus accumbens (NAc, AP 1.25 mm; ML 1.85 mm; DV 6.00 mm), in a total volume of 400 nl per injection. For all stereotaxic implants, AP and ML were zeroed relative to bregma, DV was zeroed relative to the pial surface. Injections were performed in 10 min, unilaterally in each rat. A single 200 μm diameter, 0.39 NA fiber cannula was implanted 200 μm above the injection site at the NAc for photometry data collection. Dental silicate cement was applied to secure the fiber with the skull. Dopamine (DA) signal and behavioral data were obtained at PND 45-55 days post-surgery.

Surgery for optogenetics

Shank3$^{+/-}$ SD rats (PND 25-27) underwent stereotaxic surgery. AAV2/9-hEf1a-DIO-hChR2(H134R)-mcherry-WPRE-PA (Shanghai Taitool Bioscience Co. Ltd., titer: 1.1×10^{13} μg/ml) and AAV2/9-rTH_PI-Cre-pA (Shanghai Taitoo Bioscience, titer: 3.0×10^{12} μg/ml) were bilaterally injected into the ventral tegmental area (VTA, AP 5.50 mm; ML ± 1.00 mm; DV 7.85 mm), in a total volume of 400 nl per injection. Two 200 μm optical fibers were implanted bilaterally above NAc. Optical stimulation (470 nm, 40 Hz, 5 ms pulse width, 4 mW per site) was administered during behavioral training on PND 45-52.

Histology

For immunohistochemistry, rats were anesthetized with sodium pentobarbital and transcardially perfused with saline, followed by 4% paraformaldehyde (PFA). Brains were postfixed for 12h at 4 °C, cryoprotected in 15% sucrose followed by 30% sucrose, and sectioned frozen on a sliding microtome at 40 μm. Sections were blocked and permeabilized in PBS with 5% donkey and goat serum containing 0.3% Triton X-100 for 1h at room temperature. The following primary antibodies were used: Anti-Tyrosine Hydroxylase Antibody and mCherry/tdTom Guinea pig pAb. The following secondary antibodies were used: donkey anti-guinea pig Cy3 and goat anti-rabbit Alexa Fluor 647. Cell nuclei were counterstained using 4',6-diamidino-2-phenylindole (DAPI). Images were acquired on a Nikon Ti2 laser scanning microscope with a Plan Apochromat VC 20X objective (N.A. = 0.75), and analysis was performed using Cell Counter of ImageJ (N.I.H, Bethesda, MD). Numbers of Channelrhodopsin-2 (ChR2)-expressing TH neurons (ChR2$^+$ TH$^+$), ChR2-expressing neurons, and TH neurons (TH$^+$) in the VTA were counted.

Social behavioral recordings

Animal subjects

To establish the social behavior dataset, we used 24 pairs of juvenile wildtype SD rats (12 male pairs and 12 female pairs; PND 50-60), housed in same-gender dyads. For identification, one randomly selected rat per cage received fur dyeing.

For autism model studies, we utilized *Shank3*$^{+/-}$ rats (derived from *Shank3*$^{+/-}$ x WT breeding pairs) alongside WT controls (from WT x WT parents), co-housed in same-gender, same-genotype pairs. Behavioral testing was conducted across three developmental

stages: early-juvenile (PND 35-45, *Shank3*$^{+/-}$::WT *vs.* WT::WT, 14 *vs.* 14 male pairs, 10 *vs.* 12 female pairs), mid-juvenile (PND 55-65, *Shank3*$^{+/-}$::WT *vs.* WT::WT, 24 *vs.* 28 male pairs, 22 *vs.* 28 female pairs), and adult (PND 75-85, *Shank3*$^{+/-}$::WT *vs.* WT::WT, 19 *vs.* 20 male pairs, 18 *vs.* 17 female pairs).

Free social behavioral paradigm

All cohorts underwent identical habituation and testing procedures. During a 3-day habituation phase, rats were placed in the cylindrical testing arena with their cage mate for 15 min/day to acclimate. On experimental days, subjects were socially isolated for 4-6 hours to enhance social motivation before testing. Free social interactions were recorded for 15 minutes using a multi-camera setup, with behavior analyzed via Social-seq pipeline for quantification and clustering.

Fiber photometry in WT and Shank3$^{+/-}$ rats

Separate cohorts of *Shank3*$^{+/-}$ (7 males and 11 females) and WT (8 males and 12 females) rats underwent aseptic stereotaxic surgery at PND 25-27 for optical fiber implantation targeting the NAc. After post-operative recovery, *in vivo* fiber photometry was performed during social interaction tests at PND 45-55, capturing DA signals synchronized with behavioral video.

Liquid consumption paradigm (Sucrose and quinine)

Following viral injection at PND 28, rats (7 males and 5 females) were allowed a two-week recovery period. Subsequently, they were handled daily for 5 min and subjected to water restriction for two days prior to behavior test. To maintain their body weight at 85%–90% of ad libitum levels (keep them motivated to get liquid during the behavior test without compromsing their health), each rat received 2.0 mL of water daily via a small dish (diameter: 5.8 cm).[63] Prior to photometry recordings, each rat was placed in a cylindrical testing arena and allowed to explore freely for 3 min after being connected to an external optical patch cord. Fiber photometry signals and multi-view behavioral videos were recorded simultaneously. Three liquid stimuli (distilled water, 1 mM quinine solution, and 30% sucrose solution) were presented in a randomized order. Each stimulus was delivered at a volume of 150 μL per trial, for a total of three trials per stimulus, with an inter-trial interval of 60 s. To precisely assess DA sensor signals in response to each stimulus, the exact onset of liquid consumption was manually annotated from the synchronized videos for rats expressing the DA sensor in the NAc. During the recordings, the laser power was maintained at approximately 15 μW at the tip of the implanted optic fiber.

Closed-loop optogenetic stimulation in male Shank3$^{+/-}$ rat

Male *Shank3*$^{+/-}$ rats were implanted with integrated optical fibers targeting the NAc at PND 25–27. After arena habituation, they underwent three consecutive phases: (1) A 3-day baseline period with daily 15-minute social interactions without stimulation; (2) An 8-day training phase where closed-loop optogenetic stimulation was delivered in real-time based on classifier-identified behavioral patterns; and (3) A 9-day post-training phase without stimulation to assess persistence of behavioral modifications.

Behavioral setup

Recording arena

All social behavior experiments were conducted in a custom-built cylindrical arena (55 cm diameter × 50 cm height) constructed from 5 mm-thick transparent acrylic. The arena floor was uniformly covered with 2 cm of autoclaved corncob bedding mixed with soiled bedding from donor cages. Illumination was provided by six 30 cm red LED tubes positioned equidistantly around the arena periphery, maintaining 20 lux at the arena center to accommodate the rats' nocturnal activity patterns while enabling high-quality video capture.

Video acquisition system

Nine synchronized high-speed cameras (Kingcent KS1A293-OV9281, 1MP grayscale) equipped with 2.6–3.2 mm lenses were arranged in a 3×3 grid pattern surrounding the arena. Each camera was connected via PCIe-to-USB 3.0 interfaces (Susuyou U3412U) to a Windows workstation (32GB RAM) equipped with NVIDIA GPUs (RTX 2060/3060 validated). The system achieved 3840×2400 pixel resolution at 30 fps using OBS Studio (v2.8) with NVIDIA NVENC HEVC encoding (P5 preset, 30 Mbps bitrate), generating ∼3.5 GB/15 min of video data. Camera alignment was verified daily using a calibration ball to ensure consistent multi-view coverage.

Camera calibration

For camera calibration, we developed a pipeline combining checkerboard-based intrinsic parameter estimation and marker-based extrinsic calibration. Intrinsic parameters (focal length, principal point, and distortion coefficients) were computed using a 12×9 black-and-white checkerboard (20 mm square size) imaged from multiple orientations, processed through OpenCV's camera calibration toolbox. Extrinsic parameters (camera positions and orientations) were determined by tracking a 25 mm metal ball moved throughout the arena. Ball centroids were detected using a You Only Look Once version 8 (YOLOv8) network trained for high-precision marker localization, followed by bundle adjustment to optimize camera poses. Absolute scale and global orientation were established by aligning the reconstruction to checkerboard-derived landmarks through rigid transformation (rotation and scaling). The final calibration output—comprising intrinsic and extrinsic parameters in standard formats—is compatible with major pose estimation frameworks (Anipose, DeepLabCut, and DANNCE). Code and validation protocols are publicly available on GitHub (https://github.com/chenxinfeng4/multiview_ball_calib).

Social Conditioned Place Preference (sCPP) task

The protocol for the social conditioned place preference (sCPP) task was adapted and modified from previously published work.[8] Upon weaning, animals were socially housed (3–4 same-sex cage mates per cage) in a home cage containing corncob bedding until the predetermined age for sCPP testing. Testing was conducted in a custom-built apparatus consisting of two joined circular

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

transparent acrylic cylinders. A 10-cm-wide passage was cut at the junction to allow rats to shuttle freely between the two compartments. The base of each cylindrical zone contained a distinct novel bedding cue: soft oak granules (Niteangel, China) or paper cotton (Kaytee Soft Granule, USA). An infrared camera coupled with Open Broadcaster Software (OBS) was utilized to continuously monitor the position of the rats.

During an initial 10-min pre-conditioning test session, the amount of time spent freely exploring each zone was recorded to establish baseline preferences for the two sets of bedding cues. To evaluate the rewarding valence of social interaction, we employed a biased assignment procedure. Based on the pre-test baselines, rats were assigned to receive social conditioning in the compartment containing its initially less-preferred bedding (i.e., baseline preference ratio < 0.5). Consequently, the pre-conditioning scores inherently reflect a natural lower preference (or mild avoidance) for the designated social-paired side.

The conditioning phase spanned 5 consecutive days. On each conditioning day, the subject rat underwent 3.5 hours of acute social isolation prior to the training session. Subsequently, the rat was placed in the assigned compartment and allowed 15 min of free social interaction with an unfamiliar conspecific. Following a 2-hour interval of isolation in a standard cage with corncob bedding, the rat was placed in the opposite compartment and underwent a 15-min isolation conditioning session. Immediately after this isolation session, the rat was returned to its home cage to reunite with its original cage mates. This procedure was repeated daily. Importantly, while the social pairing was assigned to the less-preferred side for each individual, the physical chamber sides (left vs. right) and specific bedding cues were counterbalanced across the entire experimental cohort to ensure that the overall preference shift was not driven by innate environmental biases.

Following the 5-day conditioning phase, a 10-min post-conditioning test was conducted (identical to the pre-conditioning test, without the social partner) to evaluate the shift in preference for the conditioned cues. To quantify this shift, a preference score (sCPP score) was calculated as the ratio of time spent in the social-conditioned compartment to the total time spent exploring both compartments.

Social behavior clustering

Mask segmentation

To establish robust animal segmentation, we manually annotated 4,684 frames covering diverse behavioral scenarios using the Labelme annotation tool (https://github.com/labelmeai/labelme). This comprehensive dataset captured various interaction states ranging from isolated individuals to complex social contact situations, ensuring the models could handle real-world variability.

For offline high-precision analysis, we implemented a Mask Region-based Convolutional Neural Network (Mask R-CNN) model through the MMDetection framework (https://github.com/open-mmlab/mmdetection), utilizing a ResNet-101 backbone trained over 12 epochs on images resized to 832×512 resolution. The model incorporated strategic 21×21 kernel dilation of binary masks to enhance contour visibility for subsequent pose estimation.

Parallel to this, we developed a real-time variant using YOLOv8-seg (https://github.com/ultralytics/ultralytics), optimized for 640×480 pixel input with 7×7 mask dilation, achieving sub-20ms processing times per frame on our GPU configuration while maintaining adequate precision for behavioral tracking.

3D pose estimation

The pose estimation pipeline began with meticulous 3D annotation of 3,638 pose instances through Label3D (https://github.com/diegoaldarondo/Label3D), defining a 14-point skeletal model that captured essential body landmarks including: head (Snout, Ear-L, Ear-R), trunk (Neck, Back, Tail-Base), limbs (Elbow-L, Knee-L, Elbow-R, Knee-R), and claws (Wrist-L, Wrist-R, Foot-L, Foot-R). The 3-Dimensional Aligned Neural Network for Computational Ethology (DANNCE) architecture formed the core of our 3D reconstruction system, processing $64\times64\times64$ volumetric grids centered on triangulated mask centroids. Through 300 epochs of training with comprehensive data augmentation, the network learned to predict precise 3D probability distributions for each keypoint, with final positions determined by local maxima extraction. The system demonstrated strong reconstruction consistency across viewpoints, while handling the challenges of occluded limbs and dynamic social interactions.

Jitter refinement

We implemented a multi-stage jitter correction pipeline to ensure kinematic plausibility of 3D pose trajectories. The process began with raw 3D keypoint positions estimated by DANNCE, which were visualized alongside SmoothNet-derived fits (3-layer FC network, 64-frame window) across all spatial dimensions (X/Y/Z axes) to assess movement dynamics. Significant jitter events were systematically identified through exceeding specific standard deviations (2.5 SD for head, truck and limbs; 3.5 SD for claws) of the baseline movement variability. The correction algorithm first computed residual displacements between raw and smoothed trajectories for each spatial dimension, then applied threshold-based detection using magnitude when the residual exceeded the adaptive threshold. The final output preserved natural movement dynamics while eliminating artifactual discontinuities, as verified through manual inspection of 3,690 seconds of annotated samples across all behavioral conditions.

Behavioral feature processing

We designed 32 features to quantify posture, locomotion, and social interactions between pairs of rats. To strictly control body-size-related kinematics during development, we implemented scale-invariant normalization throughout our tracking pipeline. Specifically, velocity was normalized by the individual's body length, absolute interaction distance was normalized by the average body length, and body occlusion (overlap area) was normalized by the square of the average body length. Consequently, all derived kinematic

metrics—including velocity, interaction proximity, and posture—are inherently scale-invariant, ensuring that downstream behavioral classifications and analyses are not trivial byproducts of animal size. The features were grouped into four categories:

(1) Individual kinematic features (n = 10)

For each rat, we computed kinematic measures including back velocity (XY-plane), back height (z-axis), limb-to-paw height differences, and the angles of the upper and lower trunk planes relative to the z-axis. The upper trunk plane was defined by the forelimbs and back, while the lower trunk plane was defined by the hindlimbs and back

(2) Social Interaction Features (n = 10)

These features captured spatial relationships between the two rats, including distances between their backs, noses, and nose-to-tail and vice versa. We also calculated their relative orientations, such as the heading direction of rat A with respect to rat B-nose or -back, and vice versa. Additionally, we measured height asymmetries, including differences in averaged paw height (rat A-paws vs. rat B-paws), back height (rat A-back vs. rat B-back), and cross-body comparisons (rat A-back vs. rat B-paws and rat A-paws vs. rat B-back).

(3) Sniffing Features (n = 3)

To quantify sniffing behavior, we measured XY-plane overlaps between the noses and tails of the two rats, specifically rat A-nose to rat B-nose, rat A-nose to rat B-tail, and rat A-tail to rat B-nose. Overlap areas were determined using circular zones sized by the average distances between keypoints (nose-ear and tail-hindlimb)

(4) Body Occlusions (n = 9)

We quantified occlusions by measuring XY-plane overlaps between three body regions: the head zone (nose to forelimbs) and the upper and lower trunk zones (quadrilaterals spanning forelimbs to hindlimbs). The direction of occlusion (positive or negative) was determined by the relative height of rat A compared to rat B.

Self-supervised behavioral embedding and clustering

We developed a computational pipeline to systematically characterize the diversity of social interactions by analyzing behavioral segments where rats were within one body length of each other. These interaction periods were divided into 0.8-sec clips based on the temporal autocorrelation properties of behavioral features, providing optimal time windows for capturing discrete social events.

For feature processing, we implemented a Seq2seq autoencoder architecture to perform nonlinear dimensionality reduction on the behavioral sequences. Each 0.8-sec clip was represented as a 32×24 feature matrix capturing 32 behavioral metrics across 24 time points. The model was trained using 90% of the dataset (60,000 clips) with the remaining 6,666 clips reserved for testing, employing a learning rate of 0.01 and batch size of 512 until either the test set loss decreased to 0.16 or 400 epochs were completed. The 128-dimensional latent representations generated by the decoder were subsequently reduced to 12 PCs through Principal Component Analysis (PCA), effectively preserving over 90% of the variance while eliminating redundant dimensions.

The clustering analysis proceeded through a hierarchical approach beginning with an initial K-means clustering (K=20) of the full dataset, which identified the smallest cluster containing 23280 clips for balanced subsampling. We then performed a second K-means clustering (K=100) on this subsampled data to determine the optimal number of behavioral clusters, ultimately identifying 42 robust clusters that each contained more than 1% of the samples. A Support Vector Machine (SVM) classifier trained on this balanced subset achieved strong performance when applied to the full dataset.

To generate the final Uniform Manifold Approximation and Projection (UMAP) histogram, we embedded our behavior latent feature (12D) into two-component UMAP manifold (umap-learn.umap.UMAP with n_neighbors = 200 and min_dist = 0.5). The embedded space was binned into a 200×200 histogram and smoothed with a 2D gaussian kernel ($\sigma = 3$ for Figure 4A; $\sigma = 6$ for Figure S5A). The decision boundaries in UMAP histogram were delineated based on the peak cluster label within each bin. The histogram's entropy was calculated from its flattened density (40,000 elements) using scipy.stats.entropy.

As UMAP is non-parametric, a model was necessary to learn the mapping from the latent feature to this 2D UMAP space. A multilayer perceptron (sklearn.neural_network.MLPRegressor, hidden layer size of $100 \times 100 \times 20$ units) was used for this purpose, allowing different datasets sharing the identical 2D UMAP space.

Manual annotation of clusters

Human annotators reviewed video clips randomly sampled from each behavioral cluster. Behavior names were typically assigned from the rat-s's perspective, referring to guidance.[64] Considering the consistency matrix, pairs of reciprocal clusters were given complementary names (e.g., chasing versus being chased). Additionally, clusters with closely related dendrogram branches were categorized under the same coarse social category while emphasizing their differences (e.g., fast chasing versus slow chasing).

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

Closed-loop, behavior-triggered optogenetic stimulation

Real-time behavioral detection

The real-time processing pipeline began with multicamera video acquisition (3840×2400 resolution at 30 fps) streamed to cloud servers using OBS Studio with MPEG encoding at 10 Mbps via RTSP protocol. The media server (Mediamtx, https://github.com/bluenviron/mediamtx) with ffmpegcv tool decoded incoming streams within 133-ms latency, feeding the video data to a parallelized processing architecture where three GPU-accelerated models operated simultaneously: YOLOv8n-seg performed instance segmentation, DANNCE reconstructed 3D poses, and a Bidirectional Long Short-Term Memory Fully Connected Network (BiLSTM-FCN) classifier (12-frame window) identified behaviors from 36 distinct categories. Model outputs were temporally smoothed using a 7-frame median filter and validated against offline labels from 40-video dataset.

Latency measurement

System latency was measured through timestamped virtual frames generated by pyvirtualcam. A custom-built Optical Character Recognition (OCR) pipeline automatically detected and decoded numeric timestamps embedded in each video frame, with the extracted timestamps cross-referenced against the system clock to quantify processing delays across all pipeline components (detailed latency breakdown in Figure S7C). Behavior classifications were transmitted to the client computer via custom socket communication (port 8001), where a GUI allowed researchers to select target behaviors for optogenetic triggering.

Optogenetic stimulation protocol

Closed-loop: Upon target behavior detection, an Arduino-controlled laser system delivered precisely timed stimulation (470 nm, 4 mW, 40 Hz, 0.5-s pulses). Open-loop: control animals received matched total stimulation (11%–15% of session time) randomly distributed, with no delivery in the 2-second window following detected play behaviors. All sessions were recorded with color-coded virtual indicators marking stimulation events, enabling offline verification of behavior detection accuracy, stimulation timing precision, and animal responses. This integrated system achieved closed-loop control with 266 ± 20 ms end-end latency, establishing a powerful platform for investigating behavior-specific neural circuit function. Complete implementation details are available at Github (https://github.com/chenxinfeng4/social-seq-live-client-GUI).

To validate the sub-second end-to-end latency, we performed offline analyses on closed-loop video recordings. Behavior prediction was conducted on 0.8-s video clips, assigning labels to the central frame in a frame-by-frame sliding manner. Consequently, trigger latency was quantified as the interval between the algorithmic detection of the targeted syllable (e.g., pouncing) and the subsequent laser pulse delivery.

Real-time system latency and stimulation timing

A critical parameter of our closed-loop system is the temporal alignment between behavioral onset and optogenetic stimulation. Our real-time behavioral classifier evaluates a sliding history window of 400 ms. Furthermore, the end-to-end computational pipeline—comprising image acquisition, YOLOv8 segmentation, DANNCE 3D pose estimation, BiLSTM-FCN inference, and TTL pulse generation—introduces a processing latency of approximately 266 ms.

Cumulatively, this results in an effective system delay of ∼666 ms from the true physical initiation of a behavior. Because a typical targeted proactive play bout (e.g., pouncing/pinning) typically lasts well beyond this initial timeframe, the optogenetic stimulation reliably begins during the ongoing execution phase of the behavioral bout. Crucially, the stimulation delivery is behavior-contingent: the laser remains active only as long as the targeted behavior is continuously detected, with a stimulation bout duration of 0.5 s, and it terminates immediately upon the cessation of the behavior.

Fiber photometry for GRAB-DA recordings

Hardware setup

Photometry and behavioral data were collected simultaneously. The photometry system was assembled following previously described methods.[62] Briefly, a 405 nm (UV), a 480 nm (blue) LED and a 565 nm (yellow) LED were alternately illuminated at a toatl frame rate of 30 Hz (10 Hz per channel). Emitted fluorescence was collected through the same optical fiber and recorded at 30 Hz using a scientific CMOS camera (Hamamatsu). LED pulses and CMOS exposure were synchronized via a customized Arduino-based timing controller.[65]

For signal processing, fluorescence intensity was extracted by calculating the mean pixel value within the fiber region of interest (ROI). Because fluorescent sensors typically exhibit broad emission spectra that can cause optical bleed-through between channels, we employed time-division illumination by alternately pulsing each excitation light source. The composite recorded signal was then computationally demodulated (temporally demultiplexed) into separate emission traces evoked by each specific excitation wavelength: 405 nm (isosbestic control), 480 nm (DA3m) and 565 nm (rDA3m), based on their distinct illumination timing.[66] Signals from these channels were median-filtered with sliding 0.5-s windows, then bleach-corrected by subtracting double-exponential fits. To rigorously control for motion artifacts and body occlusion—especially during intense physical contact—signals were robustly regressed using the 405 nm isosbestic channel, which reflects dopamine-independent fluorescence. Specifically, the 405 nm reference trace was scaled to fit the 480 nm (DA3m) signal via robust linear regression, and the fitted baseline was subtracted to compute the dF/F.

Synchronization with video recording

We developed a custom synchronization system to achieve millisecond-level temporal alignment between photometry recordings and multi-camera behavioral video. The photometry system initiation served as the master trigger, activating a synchronization

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

controller through local socket communication (port 20169). This controller then coordinated the simultaneous start of behavioral video recording by sending commands via OBS websocket protocol (port 4455). The synchronization software is publicly available at https://github.com/chenxinfeng4/f2_synchronizer.

Temporal auto-correlation analysis of behavioral features

To quantify the temporal structure of social interactions, we performed autocorrelation analysis on each behavioral feature using the statsmodels.api.tsa.stattools.acf function in Python. Feature time series were concatenated across all experimental sessions to compute correlation distances between time-lagged data points. The resulting autocorrelograms were modeled as exponentially decaying processes using nonlinear least-squares optimization (scipy.optimize.curve_fit) with the function:$y = (1 - b) \times \exp(-x / \tau) + b$ where y represents the autocorrelation coefficient at time lag x, and b is a consistent offset. The characteristic decay time constant (τ) represents behavioral features persistence during social interactions.

Similarity analysis of behavioral clusters

To quantify relationships between behavioral clusters, we developed a correlation-based similarity metric in the dimensionally reduced latent space (12 PCs). For each cluster pair, we randomly sampled 200 behavioral clips per cluster (with replacement for clusters containing fewer samples) and computed pairwise Pearson correlation coefficients between their latent representation. This generated a symmetric K × K (K= number of behavioral clusters) correlation coefficient matrix (CCM), where each entry reflected the mean cross-cluster similarity. In Figure 2G, cluster distances were determined by computing the Euclidean distance between the CCM representations of each cluster (as K×1 vectors), with hierarchical relationships visualized using complete-linkage clustering (scipy.cluster.hierarchy.linkage). The resulting dendrogram reveals the topology of behavioral relationships, where branch lengths correspond to the degree of latent space divergence between clusters.

Active-learning based refinement of behavioral clusters

Clip labeling from different perspectives

To establish a robust behavioral classification system invariant to observer perspective, we implemented a multi-stage validation and refinement protocol. The analysis was performed on 800-ms behavioral clips derived from proximal interactions (<1 body length distance) between paired rats. For each behavioral clip, we computed two distinct feature representations corresponding to the egocentric perspective of each rat. These representations were then classified using our initial behavioral clusters, generating perspective-dependent label assignments.

Consistency matrix construction

To quantify inter-perspective consistency, we constructed a probability consistency matrix (PCM; dimensions K × K), where each element $PCM(i, j) = N(i \cap j)/N(i)$. Here, $N(i)$ represents the count of frames assigned to cluster "i" from rat w's perspective, and $N(i \cap j)$ represents the frames co-occurred in cluster "i" (rat w) and cluster "j" (rat s). Thus, $PCM(i, j)$ provides a normalized measure of cross-perspective agreement.

Identification of reliable behavioral clusters

Clusters demonstrating >50% inter-perspective agreement ($PCM(i, j)>0.5$) were classified as reliable. These reliable clusters were further categorized into: (1) Mutual behaviors: Clusters showing identical labeling from both perspectives (e.g., head-to-head approaching); (2) Reciprocal role pairs: Reciprocal-symmetric clusters representing complementary social roles (e.g., chasing versus being chased).

Automatic classification refinement

To improve classification robustness, we implemented an iterative refinement procedure using a hybrid Sequence-to-Sequence Fully Connected Network (Seq2Seq-FCN) architecture. The model combined a Seq2Seq autoencoder (128-dimensional latent space) with a fully-connected classifier network (hidden layer size of 64×36 units with ReLU activation). Beginning with label-consistent clips in each reliable cluster (K = 36), the system generated pseudo-labels for the full dataset which were then used to retrain the model. This active-learning process converged after four iterations, achieving 89.2% perspective-invariant classification in proximal interaction (<1 body length distance) dataset.

The trained Seq2seq-FCN model exhibited robust generalization capabilities when applied to behavioral context beyond its original training distribution. For interactions occurring at distances greater than one body length, the model maintained 86.5% perspective-invariant classification accuracy (Figure S3E), demonstrating effective transfer of learned behavioral representations to distal social interactions. Furthermore, the model performed consistently across different animal pair configurations, achieving overall 87.4% perspective-invariant classification accuracy for WT::WT and *Shank3*$^{+/-}$::WT pairs with optic fiber implants respectively, indicating its robustness to both anatomical and genetic variations.

Behavioral Metrics Analysis

Behavioral proportion quantification

The behavior proportion of each behavior cluster was determined by counting the labels of video clips (each 0.8 s). To highlight differences in behavior proportion in the same cluster between different groups (e.g., A *vs.* B), we calculated the normalized fold change (ranging from -1 to 1) as follows:

$$normalized_fold_change = \frac{P_A^k - P_B^k}{P_A^k + P_B^k}$$

Note that P_A^k is the behavior proportion by the group A in cluster "k". A normalized fold change of 0 indicates equal behavior proportion by both groups. Positive values signify that group A had a higher proportion in the cluster relative to group B, and negative values indicate that group B had a higher proportion relative to group A.

To isolate relatively pure approaching behavior in Figures 4B, 4C, 5, and S5, head-to-head approaching events were classified as "approaching" only when the subject rat's mean velocity during the event exceeded 0.4 times the mean of the maximum velocities of all rats across all videos. "Being chased" was classified as a passive state when the partner persistently pursued the subject at high velocity; and "leaving" was defined as a deliberate, continuous increase in inter-animal distance initiated by the subject to disengage from an ongoing proximal interaction.

Behavior transition analysis

Behavioral label sequences were temporally ordered and concatenated separately for each group. Both sequences were filtered to retain only relevant behaviors, with gaps omitted. Transition probabilities PT($j|i$) is computed as the number of transitions from behavior i to j normalized to the count of the initial behavior i. However, for infrequent behaviors (e.g., pinning), we normalized to the count of the target behavior j. Transition diagrams depict only behaviorally relevant transitions (self-transitions excluded). Statistical significance was assessed using Kruskal-Walls and post hoc Dunn's test with Benjamini-Hochberg false discovery rate (FDR) correction ($p < 0.05$) for each transition pair. In the results, significant differences are highlighted for emphasis.

Low-dimensional behavioral profiling

We quantified behavioral phenotypes by representing each video recording as a 36-dimensional feature vector, where each element corresponded to the relative proportion of a specific behavior cluster (Figures 4A, 5B, 5D, and 5F). Following min-max normalization (0–1 range) of these behavioral features, we performed PCA to project the high-dimensional data into an interpretable 2D space while preserving maximal variance. This dimensionality reduction allowed each video session to be visualized as a single point in the principal component space, with spatial position reflecting the animal's overall behavioral profile.

To statistically assess group differences in this reduced PC space, we implemented a linear SVM classifier. The classifier's discriminative performance was rigorously evaluated using leave-one-out cross-validation, with model accuracy quantified through the F1 score – a harmonic mean of precision and recall that provides balanced assessment of classification performance, particularly valuable for datasets with potential class imbalances. This approach enabled sensitive detection of systematic differences in behavioral patterns between experimental groups while accounting for individual variability.

To visualize sexual dimorphism on genotype-related behavioral profiles, we applied Linear Discriminant Analysis (LDA) for dimensionality reduction (Figure 5H).[67] To mitigate overfitting, we first projected 36D behavioral vectors into 18D PC space before applying LDA to further reduce to 2D. Cluster ellipses, computed via Singular Value Decomposition (SVD), delineate 2.1 σ boundaries (89.0% confidence intervals) to visualize group dispersion.

Behavioral Profile Similarity Analysis

To quantitatively assess inter-group behavioral differences, we maintained the original 36-dimensional normalized feature vectors representing the proportional distribution of behavior clusters for each video session. For each individual subject, we computed two key similarity metrics using cosine distance: (1) the mean distance to all other videos within the same genotype (within-group similarity) and (2) the mean distance to videos of the opposite genotype (between-group dissimilarity) (Figure 5I).

Analyzing behavior-related DA fluctuations

Photometry data preprocessing

We implemented a standardized preprocessing pipeline for both wired and wireless photometry recordings with modality-specific optimizations (https://github.com/ThomasAkam/photometry_preprocessing). For wired recordings (10 Hz acquisition), raw fluorescence traces first underwent median filtering (5-frame window) to attenuate high-frequency noise, followed by photobleaching correction through double-exponential fit subtraction. Motion artifacts were mitigated via linear regression of the 405 nm isosbestic signal against the DA3m channel. High-fluctuation artifacts were identified by convolving motion-corrected traces with a symmetric triangular kernel (5-s window) and applying a threshold-based detection algorithm, with artifact periods replaced using median-filter interpolation. ΔF/F normalization was performed using the baseline fluorescence (F) derived from the photobleaching fit. Wireless recordings (100 Hz single-channel) were processed similarly but omitted motion correction due to the absence of an isosbestic channel and employed a shorter detection window (0.5 s) for artifact identification. Processed signals were downsampled to 10 Hz for cross-modality consistency. All traces underwent session-specific z-scoring, with quality control excluding recordings exhibiting either maximum z-scores < 2.0 or inter-frame fluctuations > 6 SD.

Behavioral classification and event detection

We implemented a hierarchical behavioral analysis system beginning with frame-level classification of 0.8-second clips (sampled at 100 ms intervals) using our pre-trained Seq2Seq-FCN model. The initial 36 behavioral clusters were consolidated into 15 meaningful categories based on kinematic and postural similarities, organized by increasing social intensity (Table S2). Given the relatively low occurrence frequency of chasing and play-related behaviors in some video recordings, we further consolidated slow and fast chasing into a single "chasing" category, and similarly combined pouncing and pinning into a unified "pouncing-pinning" module.

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

Furthermore, specific criteria were established for directed spatial interactions: "approaching" was defined as the subject initiating movement toward the partner, parameterized specifically by the subject's own locomotor velocity; "being chased" was classified as a passive state when the partner persistently pursued the subject at high velocity; "evading" was distinguished by the subject's active, rapid acceleration to escape the partner's proximal approach; and "leaving" was defined as a deliberate, continuous increase in inter-animal distance initiated by the subject to disengage from an ongoing proximal interaction. This consolidation ultimately yielded 10 core behavioral modules for dopaminergic correlation analysis.

To ensure temporal continuity, brief behavioral fragments (< 0.3 s duration) were reassigned to preceding labels. Additionally, sequences of the same behavioral category separated by brief intervals (< 1.6 s) were merged into a single continuous bout only if the intervening behaviors exhibited a lower social intensity. For instance, if two high-intensity pouncing events were interrupted by a brief (< 1.6 s) lower-intensity chasing or sniffing episode, the entire sequence was concatenated into a single pouncing bout. Conversely, if two lower-intensity sniffing events were separated by a brief, higher-intensity pouncing episode, they remained distinct and were not merged. Final behavioral bouts were selected using duration thresholds ($\geq$ 0.8 s for chasing and approaching, $\geq$ 1.0 s for others) to capture sustained behavioral states.

Behavioral initiation points analyses

We established a comprehensive pipeline for detecting behavioral initiation events across both social and non-social interaction types. For dynamic social behaviors such as chasing, we first applied a 1.0-second mean filter (10-point window) to their corresponding kinematic features (e.g., individual velocity or structural overlap) within each experimental session. Specifically, initiation points were identified within the window from 500 ms prior the behavioral label onset to label offset when the smoothed kinematic values first exceeded the following thresholds above the session mean: 0.8 SD for chasing-related events (including active chasing, being chased, and evading) based on velocity; 0.1 SD for approaching based on the subject's velocity; 1.8 SD for leaving based on the change in inter-animal distance over a 0.4-second interval; and 0.5 SD for both sniffing (based on nose-tail overlap) and pouncing-pinning (based on body-body overlap).

For stationary behaviors such as rearing, we implemented an alternative detection scheme where initiation points correspond to local maxima of their characteristic kinematic features (e.g., height). This dual detection strategy ensured accurate and behaviorally meaningful initiation timing across all interaction types. The complete mapping between distinct behavioral modules and their respective kinematic features, along with detailed validation criteria, are provided in Table S2.

Dopaminergic dynamics analysis

For each behavioral module, DA transients were analyzed within precisely defined temporal windows relative to feature-aligned behavioral onset (t = 0). The response magnitude was calculated as the difference between the peak z-scored $\Delta F/F$ value in the post-onset period and the baseline mean during the pre-onset window (-1.0 to -0.5 s). The post-onset window was 0–1.0 s for behaviors including sniffing, chasing, and pouncing-pinning, and a 0.5-s lagged window (0.5–1.5 s) for rearing, approaching, evading, and leaving behaviors to account for prolonged arousal effects. To ensure robust statistical power while accounting for behavioral frequency differences, we implemented minimum occurrence thresholds: $\geq$ 3 events per experiment for all behaviors. For each valid behavioral module within an experiment, we computed the average DA response profile across all qualifying events. Peak latency for each behavioral module was defined as the interval from behavioral onset to the peak response within the corresponding post-onset window.

Dopamine-kinematics behavioral relationship analysis

The relationships between dopamine responses and behavioral kinematics were explored during social interactions between WT male rats. For each behavioral module event, we extracted z-scored DA signals spanning 1.0 s before to 1.0 s after onset, normalized by subtracting the mean within this window. Concurrent kinematic features for initiation analysis (Table S2) underwent the aforementioned temporal alignment and normalization procedures. Within each experimental session, we concatenated these time-aligned DA signals and kinematic traces across all occurrences of a given behavioral module, maintaining the minimum event threshold of $\geq$ 5 repetitions per module.

Pearson correlation coefficients were then computed between the concatenated DA signals and each kinematic feature on a per-experiment, per-module basis. This approach preserves the temporal resolution needed to detect transient dopamine-kinematics coupling while aggregating sufficient data points for robust correlation analysis. The resulting correlation matrices capture how specific movement parameters (e.g., velocity, body occlusion) covary with dopaminergic signals during distinct social interaction types.

Dopamine-magnitude behavioral duration relationship analysis

To examine the relationship between DA transient magnitude and behavioral duration, analyses were performed across all genotypes and sexes. For each behavioral module, the proportion of session time occupied and the mean DA response amplitude (averaged across all qualifying events within the session) were calculated. Pearson correlation coefficients were then computed between the mean DA response and the behavioral duration percentage for each module. These correlation matrices were used to assess whether the enrichment of specific behaviors (e.g., rearing, sniffing) covaries with dopaminergic signals.

Analysis of the relationship between dopamine magnitude and antecedent behaviors

To examine the relationship between DA transient magnitude and behavioral sequences, analyses were performed across all genotypes and sexes. For a given behavior module B (e.g. pouncing), the mean behavioral DA response amplitude was computed as the average across all qualifying events within a session. To assess the influence of preceding behavioral context, the antecedent probability of module A (e.g. approaching) preceding B was calculated as the proportion of events in which A immediately preceded B

relative to the total number of B events. Pearson correlation coefficients were then computed between the mean DA response and this antecedent probability for each module–module pair.

Multi-agent reinforcement learning model

Agents

In our settings, we simulated rats using Proximal Policy Optimization (PPO) agents of RLlib. During each experiment, the PPO agents of different reward settings stands for different types of rat (*i.e.* male WT, Male *Shank3*, Female WT, and Female *Shank3*). Unlike standard pursuit-evasion paradigms, we equipped each agent with a decoupled, multi-discrete action space that simultaneously outputs spatial navigation (five directional movements: stay, up, down, left, right) and explicit social intents (four discrete actions: none, sniff, pounce, escape).

Environment and observation space

We conducted experiments to simulate several social interactions between two agents within a 4×4 grid world environment, and recorded the proportion of the behavior for each agent. Since the biology and simulated environments are both small, we assumed that the observation space is fully observable, providing each agent with a flattened one-hot encoded state vector representing the exact spatial coordinates of itself and its interacting partner.

Action space and social interactions

To better match social interactions in biology experiments, we separated action space of the agents into movement actions and social actions so that the agent can output spatial navigation (five discrete directional movements: stay, up, down, left, right) and explicit social intents (four discrete social actions: none, sniff, pounce, escape) simultaneously. Moreover, the social interactions are evaluated based on Manhattan distance between the agents. High-intensity contact behaviors, such as Pouncing (active) and Pounced (passive), were only evaluated when agents occupied the exact same coordinate (Manhattan distance = 0). Proximal social behaviors, Sniffing (active) and Sniffed (passive), were only evaluated when agents were located at adjacent coordinates (Manhattan distance ≤ 0). In addition, to model mutual social interactions, simultaneous executed sniff actions lead to active sniffing states for both agents while simultaneous executed pounce actions lead to one pouncing state for a random agent and one pounced state for the other.

For distant social interactions, a Chasing and Chased event pair was defined strictly when both agents exhibited identical movement vectors. Furthermore, to distinguish purely spatial repositioning from socially motivated movement, we defined an Approach event as any dynamically driven action that resulted in a reduced absolute Manhattan distance between the two agents, whereas a Leave event was defined as an action that actively increased this inter-agent distance. To prevent behavioral overlap in statistical evaluation, proximal social interactions (e.g., Pouncing and Sniffing) strictly superseded these distal movement states via a priority-lock mechanism.

Reward shaping

The reward-shaping parameters for different agent groups were calibrated using DA signals measured in the biological experiments. For the behavioral states central to our comparison, particularly the pouncing and pounced states in WT and *Shank3* animals, we directly used the corresponding empirical DA values. Rewards for the remaining ancillary behaviors were calibrated to remain within the empirically observed DA range and to preserve the qualitative trends of the biological signals. To reflect the variability observed in the dopamine recordings, a zero-mean stochastic fluctuation was added whenever a configured behavioral reward was delivered. The fluctuation was independently drawn between -0.5 and $+0.5$ reward units at each reward event. Sampling was controlled by the run-specific random seed and was not manually assigned to individual behaviors or trials. The same fluctuation range was applied across behaviors as a simplifying model assumption. Consequently, the reinforcement learning agents reproduced the key genotype- and sex-dependent patterns of social interaction observed in the biological animals. Detailed comparisons of the empirical mean values of DA fluctuations (top, blank) against the final reward values utilized in the multi-agent reinforcement learning (MARL) agents (bottom, colored) are detailed in Table S3.

Agent architecture

The input to each agent was a flattened, concatenated vector representing the fully observable spatial state of the environment. For the 4x4 grid world, each spatial coordinate was flattened into a vector of length 16. The final input was the concatenation of two such vectors: one representing the one-hot encoded location of the focal agent (self), and the other representing the one-hot encoded location of the interacting partner. Thus, the total input state vector x_t had a dimension of 32. We utilized a fully connected Multi-Layer Perceptron (MLP) actor-critic architecture provided by the RLlib framework. By default, the policy (actor) and the state-value function (critic) were parameterized by independent neural networks with identical structural hyperparameters to prevent representation bottlenecking. For both the actor and the critic networks, the input x_t was fed into two consecutive hidden layers, each consisting of 256 hidden units equipped with hyperbolic tangent (tanh) activation functions. The feedforward propagation for the hidden representations is defined as:

$$h^{(1)} = tanh\left(W^{(1)}x_t + b^{(1)}\right)$$

$$h^{(2)} = tanh\left(W^{(2)}h^{(1)} + b^{(1)}\right)$$

Please cite this article in press as: Chen et al., Sub-second dopamine reinforcement shapes juvenile social play, Cell (2027), https://doi.org/10.1016/j.cell.2026.10.040

where $x_t \in R^{32\times1}$, $W^{(1)} \in R^{256\times32}$, $b^{(1)} \in R^{256\times1}$, $W^{(2)} \in R^{256\times256}$, $b^{(2)} \in R^{256\times1}$

For the critic network, the state value readout v_t is computed via a linear layer applied to its respective second hidden layer $h^{(2)}_{critic}$:where $W_{value} \in R^{1\times256}$, $b_{value} \in R^1$, and $v_t \in R^1$is the scalar baseline value used for advantage estimation.

For the actor network, to accommodate the decoupled spatial and social decision-making, the action readout maps the hidden representation $h^{(2)}_{actor}$ to a joint logit space:

$$a_t = W_{action}h^{(2)}_{actor} + b_{action}$$

where $W_{action} \in R^{9\times256}$, $b_{action} \in R^{9\times1}$, $a_t \in R^{9\times1}$

Crucially, because we formulated the action space as a multi-discrete distribution of [5, 4], the 9-dimensional raw logit output a_t, was deterministically partitioned into two independent components. The first 5 logits were passed through a softmax function to sample the spatial movement action ($a_{mov} \in \{0,1,2,3,4\}$), while the remaining 4 logits were independently passed through a parallel softmax function to sample the explicit social intent ($a_{int} \in \{0,1,2,3\}$). This branched readout allowed the agents to simultaneously navigate the spatial grid and broadcast proximal social intents without cross-channel interference.

Agent training and evaluation

We trained the multi-agent models using the Proximal Policy Optimization (PPO) algorithm implemented in the Ray RLlib framework. The training process spanned up to 120,000 environment timesteps to allow the joint policies to plateau and converge toward stable social dynamics. Data collection was partitioned into iterative rollouts, with each iteration explicitly gathering a batch of 2,000 timesteps (equivalent to 20 full episodes, given the strict episode length limit of 100 timesteps). During each iteration, the collected batch was divided into minibatches of size 128, and the actor-critic networks were updated over 10 Stochastic Gradient Descent (SGD) epochs. To stabilize the policy updates, we applied a clipping parameter of $\epsilon = 0.2$. Because multi-agent environments often suffer from premature convergence to suboptimal local equilibria (such as degenerate static behaviors), we introduced an entropy coefficient of 0.01 to encourage continuous exploration of the decoupled spatial-social action space. The discount factor γ was set to 0.95 to appropriately prioritize immediate behavioral feedback during proximal interactions, and the learning rate was fixed at 3×10^{-4} using the Adam optimizer. Multiple independent runs from different random seeds were conducted to ensure the reproducibility of the emergent behaviors. Although we trained two PPO agents until they got stable policies, we only evaluated them at intermediate step (e.g. 50,000 step) since we wanted to simulate the agents when their policies varied the most. During evaluation, the agent pairs were evaluated over 30 independent rollout episodes as actions were sampled stochastically from the probability distribution computed by the actor network to accurately capture the behavioral variance.

Software packages

The Social-seq-Live platform was built on Python 3.7.16 within an Ubuntu 18.04 environment. The computational backbone consisted of an Intel Xeon 4314 CPU and four NVIDIA RTX 3090 GPUs, leveraging CUDA 11.4, cuDNN 8.8 and TensorRT 8.6 for deep learning operations through TensorFlow 2.7.0 and PyTorch 1.11.0 frameworks. The system incorporated OpenCV-Python 4.8.0 for core image processing tasks, enabling efficient handling of multi-camera video streams.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of the statistical analysis including sample size were provided in the results or figure legends. Both parametric and non-parametric tests were used wherever appropriate. The FDR was used to conceptualize type I errors when conducting multiple comparisons. Statistical significance was defined as follows: *ns*. not significant $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Box plots adhere to standard conventions: the box boundaries represent the first and third quartiles, while whiskers extend to include the extreme data point within 1.5× interquartile range of either the first or third quartile. Bar plots and other quantitative data were represented as mean ± standard error of the mean (SEM) unless otherwise specified. Two-Way Repeated Measures ANOVAs (Factors: Sex×Behavioral Role [Active vs. Passive]) were conducted to assess the main effects and interactions. All graphs and analyses were performed using custom code written in Python 3 based on standard libraries: matplotlib, numpy, scipy and statsmodels.