Behavioral/Cognitive

Emergent Exploration via Novelty Management

Goren Gordon,[1] Ehud Fonio,[2] and Ehud Ahissar[1]
Departments of [1]Neurobiology and [2]Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel

When encountering novel environments, animals perform complex yet structured exploratory behaviors. Despite their typical structuring, the principles underlying exploratory patterns are still not sufficiently understood. Here we analyzed exploratory behavioral data from two modalities: whisking and locomotion in rats and mice. We found that these rodents maximized novelty signal-to-noise ratio during each exploration episode, where novelty is defined as the accumulated information gain. We further found that these rodents maximized novelty during outbound exploration, used novelty-triggered withdrawal-like retreat behavior, and explored the environment in a novelty-descending sequence. We applied a hierarchical curiosity model, which incorporates these principles, to both modalities. We show that the model captures the major components of exploratory behavior in multiple timescales: single excursions, exploratory episodes, and developmental timeline. The model predicted that novelty is managed across exploratory modalities. Using a novel experimental setup in which mice encountered a novel object for the first time in their life, we tested and validated this prediction. Further predictions, related to the development of brain circuitry, are described. This study demonstrates that rodents select exploratory actions according to a novelty management framework and suggests a plausible mechanism by which mammalian exploration primitives can be learned during development and integrated in adult exploration of complex environments.

Key words: **active sensing; hierarchical model; intrinsic motivation; reinforcement learning; whisker system**

Introduction

Mice and rats show a structured buildup in the extent and complexity of their exploratory behavior mediated by whisking and locomotion. In unfamiliar environments free of objects or before contact, rodents have been observed to exhibit periodic exploratory whisking (Gao et al., 2001; Berg and Kleinfeld, 2003; Kleinfeld et al., 2006; Diamond et al., 2008; Towal and Hartmann, 2008). Upon contact, a larger repertoire of behaviors is observed, such as rapid cessation of protraction upon initial contact (Mitchinson et al., 2007; Grant et al., 2009), touch-induced pumps (Deutsch et al., 2012), and repetitive palpation (Knutsen et al., 2006) accompanied by touch with their snout (Welker, 1964). Freely moving mice exhibit complex exploration behaviors that grow in dimensionality when exploring a novel arena, from circling in place, through wall following to open-space exploration (Fonio et al., 2009; Benjamini et al., 2011). Modeling efforts had been focused so far on specific scenarios (Tchernichovski et al., 1998; Harish and Golomb, 2010) or high-level information theoretical aspects (Polani, 2009; Still, 2009; Friston, 2010; Tishby and Polani, 2011) of these behaviors. However, a quantitative generic model, based on information-theoretical grounds, that accounts for the underlying principles of exploration behavior and is applicable across modalities, species, and timescales is still lacking.

An emerging modeling effort attempts to model exploration behavior and implement it in artificial agents using intrinsic motivation and rewarding novelty (Harlow, 1950; Schmidhuber, 1990; Kaplan and Oudeyer, 2007; Oudeyer et al., 2007; Singh et al., 2010; Baldassarre, 2011), where novelty is taken to be a quantitative measure of how much information about agent–environment interaction is gained (Friston, 2010; Little and Sommer, 2013). While novelty seeking was often modeled, typically, in combinations with other goals (Simpkins et al., 2008; Berger-Tal et al., 2014), novelty management (i.e., controlled alternation between neophilia and neophobia; Barnett, 1958; Misslin and Cigrang, 1986; File, 2001; Bahar et al., 2004; Hughes, 2007), also referred to as approach-avoidance behavior (Elliot, 2006; Hughes, 2007), has not been modeled thus far. We have recently developed a model wherein novelty is managed by alternate switching between efficient novelty seeking (Berlyne, 1960) and reflexive-like novelty-aversive (Hughes, 2007) motor primitives (Gordon et al., 2014).

Continuing this line of development, we consider here agents to engage in active perception of the structure of their world by acquiring novel sensory information. Exploration is composed of a sequence of excursions into an environment: whisking in the space around the snout (Deutsch et al., 2012) and locomotion into an approachable arena (Fonio et al., 2009; Benjamini et al., 2011). We show that our model can explain a large repertoire of behaviors, on multiple timescales, and predict novel behaviors. This is first presented in simulation results of artificial agents in the whisking and locomotion systems, followed by model-based

Received May 8, 2014; revised July 22, 2014; accepted Aug. 1, 2014.

Author contributions: G.G., E.F., and E.A. designed research; G.G. and E.F. performed research; G.G. and E.F. analyzed data; G.G., E.F., and E.A. wrote the paper.

This work was supported by the United States-Israel Binational Science Foundation Grant 201143; the Minerva Foundation, funded by the Federal German Ministry for Education and Research; the office of the Chief Scientist, Israeli Ministry of Health; the Weizmann-UK Joint Research Program; and a fund from Lord Alliance for the Weizmann-Manchester Life Science Collaboration. E.A. holds the Helen Diller Family Professorial Chair of Neurobiology.

The authors declare no competing financial interests.

Correspondence should be addressed to Ehud Ahissar. Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: ehud.ahissar@weizmann.ac.il.

DOI:10.1523/JNEUROSCI.1872-14.2014

 0270-6474/14/3412646-16$15.00/0

analysis of empirical data, showing that rodents choose their actions according to a single underlying principle: maximization of novelty signal-to-noise ratio (SNR; Saig et al., 2012; Gordon et al., 2014). We finally verify the hypothesis that novelty management also occurs between modalities in a novel experimental design.

Materials and Methods

We first briefly summarize the model presented by Gordon et al. (2014), followed by a description of the specific model extensions that enabled simulation of behavioral data for whisking and locomotion (see Figs. 2, 3, 4). We then describe the quantification of novelty from experimental data in both the whisking and locomotion system (see Fig. 5). Finally, we describe the experimental setup of the first touch of life (see Fig. 6).

Whisking system model simulation

The full mathematical description of the framework (Fig. 1), as well as the description of the whisking system model simulation is given in the article by Gordon et al. (2014).

Locomotion model simulation

Experimental setup. We briefly describe the experimental setup used in the study by Fonio et al. (2009). It consists of a 250-cm-diameter circular arena illuminated with infrared (IR) lights and dim white lights (~1 lux) placed on the ceiling above arena center, simulating moonlight. The arena is surrounded by a continuous wall with a single 4 × 5 cm doorway leading to an infrared-lit Plexiglas home cage (30 × 40 × 50 cm) containing wood shavings, and food and water *ad libitum*. A small Plexiglas box attached to the home-cage doorway on its inner side forces the mouse to pass through it on its way into the arena without carrying along shavings that might distract the tracking system. Mice were free to exit the home cage and explore the circular arena, wherein their movements were automatically recorded.

Model assumptions. Since the experimental arena was circular, we chose to represent the spatial state in the discretized polar coordinates, where $\theta \in \{2\pi i/N_\theta\}$, $i = 1,\ldots,N_\theta$ is the angular coordinate and $\rho = \{j/N_\rho\}$, $j = 1,\ldots,N_\rho$ is the normalized radial coordinate. We have excluded $\rho = 0$ to disambiguate state transitions in the center of the arena. The circular arena walls were represented by forbidden transitions from one state to an adjacent one. It resulted in $\rho \leq P\,\forall\,\theta$, where $p < 1$ is the radius of the arena.

We modeled the Plexiglas box attached to the home-cage doorway as a singular point or a niche in the arena wall, $cage = \{\theta = 2\pi, \rho = P + 1/N_\rho\}$, surrounded by walls in both angular directions and in the increasing radial direction. We also assumed that the vision of the mouse is lacking, thus entertaining only proximal sensors (e.g., whiskers and snout that can sense the walls).

State and action spaces. The full state of the agent in each time step was thus given by $s = \{\theta, \rho, d, v\}$. For example, $s = \{\theta = 2\pi \cdot 3/24, \rho = P, d = \rho^+, v = \{1, 0, 0\}\}$ means that the agent is facing the arena wall and detecting it with its front sensor. The arena itself could have different sizes and/or discretization (i.e., N_θ, N_ρ) may differ from one exploration episode to another.

Exploration critic and actors. In our implementation of the hierarchical curiosity loops, we used as the reward the information gain after each time step (i.e., after the updates of all three sensors). Furthermore, the local critics and actors depended only on the sensor information v. Hence, the critic $\hat{V}^\pi(v;\mu)$ was represented by $\|\mu\| = 8$ parameters, and the actor $\pi(a|v;\lambda)$ was represented by $\|\lambda\| = 8 \times 4$ probabilities, for forward (F), backward (B), left (L), or right (R) movements.

The critics and actors were each represented by 8 values and 32 probabilities, respectively, which was too complex to visualize. Hence, we wished to have a descriptive numerical representation of the behaviors they represent. Since both critics and actors depended solely on the proximity to walls, we defined the wall parameter $w(v) = \Sigma_{i=F,L,R} v^{(i)}$ and the following three behavioral regimes. (1) Corners were represented by contact with more than one wall, $w(v) \geq 2$, which accounted for four possibilities, namely three options with two walls (forward-left, forward-right, and left-right) and one option with three walls (forward-left-right). Hence, we defined the probability of staying in corners as 1 − the probability of moving away from the detection of multiple walls: $p_{\text{corner}} = 1 - \frac{1}{4}\Sigma_{i=F,\text{R,L,B}}\,\pi(\text{a} = i|v^{(\bar{i})} = 0, w(v) \geq 2)$, where $\bar{i}$ denotes the opposite direction (e.g., $\bar{F} = B$). (2) Single walls, $w(v) = 1$, accounted for three options, namely, wall to the front, right, or left. Hence, we defined the probability to stay near walls as 1 − the probability of moving away from detection of a single wall: $p_{\text{wall}} = 1 - \frac{1}{3}\Sigma_{i=F,R,L}\pi(a = \bar{i}|v^{(i)} = 1, w(v) = 1)$. (3) Open space was represented by no contact with walls, $W(v) = 0$, which accounted for a single possibility. Here we defined the probability of going forward in open space as the directed exploration when there are no detected walls: $p_{\text{open}} = \pi(a = F|w(v) = 0)$.

Exploration novelty management. We implemented a four-loop model, wherein each loop contains a critic and an actor, $\pi^{l=1\ldots4}(a|v;\lambda)$. In each new episode, the arena size was chosen randomly, $N_\theta \sim [9,13]$, $N_\rho \sim [9,11]$, signifying a variable environment and emphasizing the fact that the actors and critics are invariant to arena size and discretization.

Retreat primitive. The retreat motor primitive in the arena exploration implementation used the information already learned about the existence of a wall to return to the home base. In other words, the next action was selected such that there was decreased probability of hitting a wall, and it brought the agent closer to the home cage.

Model parameters and cluster analysis. We set the sensor uncertainty to be identical for all sensors, and varied σ to explore its effects on the emerging behaviors. In our numerical implementation, we varied the prior (P_0) to analyze its effects on the emerging behaviors. We have clustered the parameter space by the 12 converged policy probabilities, namely, $p^{i=1,2,3,4}_{\text{corner,wall,open}}$ [i.e., each parameter pair (σ, P_0) was a single 12-dimensional point]. We then used the k-means algorithm, with $k = 3$, and ended up with regions 1, 2, and 3.

Multimodal model

We used both aforementioned models to generate a simulation of an exploring mouse that both moved in a novel arena and interacted with the objects in it via whisking. The body of the mouse thus moved according to the exploration novelty management behavior, which determined which tactile objects were in its vicinity. The model also consisted of two (uncoupled) whiskers, one to the left and one to the right, each governed by its own novelty management behavior. The interplay between these two levels of description occurred according to the following principles. The first principle was that each position and orientation of the body of the mouse had its own whisker perceiver (i.e., perception of a tactile object such as a wall was defined by position, orientation, and whisker angle during contact); in the arena setup, the model assumes that when there was a wall to the left/right, the object was in the middle of the whisker field, and when there was a wall in front, it was at 80% of the whisker field. The second principle coupled the time between the two levels; moving to another position occurred only after accumulated novelty of both whiskers had increased beyond a certain threshold, w_{th} (measured in bits). A pole, modeled as an object or wall, was placed right outside the arena. Thus, whenever the modeled mouse exited the home cage, it encountered the pole.

Quantifying novelty

Data. As whisking exploratory episodes, we analyzed trials of head-fixed rats whisking against a moving vertical pole ($n = 144$ trials, from 7 male albino Wistar rats; data are from Deutsch et al., 2012). The pole (2.8 mm diameter) moved horizontally toward the rat at a speed of 1.1 cm/s. We considered only the contacting whisker, which was always C2. Whisker tracking and segmentation were identical to those described by Deutsch et al. (2012). As locomotion exploratory episodes, we analyzed sessions of freely moving BALB/c male mice exploring an arena ($n = 11$ sessions, each by a different mouse; data are from Fonio et al., 2009). The arena was circular, with a 250 cm diameter. For each mouse, the exploratory episode included all excursions in the arena from the first entrance to the arena until the first arrival at the center of the arena. Excursion segmentation was identical to that described by Fonio et al. (2009).

Perception and novelty. In each modality, we defined a perceiver as a probability distribution over states, given the previous state and ac-

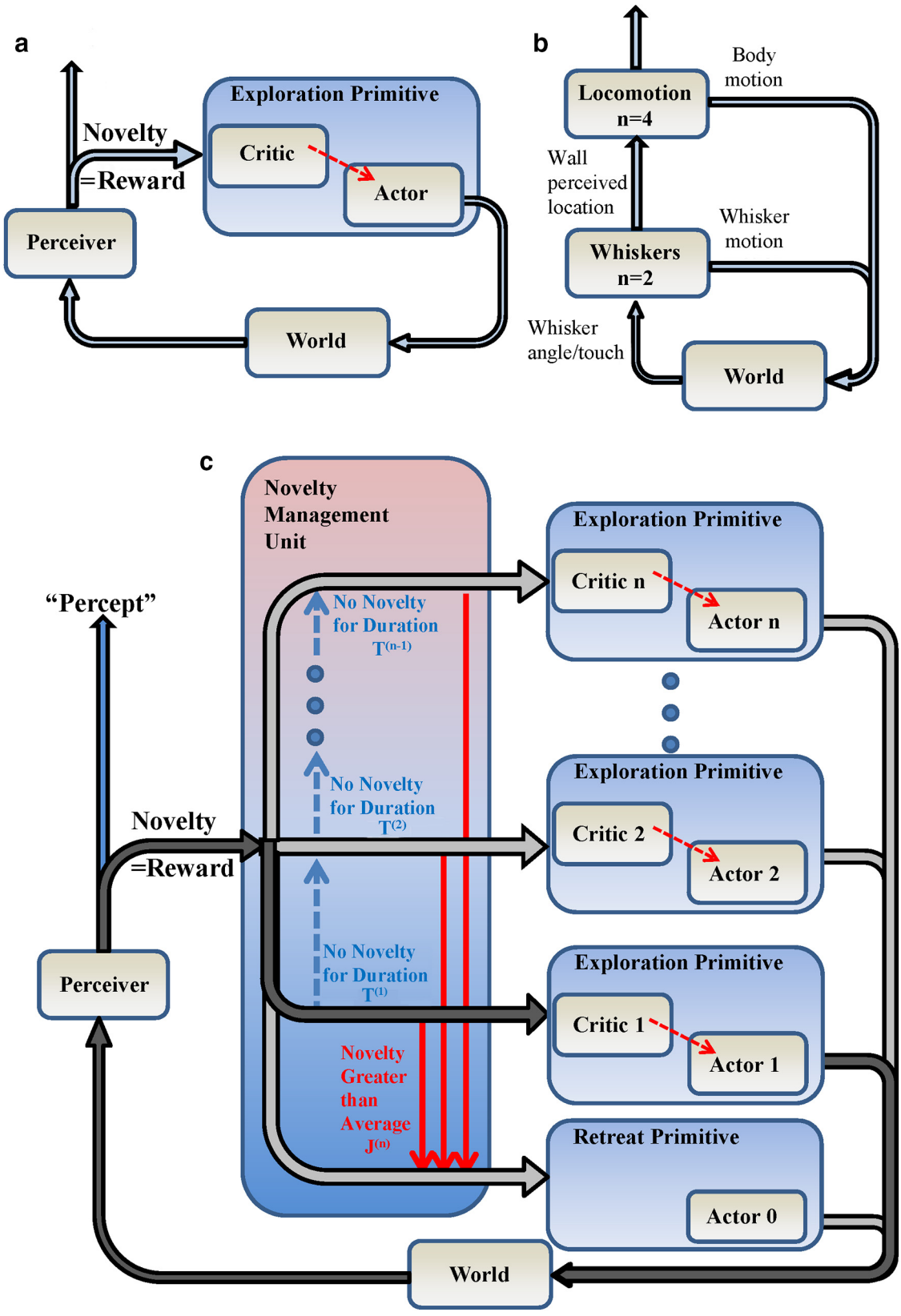

Figure 1. A schematic diagram of the model (adapted from Gordon et al., 2014) extending to arena exploration and across both sensory systems. ***a***, Basic curiosity loop. The agent actively perceives the world through its sensors and learns to predict the next state of its sensors from the current state and the action performed by the actor. Novelty, measured as information gain, is the intrinsic reward for an AC module that implements temporal difference error (dashed red arrow) reinforcement learning. ***b***, A model of an exploring rodent that moves its whiskers to perceive walls and moves its body to perceive an arena. The whiskers modality is composed of two loops, whereas the locomotion modality is composed of four loops. ***c***, A hierarchical model of an active perceptual modality that contains n AC modules and one retreat primitive. At any time, only a single loop is closed (dark arrows); if at any time novelty is higher than the average of the active module, $J^{(n)}$, the retreat primitive is activated (red arrows); if novelty is lower than the average for the duration of the active loop, $T^{(n)}$, the next loop is activated (blue arrows).

tion (i.e., the forward mode $p(s'|s, a)$), where s is the state (e.g., the pole position in the whisking system and the wall location in the locomotion system) and a is the action (e.g., protraction or retraction in the whisking system and egocentric motion in the locomotion system). The perceiver was updated after each observation (o) according to Bayes rule. Novelty flow was defined as the information gain and was calculated with the following Kullback–Leibler divergence: novelty flow $= D_{KL}(p(s|o)\|p(s))$. Novelty was defined as the accumulated novelty flow during an excursion.

Whisking system. Raw data from the study by Deutsch et al. (2012) were processed in the following manner. Whisker angle was automatically tracked from the high-speed videos, and contacts with the pole were manually tracked and marked. The trajectory of whisker angle θ was low-pass filtered at 80 Hz, and discretized to integer degrees, whereas contact c was taken to be binary, where $c = 0$ means no contact and $c = 1$ means contact. Actions (a) were also binarized to either protraction or retraction (i.e., the sign of the angle derivative). The discretization level was chosen such that the animal had controllability at each time step. Higher discretization would include uncontrolled movements due to inertia, whereas lower discretization would exclude known controlled behaviors, such as touch-induced pumps.

Excursions (i.e., whisks) were delineated based on the process described in the study by Deutsch et al. (2012). Only whisks that contained contacts with the pole and had novelty >0.02 bits were analyzed (i.e., only whisks that contained any kind of information gain).

Perception of the environment was modeled here by a probability distribution over angles and contacts, given the previous angle, contact, and the action performed (i.e., the forward model of whisker-pole dynamics; for more details, see Gordon et al., 2014): $p(\theta_t, c_t|\theta_{t-1}, c_{t-1}, a_t)$.

Locomotion system. The raw data from the study by Fonio et al. (2009) were processed in the following manner. The position of the mouse inside the arena was given in polar coordinates, angle θ and radius ρ inside the arena, where (1) the angle was discretized to 12 equally spaced values; (2) the radius was discretized to 6 equally spaced values; (3) stay-in-place was not allowed, hence the time steps in which the discretized angle and radius were not changed were bunched together; and (4) the initial position was a special position (i.e., an excursion that had only one discretized data point) was extended to include a change of direction to allow circling near the exit of the arena. These stages ensured the controllability of the actions of the animal and the tractability of the analysis.

Each point in the trajectory was characterized by the angle and radius; a postural direction expressed in polar coordinates, $d \in \{\rho^+, \rho^-, \theta^+, \theta^-\}$, where $\rho^{\pm}$, where $\rho^{\pm}$ indicates outward/inward direction in the arena and $\theta^{\pm}$ indicates clockwise/anticlockwise direction. The observations were sensor information that represents the snouts and vibrissae of animals, denoted by $v = \{v^F, v^L, v^R\}$, where $v^{F,L,R} = \{0,1\}$ indicates no contact/contact with an object or wall in the front/left/right sensor. The actions were taken to be egocentric (i.e., $a =$ forward/backward/left/right).

The perceiver was given by $p(W|\theta, \rho, d, a)$, where $W = \{0,1\}$ means no wall and wall, respectively. In each time step, the perceiver was updated according to the information from the three sensors. The update was performed using the perceived sensor, $v^{(i)}$, information that substitutes the action in direction $i =$ F, L, R.

$$p(W|\theta, \rho, d, a, v^{(i)}) = \frac{p(W|\theta, \rho, d, a)}{\sigma + p(v^{(i)}|\theta, \rho, d, a)(1-\sigma)} \times \begin{cases} 1 & W = v^{(i)} \\ \sigma & W \neq v^{(i)}, \end{cases}$$

where $\sigma = 0.125$ was the sensor uncertainty. The perceiver was updated thrice, once for each sensor, where the update rule means that the agent learns whether there is a wall in front/right/left of it from the front/right/left sensor, and updates the respective probabilities.

Controls. We performed the following three controls on the data: (1) shuffled: we shuffled the excursion order, but maintained their internal dynamics; (2) random: we randomized the actions performed within each excursion, but maintained the excursions order; and (3) shuffle & random: we shuffled the excursion order and randomized the actions within each excursion. Each control was performed 20 times per exploration episode (whisking session and exploring animal).

Statistical analysis. Bootstrap analysis was applied to test the significance of our results. For each exploration episode, we performed 20 repetitions of the controls (i.e., shuffling the excursion order, randomizing the actions, and both). For each control repetition, we calculated the difference between it and the corresponding data point. Hence, we had 2880 and 220 differences, respectively, for the whisking and locomotion systems. For each experimental-control comparison, we performed t tests on these differences and report the p values of the hypothesis that they are not different than 0.

Retreat calculation. In the whisking system, retreat was measured as the angle change following the first point of maximal novelty flow in each excursion (see Fig. 5*e*,*f*). In the locomotion system, retreat was measured as the distance from the exit following the first point of maximal novelty flow.

Novelty hierarchy exploration. In the locomotion system, novelty was directly related to the number of adjacent walls. Hence, the area could be divided into several special novelty-defined positions. The home cage, denoted High, was at the maximal radius and a specific angle, had three adjacent walls, and hence had the highest novelty density. The circumference of the arena, denoted Medium, was at a smaller radius and all angles, had a single adjacent wall, and hence had medium novelty density. The center of the arena, denoted Low, was everywhere else, had no adjacent walls, and hence had low novelty density. A special place was the garden, which is an open space just beyond the exit, and must be encountered before reaching the walls. In the calculation below, we ignored the garden and treated it as part of Medium. In each exploration time step, the position defined the novelty zone. We defined the zone of each excursion as the zone with the minimal amount of novelty. For example, an excursion starting from the home cage and going to the open space was taken to be a Low excursion. An excursion with only a single time step was not considered, hence it was not trivial that the first visited zone was High, since the mouse must stay and explore the exit of the home cage during the entire excursion for that excursion to be considered High.

First-touch-of-life experimental setup

Setup description. Two families of C57BL/6 mice (2 mature females plus 3 and 9 pups correspondingly) were raised in a "natural habitat" apparatus in which the animals were allowed to freely traverse between a secure shelter (nest-cage) and a novel environment. Pregnant females were purchased and shipped from Harlan, and after an acclimation period of 13–15 d they gave birth. The females and newborn pups [postnatal day 4 (P4)] were housed in the special Perspex nest-cage that was partitioned into two compartments. The first compartment contained standard wood chips and dry food pellets that were given *ad libitum*, and in the second compartment, connected by a shallow angle ramp (30°), there was a free access to water. During the first weeks, the pups huddled in the first compartment until reaching the age of P21–P23 when they began moving around the nest-cage. At that time, a small door (3 × 3 cm), located at the far end of the second compartment and leading to a novel area, was opened, thus allowing free passage between the nest-cage and the novel region, and exploration of the floor and three vertical metal poles (2 mm diameter, 15 cm in length).

The poles were located at three different positions representing the following corresponding contexts: (1) the right pole (with respect to mice coming out from the nest-cage) was the closest to the door (8 cm); (2) the front pole was positioned 15 cm in front of the door; and (3) the left pole was positioned at an intermediate distance (13 cm) from the door and outside the floor range (2.5 cm), close enough to ensure comfortable reach by all mice.

The behavior during entries to the novel area was recorded using a high-speed video camera (CL600x2, Optronis) 1280 × 1024 pixels at 500 frames/s. The setup was designed to allow capturing high-resolution video recordings for long durations of ~12 h (~3 h of continuous recording). The behavior was recorded during nights, which is the active phase of these nocturnal rodents, and in a dark room. A 9 × 9 inch IR illuminator (Metaphase Technologies) was placed below the transparent

novel area and was used as a source of uniform backlight illumination for the camera.

Animal maintenance and all experimental procedures were conducted in accordance with the guidelines of the National Institutes of Health and The Weizmann Institute of Science (Institutional Animal Care and Use Committee approval #01400310-2).

Data acquisition and analysis. We collected in total 25 movies of variable durations, summing up to 177,615 tracked frames (355.23 s). Several body features of the animal were automatically extracted by using the Whisker Standard Tracker (BIOTACT consortium; Perkon et al., 2011), including whisker angular position, head position, head orientation, snout contour, and the position of the tip of the nose (see Fig. 6*a*). All whisker contacts upon poles were detected manually, frame by frame, by a human observer. Identification of touch/no touch was performed for all epochs, whereas for 17 movies of 5 animals a higher-resolution analysis was performed, specifying the exact column to which the touching whisker belongs was given (see Fig. 6*a*).

Contacts that occurred during the same approach episode, determined by the distance of the head from the pole (see Fig. 6*a*), were pooled together and termed "Palpation." Since the poles differ one from the other, the analysis include only mice whose first three palpations were performed upon the same pole (11 animals).

Results

Modeling novelty management within hierarchical curiosity loops

Our main research question is as follows: what general principles guide the selection of actions of exploring rodents? We developed a model that is based on alternation between two types of closed-loop control, namely, curiosity loops and retreat loops (Gordon et al., 2014). This section summarizes the principles and developments of our model that are relevant to the results presented here.

The first principle of maximizing novelty during outbound exploration was modeled by the basic curiosity loop (Schmidhuber, 1990; Tishby and Polani, 2011; Gordon and Ahissar, 2012; Fig. 1*a*), which was composed of (1) a perceiver, which was a Bayesian learner (Bastos et al., 2012) that attempted to predict the next sensory state given the current one and the action performed (i.e., a "forward model"; Kawato, 1999; Lalazar and Vaadia, 2008); (2) a critic, which attempted to predict future accumulated rewards; and (3) an actor, which determined the next action given the current state using a stochastic policy (Sutton and Barto, 1998). The intrinsic reward (Harlow, 1950; Schmidhuber, 1990; Baldassarre, 2011), as opposed to external, goal-directed rewards (Berger-Tal et al., 2014), was given by novelty, and was quantified by the information gain of the perceiver upon sensory updates. The structure of the curiosity loop ensured the emergence of an exploration motor primitive that maximized novelty during its activation.

Actor convergence occurred on a slow developmental timescale, measured in units related to actor-critic (AC) update (αt, where α is the AC learning rate), via interactions with different instances of the same type of environment. We considered learning on exposure to new environments with and without prior learning in previous environments. When transferring a (simulated) agent from an old to a new environment, the parameters of the perceiver were reset, but the parameters of the critic and actor were retained. Crucially, the AC module depended on a subset of invariant states that were common to all environments considered (e.g., whisker–object touch in the whisking system and agent–wall touch in the locomotion system). This independence of actors from the variable states (e.g., whisker-field and arena sizes) enabled us to examine how the emergence of exploratory behavior generalized to different environments, even when each new environment had to be learned.

The novelty-based withdrawal-like behavior was carried out by the retreat motor primitive, triggered whenever novelty exceeded an adaptive threshold. It planned and executed a retreat to the initial position of that excursion by using the already perceived environment (Moldovan and Abbeel, 2012). The retreat motor primitive selected the next action so as to reduce the distance to the initial position while minimizing novelty.

The appearance of more complex exploration behavior was modeled by a single perceiver and several actor-critic (Bhatnagar et al., 2007) modules (Fig. 1*c*). Each actor started from a uniform random action distribution, thus randomly exploring different features of the environment. Since reward was given proportional to novelty, the first module converged to a guided exploration motor primitive that learned the most novel feature in the agent–environment coupling, the second module to the next most novel feature, etc. As a result, each actor converged to an idiosyncratic exploration primitive that acted to perceive a specific feature in the environment (Fig. 1*c*). This architecture resulted in the emergence of exploration behavior. Thus, our model incorporates both random or stochastic exploration (Daw et al., 2006; Cohen et al., 2007; Frank et al., 2009; Humphries et al., 2012) during development and guided exploration during adulthood via the converged exploration motor primitives.

During on-line active perception, the AC modules explored their respective features by executing their current policies (actors). Engaging specific actors in perception was postulated here to be controlled by a novelty management unit (NMU). The NMU switched between the ACs and a safe retreat policy (Moldovan and Abbeel, 2012). While the former ensured an increase in novelty flow, the latter guaranteed a decrease via a planned retreat to the initial familiar base state. Novelty management was thus based on the following dynamics: (1) the agent first acted according to the first AC by forming a curiosity loop via that AC (Fig. 1*c*, dark NMU arrow); (2) whenever the currently active loop resulted in novelty greater than the average novelty of that loop ($J^{(n)}$), the NMU switched to the retreat policy until the agent safely returned to the base state (red NMU arrows); and (3) if the current loop (loop n) was active for a (loop-dependent adaptive) duration ($T^{(n)}$) without retreat, the NMU switched to the next loop in the sequence (loop $n + 1$; Fig. 1*c*, blue NMU arrow). These transition principles guaranteed that novelty accumulation was bounded due to retreats, yet did not diminish due to advances to higher loops, thus ensuring a "continued bounded novelty flow" (Saig et al., 2012) and the maximization of the novelty signal-to-noise ratio (see below).

The model had only a few free parameters (Gordon et al., 2014), since all thresholds ($J^{(n)}$ and $T^{(n)}$) and learning rates (α) were adaptive and converging. This constituted an extremely robust and general framework for the emergence of exploration behavior in different scenarios, each defined by the perceiver in question. The influence of the parameters of the Bayesian perceiver, namely, the initial prior (P_0) and sensory uncertainty (σ), on the emergent behavior were analyzed for each scenario implemented.

The model assumptions and parameters are summarized as follows: (1) the model assumes that after novelty becomes higher than average, the retreat primitive is activated, yet the actions of this primitive are not predetermined, but rather determined on-line based on perception of the current environment; (2) the model assumes that, after a certain amount of time with novelty lower than average, the next exploration motor primitive is activated, yet the actions taken by each motor primitive are learned; (3) the novelty threshold and the number of time steps before

advancements are adaptive and convergent, thus can acquire any value, representing no assumptions on the timing of motor primitive switching; and (4) the model assumes that the initial actors are uniform distributions over actions, yet the converged motor primitives completely depend on the agent–environment interaction and were not hand coded.

We implemented the model based on the following two basic exploratory modalities of the rodent: the whiskers and locomotion systems. The whiskers system controlled the motion of the whiskers as they interacted with objects in the environment. The locomotion system used walls perceived by the whiskers system to create a map of novel arenas it encounters. We first analyze each system separately and show that a large repertoire of observed behaviors can be qualitatively and quantitatively explained by the model (simulation results of the whisker system, first presented in the study by Gordon et al., 2014, are shown to draw a complete picture of the multimodal power of our model). We further show that the prediction of the model of emergent exploration primitives during development agrees with recent experimental results. We then compare the prediction by the combined model for the behavior of mice during their very first encounter with novel objects using their whiskers.

Whisker dynamics emerge as novelty-managed exploration

In the following simulation results, we had implemented our model with two curiosity loops (Fig. 1*c*, $n = 2$), where the first perceives the whisker self-dynamics (i.e., with no objects present), and the second perceives the location of external objects (Curtis and Kleinfeld, 2009). The sensory information was given by the (discretized) whisker angle (4 bits) and the presence of whisker–object contact (1 bit). The perceiver attempted to predict the next sensory information, given the current information and the action performed (protraction or retraction, 1 bit). The outputs of the AC modules depended only on current and previous contact information, carried by the following four types of mechanoreceptors: whisking (no contact), contact, pressure (continued contact), and detach cells (Szwed et al., 2003). Loop 1 perceived self-motion of the whiskers and was independent of mechanoreceptors signal; its policy could be described by a single whisker protraction probability. Loop 2 perceived objects and depended on whisker–object contact information carried by all four mechanoreceptors; it could thus be described by four protraction probabilities (for more details, see Gordon et al., 2014).

The actors of each of these two loops, described in terms of protraction probability given the current mechanoreceptor activations, converge during development to the following exploration primitives (Fig. 2*a*): loop 1 converges to deterministic protraction (Fig. 2*a*, solid blue line), whereas loop 2 converges to protraction when there is no object present (Fig. 2*a*, solid red line), protraction upon initial contact (Fig. 2*a*, dashed red line), retraction upon activation of the pressure cells (Fig. 2*a*, dotted dashed red line), and equiprobable protraction–retraction upon the activation of detach cells (Fig. 2*a*, dotted red line). This behavior scans the activation of all mechanoreceptor types in the shortest possible duration, with four steps for four types.

Each AC module converged autonomously on its observed behavior. The converged behavior depended solely on the basic principles described above, and thus on only two parameters, sensory uncertainty (σ) and object probability (p_{obj}), where the latter determines how cluttered the environment will be. The robustness of our model is thus demonstrated through the dependence of the converged behavioral policies on these two parameters. The convergence time of loop 1 is $10^3 \alpha t$ and is independent of model parameters (data not shown), whereas that of loop 2 increases with increased σ and decreases as objects become more abundant (increased p_{obj}; Fig. 2*a*, inset). The efficiency of the converged policies, in terms of time and accuracy of perception, is compared with a random behavior (i.e., without prior learning in previous environments), all combined with the following novelty management principle (results are presented as the mean ± SEM; $n = 1100$ simulations): converged loop 1 is slower (267 ± 2 time steps) than random action (123 ± 4 time steps, $p < 0.001$ *t* test) in perceiving the whisker self-dynamics, but is more accurate (0.0073 ± 0.0005 MSE vs 0.2379 ± 0.0005 MSE, $p < 0.001$ *t* test); and converged loop 2 perceives objects faster (80 ± 1 time steps) and more accurately (0.0642 ± 0.0004 MSE) than its random counterpart (127 ± 4 times step, 0.2281 ± 0.0004 MSE, $p < 0.001$ *t* test).

The converged policies form exploration primitives, which, when combined with novelty management, result in natural-like whisking patterns (Fig. 2*b*). In free air, novelty management results in gradually extended whisking (Fig. 2*b*, B1), during which novel whisker angles induce immediate retreats. This behavior can serve as autocalibration of whisker dynamics, which may occur during whisker twitching (Semba et al., 1980; Nicolelis et al., 1995; Fanselow et al., 2001), during which the animal does not move in space and thus reduces the chance for contact with objects. A novel prediction of the model is that if twitching is used for calibration, its amplitude should gradually increase toward its transition to exploratory whisking. The same novelty management mechanism determines that full-amplitude periodic exploratory whisking (Berg and Kleinfeld, 2003; Ahissar and Knutsen, 2008) is controlled by loop 2. When novelty obtained by loop 1 saturates, the protraction of loop 2 induces a full sweep of the whisker field in search of contact, followed by retraction of the retreat policy (Fig. 2*b*, B2). Our model thus predicts that exploratory whisking is controlled by two distinct policies, one for protraction and another for retraction. This prediction, first presented in the study by Gordon et al. (2014), is consistent with recent results suggesting that two distinct brain regions serve as whisking-related central pattern generators, one for rhythmic protraction [the vibrissal zone of the intermediate reticular formation (vIRt)] and one for retraction (the Bötzinger complex; Moore et al., 2013). It is important to emphasize that, based on the assumptions and the adapting parameters of the model, this behavior emerges from the agent–environment interaction (see also Fig. 4*a*) and is not hand coded.

When the whisker touches an object, novelty management induces a gradual increase in palpation complexity, where initial immediate retreats upon contact (Fig. 2*b*, B3, *c*), reminiscent of the rapid cessation of protraction (Mitchinson et al., 2007; Grant et al., 2009), are gradually replaced by exploration of all mechanoreceptor activation through brief protraction–retraction cycles (Fig. 2*b*, B4, *c*), which resemble touch-induced pumps (TIPs; Deutsch et al., 2012). Furthermore, as described by Gordon et al. (2014), our model predicts that whiskers should exhibit "phantom touch" (Fig. 2*b*, B5; i.e., contact-like behavior when objects disappear from the whisker field) due to the fact that the rapid cessation of protraction is not due to contact, but rather is due to novelty.

A novel prediction of our model is that during whisker twitching, here hypothesized to represent the first exploration motor primitive, contact with objects will induce rapid cessation of protraction, which represents the retreat primitive, but not TIPs, which represent the second exploration motor primitive.

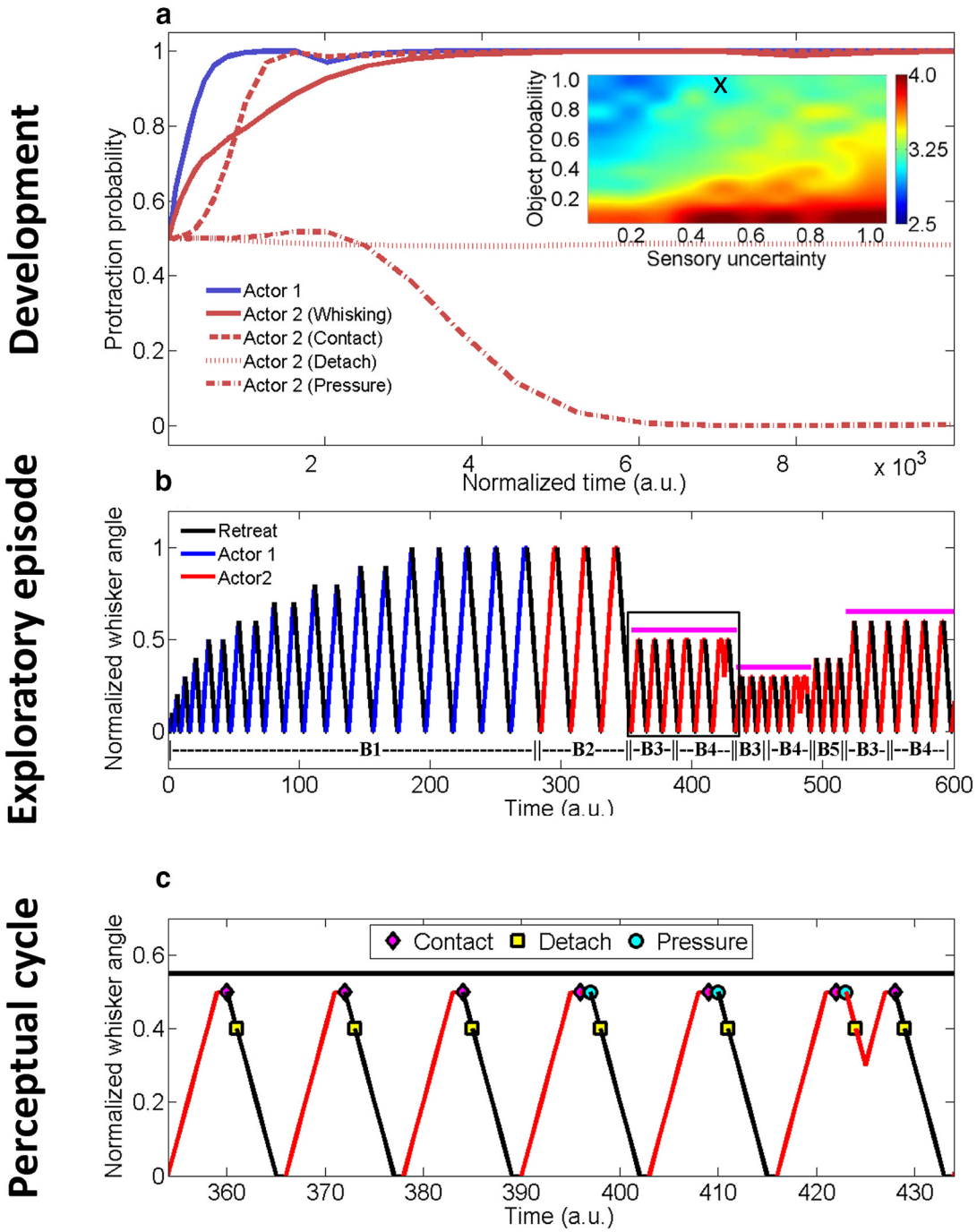

Figure 2. Model implementation on the whisker system across different timescales. ***a***, Adapted from Gordon et al. (2014, their Fig. 5). Developmental convergence dynamics of actors, presented as protraction probability as a function of normalized time (αt, where α is the learning rate), averaged over 10 runs for $\sigma = 0.5, p_{obj} = 1.0$. The protraction probability of the first actor (blue line) does not depend on contact information; the protraction probabilities of the second actor depend on whisking (solid red line), contact (dashed red line), detach (dotted red line), or pressure (dotted dashed red line) inputs. ***a***, Inset, Logarithm (base 10) of normalized convergence time of the second AC module, as a function of σ and p_{obj}. *x* marks parameters for ***a–c***. ***b***, Adapted from Gordon et al. (2014, their Fig. 7). Exploratory episode behavior of the entire converged model; whisker angle is depicted as a function of time, where color denotes the active actor. Magenta horizontal lines denote the angular position of an object. B1, actor 1 protracts the whisker and the retreat primitive retracts the whisker whenever a new angle is reached. B2, initially there are no objects in the whisker field and it protracts, whereupon experiencing no novelty, the NMU switches to the retreat policy (retraction). When objects are present, the initial contact is novel and immediately followed by retreat (B3), whereas the following contacts slowly exhibit the full dynamics of the converged actor 2 (B4). B5, when an object is removed from the whisker field, retreat follows high novelty due to false prediction of its location. ***c***, Perceptual cycle of object location (***b***, enlarged box): protraction upon contact (magenta diamond), retraction upon pressure (cyan circle), and either retraction ($t = 426$) or protraction (data not shown) upon detach (yellow square) mechanoreceptor activation.

Emergence of thigmotaxis via novelty management

In the following simulation results, exploration of a novel circular arena via locomotion (Fonio et al., 2009; Benjamini et al., 2011; Soibam et al., 2012) was modeled using exactly the same mechanism underlying whisker exploration, wherein the animal perceives the location of objects (walls) and maps the boundaries of the arena (see Materials and Methods). The implementation of the locomotion system differed from that of the whiskers system in (1) including four curiosity loops instead of two, (2) referring to a two-dimensional rather than a one-dimensional coordinate

system, and (3) receiving input from the whisker-based perceiver in terms of wall location (left, right, or center) rather than direct mechanoreceptor signals (Fig. 1*b*). The behaviors of all four loops can be represented by the probabilities of each to remain in corners, p_{corner}, follow continuous walls, p_{wall} and explore open space, p_{open}. Six of these 12 variables capture 97% of the variance in locomotion behavior, which was assessed by a principle component analysis (Fig. 3*a*). The convergence dynamics of these six variables show that lower loops (loops 1 and 2) converge before higher ones (loops 3 and 4). This result comes from the agent–environment interaction and the convergence nature of the actors, and is not due to any assumption or prior knowledge (i.e., the agent initially moves randomly, encountering corners, walls, and open space). Since the corner is the most novel feature, the first exploration motor primitive converges to explore it. Then, it moves randomly again during the second level and encounters corners, walls, and open space. Corners are no longer novel, and walls are more novel than open space; thus, the second motor primitive converges to exploring walls, and so on. Furthermore, the order of "feature" presentation depends on the structure of the arena and the behavior of the animal. For example, if an animal goes in a straight line toward the center of the arena, the least novel feature (open space) is encountered first, not last. These results offer a novel prediction of the model, namely, that pups encountering walls for the first time would not follow them as adults do, which would contrast with predictions of theories that suggest that rodents stay near walls to be safe from predators (Prescott et al., 2006).

The converged policies are significantly more efficient than random behavior in mapping the existence of walls (Fig. 3*b*), as measured by the dynamics of the error between the perceived (Fig. 3*b*, insets) and the real arena.

The emerged behavior exhibits complex yet structured dynamics, which is composed of the following three behavioral phases of wall exploration: an initial "circling in place" and the perception of the exit from the home cage (Fig. 3*c*, left); a gradual exploration of the arena walls (Fig. 3*c*, middle); and final exploration of the center of the arena (Fig. 3*c*, right; Fonio et al., 2009; Benjamini et al., 2011). Exploration behavior is interspersed with retreat behavior to the home cage (Fig. 3*c*, black segments), as dictated by novelty management. Notice that there are four actors and not three, where two of them describe wall-following behaviors, one (loop 2) in a clockwise (Fig. 3*c*, middle) and the other (loop 3) in a counterclockwise direction (Fig. 3*c*, top of left panel).

In contrast to the simpler whisker model, the converged locomotion behavior depended dramatically on the two main parameters of the perceiver (Fig. 3*d*), namely, σ, which influences the Bayesian update, and P_0, which is the initial prediction of encountering a wall in any position in space. We performed a clustering (k-means) analysis on the converged policies, represented by the 12 aforementioned variables. Using $k = 3$ clusters, the analysis resulted in a segmentation of the σ–P_0 parameter space into three distinct regions. Examination of the emergent behavior of each region revealed a difference in the transition between behavioral phases: Region 1 exhibits extreme wall attachment and reluctance to leave it (i.e., no transition to the third phase of open-space exploration); Region 2 exhibits the full complexity of behavioral phases, where open-space exploration commences only after completion of a full circle of the arena walls; Region 3 exhibits less wall-attached behavior whereupon incursion into the arena commences before completion of the circumference exploration. The model thus predicts that different animals may exhibit different exploration behaviors, depending on their idiosyncratic manifestation of the model parameters (Fig. 4*d*; Montague et al., 2012). To test this prediction, we reanalyzed the data from the study by Fonio et al. (2009) ($n = 11$ mice) according to the time of appearance of landmark behavior patterns, namely, departure from the home cage, first roundtrip along the walls, first completion of a full circle, first incursion into the arena, and first time the animal reached the center of the arena (Fig. 3*e*). We compared these times for all 11 animals to those emerging from our model and positioned mouse behavior in its best-fitted position in the σ–P_0 phase plane (Fig. 3*d*, A-K). Figure 4*e* depicts the exploration patterns of each mouse–model pair, grouped by their corresponding region in the phase plane. The remarkable match between our model and mouse behavior suggests further testable quantitative predictions, wherein manipulating the sensor uncertainty of an animal (e.g., via perturbation or medication) and its prior whisker contact (e.g., via rearing in a cluttered or completely empty environment) should drastically influence its exploratory behavior.

Exploratory behaviors emerge in a sequence during development

We had also analyzed the developmental timescale of both the whisker and locomotion systems using model simulation. The analysis predicts the gradual and sequential appearance of retraction and then exploratory whisking in the whisker system (Fig. 4*a*). Furthermore, it predicts that whisking amplitude, measured as the maximal protraction angle, has an initial static, low-amplitude period followed by a gradual increase in amplitude as whisking appears (Fig. 4*b*). In the locomotion system, the model predicts that, since loop 1 (i.e., exploring corners) converges first due to the high novelty of corners, and because the optimal behavior to explore corners is to move laterally (as opposed to forward motion, which will take the animal out of the corner), lateral motion will emerge first. Exploring walls and open space converge later due to their lower novelty. Since the optimal behavior to explore walls is forward motion (as opposed to lateral motion, which will make the animal stick to a specific point in the wall), forward motion will emerge later. This prediction is presented in the simulation results in Figure 4*c*. Overall, these predictions show remarkable quantitative agreement with recent experiments reported by Grant et al. (2012) (compare Fig. 4*a–c*, in this article to Figs. 3 (L3), 4 (W1), 5 in Grant et al., 2012). Even though the simulation time was scaled to a real developmental scale, the fit to these experimental observations is both the sequence of appearance of different behaviors, as well as their relative proportions.

Novelty management in whisking and locomotion

The following analysis of experimental data is based on the observation that exploratory behavior is composed of discretized excursions (i.e., individual complex interactions with the environment that start and finish in a known or safe location; Tchernichovski et al., 1998; Fonio et al., 2009; Dvorkin et al., 2010). Building upon the model (Gordon et al., 2014) and simulation results presented above, we quantified perception by a Bayesian predictor of sensory information, which was updated based on newly acquired sensory information due to active interaction with the environment. Novelty flow was quantified per each such interaction as the information gained by the resulting Bayesian update (see Materials and Methods). Novelty was defined as the accumulated novelty flow during an excursion.

To address generic principles of environmental exploration, we analyzed data from two different exploration modalities,

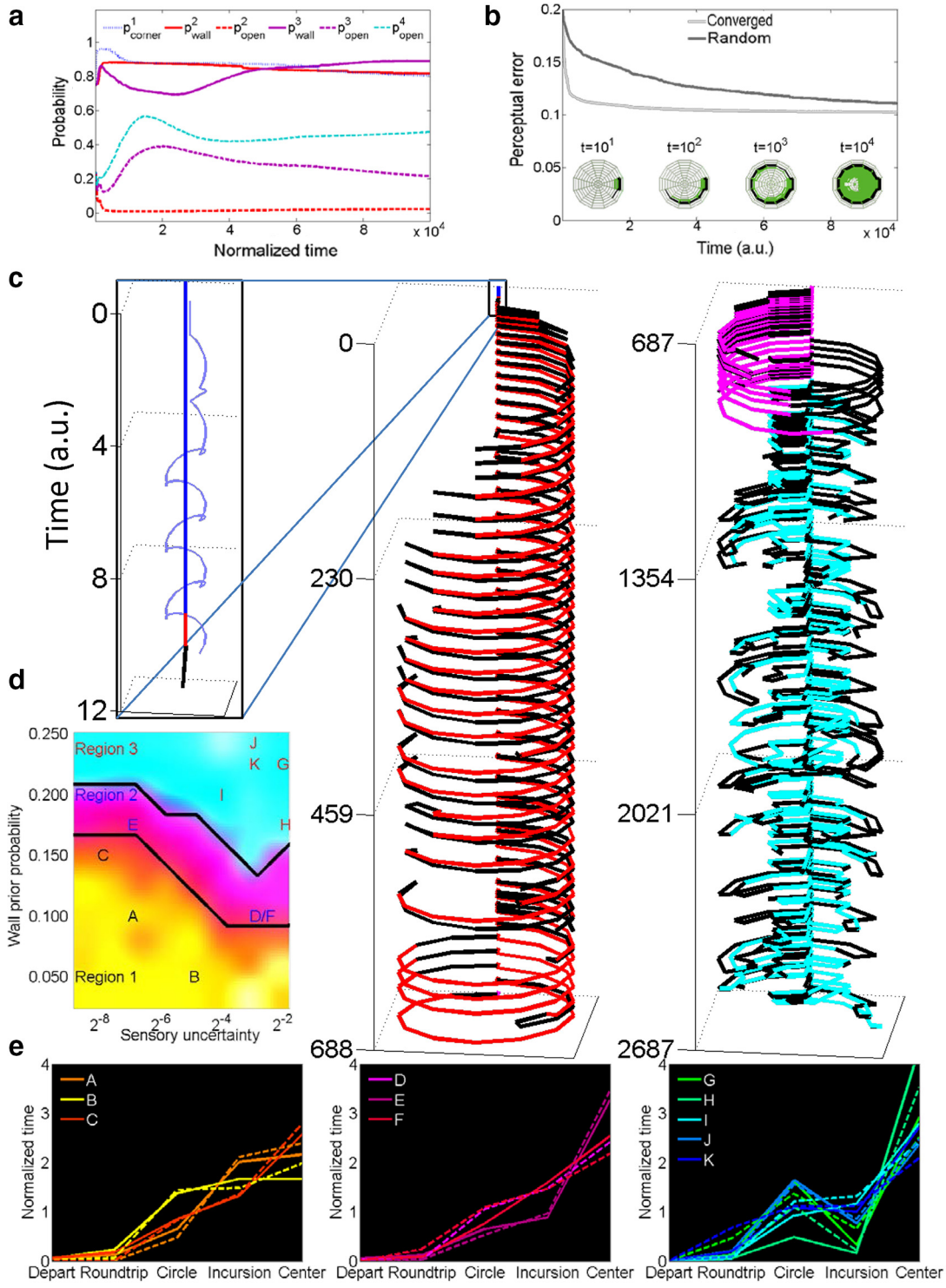

Figure 3. Model implementation on locomotive exploration of a novel circular arena (Fonio et al., 2009). ***a***, Convergence dynamics of actors of the four loops ($p^{l=1,2,3,4}$), where p_{corner} denotes the probability to stay in corners, p_{wall} denotes the probability to follow walls, and p_{open} denotes the probability to avoid walls and seek open space ($\sigma = 0.125$, $P_0 = 0.1$, averaged over 9 runs). ***b***, Perceiver dynamics when exploring with the converged primitives and novelty management. Mean perceiver error as a function of time for the converged and random actors (same parameters as in ***a***). Insets, Perceiver state at different times, where black denotes the probability of walls, green denotes the probability of no wall, and thickness denotes the distance from probability = 0.5. ***c***, Exploration behavior of a novel circular arena for the converged exploration primitives and novelty management, where color denotes the active primitive (black, retreat; blue, loop 1; red, loop 2; magenta, loop 3; cyan, loop 4) and time progresses from top to bottom. Left, Zoom in on the first steps, where the light blue line denotes orientation of the mouse. Middle, Initial exploration in which only loops 1 and 2 are active. Right, Exploration of open space with loops 3 and 4 until reaching the center of the arena. ***d***, Phase plane of model parameters, where regions were automatically discovered via clustering of the actor probabilities. Distances from cluster centroids are plotted as a function of the two free parameters of the model: σ and P_0, where red/green/blue channels denote distance from centroids of clusters 1, 2, and 3. Capital letters (A–K) denote the entire set of mice ($n = 11$) described in Fonio et al. (2009), positioned in the phase plane such that their behavior best correlated with the behavior generated by the model, given the corresponding parameters. ***e***, Transition trajectories of the experimental mice and matching model agents. Durations of exploration in each behavioral phase, normalized by their mean time, are depicted by their occurrence sequence. Dashed curves represent the behavior of individual mice (Fonio et al., 2009), and solid curves represent the behavior of model agents whose parameters are marked in ***d***.

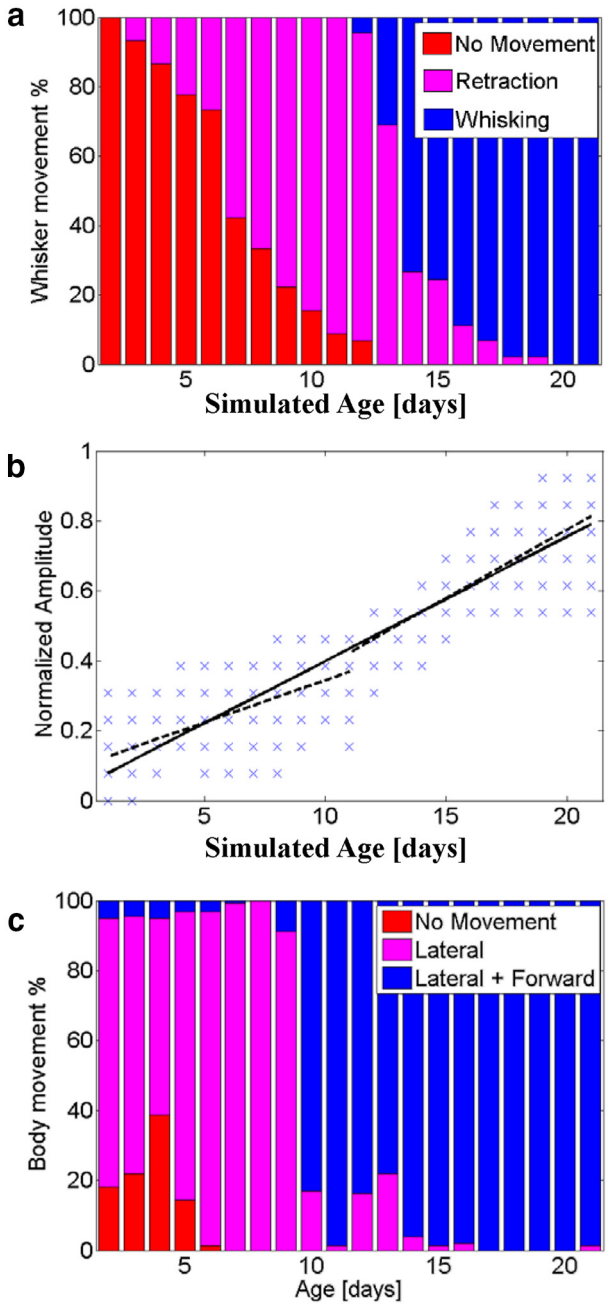

Figure 4. Developmental parameters produced by the model ($\sigma_{whisker} = 0.5$; $p_{obj} = 1$; $\sigma_{locomotion} = 0.125$; $P_0 = 0.05$). Whisker data have 1.6×10^5 time steps, automatically segmented to 941 entries, then grouped to 21 equal-sized bins, corresponding to developmental days. Locomotion data have 1.6×10^6 time steps, automatically segmented to 3375 entries, then grouped to 21 equal-sized bins. ***a***, Appearance of whisker motion patterns (retraction/protraction/whisking), which were calculated only for loop 1 (no objects): no movement, normalized whisker angle <0.25; retraction, retraction from base state; whisking, continued protraction followed by full retraction, with amplitude >0.5 the normalized angle; protraction, otherwise. Protraction was never a result. ***b***, Amplitude of whisker movement, calculated as the maximal normalized angle per entry. Comparison between a single linear fit (solid) and piecewise linear fits (dashed). ***c***, Appearance of locomotion patterns (lateral/forward). Model patterns: No movement, entry duration $<0.6\alpha t$; Forward, forward motion consists $>55\%$ of actions; Lateral, otherwise.

namely, whisking and locomotion. Rodents move their whiskers to perceive their proximal environment, and use locomotion and head movements to explore different spaces. We thus analyzed separately whisker-based exploration in head-fixed animals (rats) and locomotion-based exploration in freely moving animals (mice). The whisking data were obtained from experiments in which head-fixed rats repeatedly explored the space around their snout where a pole may be encountered (Deutsch et al., 2012). In this system, whisks were defined as excursions (Fig. 5*a*), and perception of the environment was defined as the probability distribution over observed angles and contacts, given the previous angle, contact, and the action performed (i.e., the forward model of whisker–pole dynamics). Previous studies in anesthetized rats showed that information regarding whisking and contact with objects is conveyed from the whisker follicle to the brain via three different cell types, namely, whisking, touch, and whisking-touch cells (Szwed et al., 2003). Accordingly, the perception of the environment was updated based on activation of whisking cells (representing whisker angle) and touch cells (representing contact; Szwed et al., 2003; Ahissar and Knutsen, 2008; Knutsen and Ahissar, 2009). Locomotion data were obtained from experiments in which mice exited from a home cage and explored a round arena (Fonio et al., 2009). Here, entries into the arena were defined as excursions (Fig. 5*b*), and perception of the arena was defined as a construction of a probability distribution of walls in the arena. We assumed that walls were more novel than open space (see Materials and Methods).

Excursions were automatically divided into outbound and inbound parts, where the turning point was the first point of maximal novelty flow (Fig. 5*a*,*b*, bottom panels, red crosses). With both whisking and locomotion, novelty flow had a highly dynamic nature between and within excursions (Fig. 5*a*,*b*). In the whisking system, we assumed whiskers were already "calibrated" (i.e., we analyzed starting with full knowledge of whisker angle transitions). In other words, novelty was restricted to whisker–object touch. As can be seen in Figure 5*a*, and predicted by our model, novelty flow peak is not necessarily aligned to the point of maximal protraction. It depends on the history of whisker–object interaction: during the first-touch, the novelty flow peak and maximal angle are aligned, while during later encounters they do not necessarily align, depending on the current and previous angles of contact. Furthermore, some whisking cycles show no novelty flow increase because the object was palpated in angles already explored in previous cycles. Opposed to our model assumptions, the whisking starting point did not remain the same throughout the exploration episode. While a more complex whisker model could possibly account for that, the novelty analysis is based only on whisker–object touch and hence is not affected by this discrepancy. Finally, novelty flow is not necessarily restricted to the outbound exploration part; it can accumulate during retreat, if new perceptions are encountered (see Gordon et al., 2014; Fig. 4). However, our definition of the retreat primitive is that it chooses the path of least novelty; hence, on average novelty during retreat was lower than during exploration.

To test whether exploring rodents choose their actions according to novelty management principles, we compared the novelty quantified in these experiments with the novelty quantified in the following three sets of randomized data (controls, see Materials and Methods): (1) we shuffled the order of excursions, but maintained their internal dynamics (shuffle); (2) we randomized the internal dynamics (actions performed), while maintaining the excursions order (random); and (3) we shuffled the order and randomized the internal dynamics of excursions (shuffle & random). These controls had the same action statistics, but different action order in different timescales (i.e., the shuffle control had a different long-timescale order, whereas the random control had different short-timescale order).

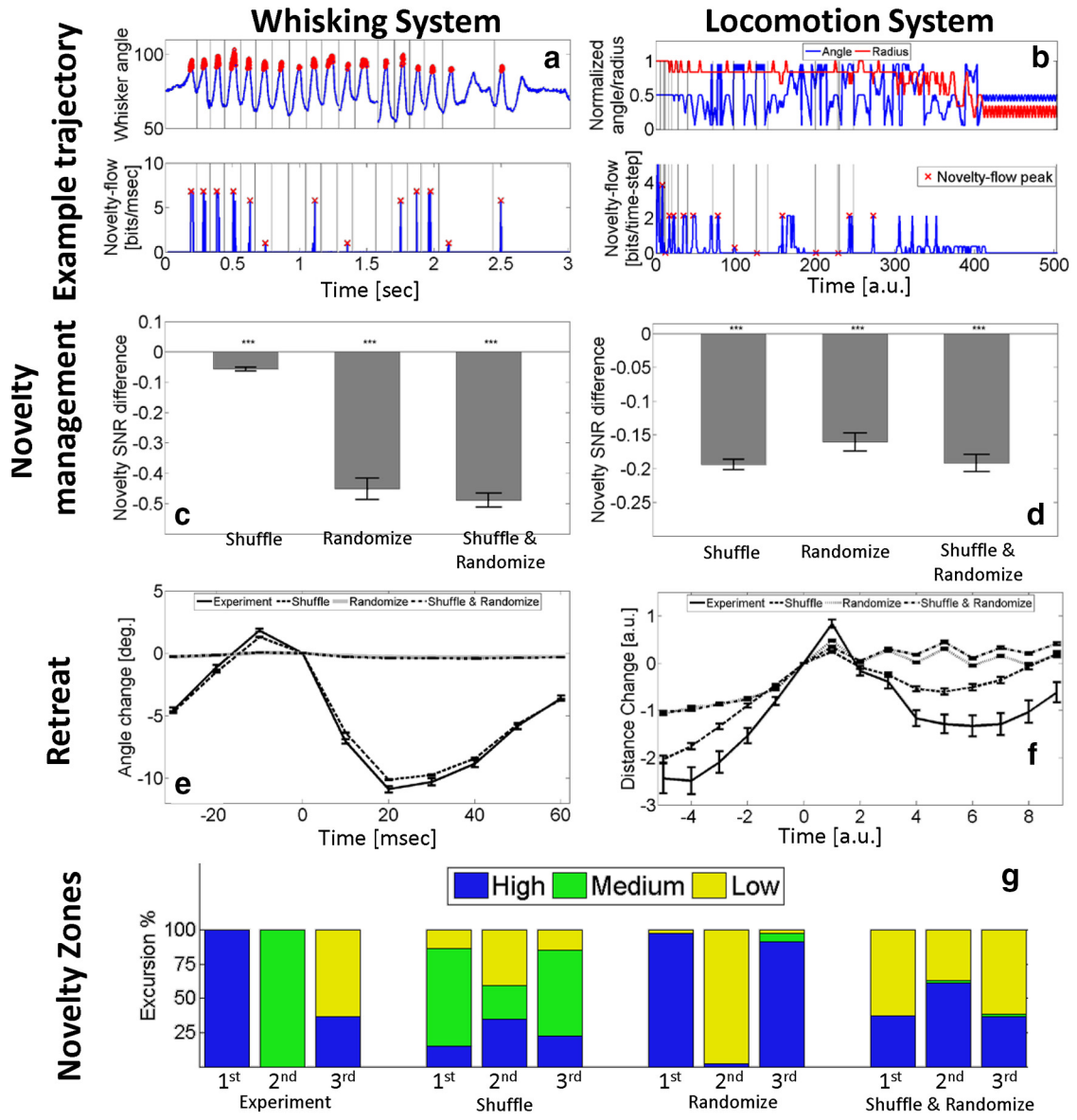

Figure 5. Novelty management principles in the whisker (Deutsch et al., 2012) and locomotion (Fonio et al., 2009) systems. ***a***, Top, Example of whisking trajectory (i.e., angle as a function of time; red circles denote contact with a pole. Bottom, Novelty flow calculated from the trajectory; red crosses denote maximal novelty flow. Vertical lines denote whisk (excursion) beginning. ***b***, Top, Example of locomotion trajectory, described in normalized polar coordinates of angle (blue) and radius (red) in a circular novel arena. Bottom, Novelty flow calculated from the trajectory; red crosses denote maximal novelty flow. Vertical lines denote entry (excursion) beginning. ***c***, ***d***, Difference between the novelty SNR of experimental and control animals in the whisking (***c***) and locomotion (***d***) systems; averaged over sessions (whisking system), animals (locomotion system), and 20 repetitions per session/animal for the controls (see text). Error bars denote SEM, $^{***}p < 0.001$. ***e***, ***f***, Dynamics of inbound movements, time aligned to the last point of maximal novelty flow, where for each data point in each excursion we calculated its spatial distance from the starting point of the excursion (error bars denote SEM). ***e***, Change in angle in the inbound portion. ***f***, Change in Cartesian distance from the home cage in the inbound portion. ***g***, Percentage of excursions according to first (left), second (middle), and third (right) visited novelty zones: the exit from the home cage is the High novelty zone (red); the circumference of the arena is Medium novelty zone (purple); the open space is the Low novelty zone (blue).

Comparing novelty-related measures between the observed trajectory and the control trajectory revealed the following characteristics [results presented as the mean (SEM)]. (1) During an entire exploration episode, the novelty SNR was maximized with respect to the action sequence, i.e. SNR, which is computed as the mean novelty divided by the SD of novelty across excursions, was significantly higher than in simulated controls, which had the same action statistics but different action order [Fig. 5*c*,*d*; whisking: novelty SNR, 0.91 (0.05); control novelty SNR differences: shuffle, −0.06 (0.01); random, −0.45 (0.03); shuffle & random, −0.49 (0.02); locomotion: novelty SNR, 0.63 (0.05); control novelty SNR differences: shuffle, −0.19 (0.01); random, −0.161 (0.014); shuffle & random, −0.19 (0.01); $p < 0.001$, for all controls]. In other words, a full exploration episode (a whisking episode or a full exploration of the arena) had a significantly higher novelty SNR when compared with other possible action ordering in all timescales (within and across excursions). Thus, our analysis did not assume that the animal was omniscient as to the disposition of the environment when planning the exploration, but rather showed, given the actions statistics it did perform, what novelty SNR other action selections could produce. Furthermore, in the whisking system, the random and shuffle & random controls show the trivial result that the whisker simply touched the object much less often. In contrast, in both whisking

and locomotion, the shuffle condition resulted in longer excursions appearing earlier and thus having large changes in novelty flow, which increased the SD and decreased the SNR of novelty. The shuffled condition shows that the order of excursions is very important to maintain high novelty SNR. (2) Novelty was maximized during outbound exploration. The novelty that accumulated during outbound movements was significantly higher than those computed in all of our controls, both for whisking and locomotion [whisking: outbound novelty, 11.99 bits (0.74 bits); control outbound novelty differences: shuffle, −0.69 bits (0.06 bits); random, −8.74 bits (0.20 bits); shuffle & random, −7.31 bits (0.21 bits); locomotion: outbound novelty, 1.13 bits (0.14 bits); control outbound novelty differences: shuffle, −0.32 bits (0.02 bits); random, −0.45 bits (0.03 bits); shuffle & random, −0.50 bits (0.03 bits); $p < 0.001$, for all controls]. (3) Retreat started shortly before or after the time of maximal novelty flow in the whisking and locomotion systems, respectively (Fig. 5*e*,*f*). We believe this discrepancy between model and experimental results (i.e., that withdrawal does not start immediately after maximal novelty flow) is due to discretization of the data in the whisking system and delays in the biological agent in the locomotion system. (4) The environment was explored in a descending order of novelty density. Mice first explored areas in which the probability of hitting a wall was high, and then gradually explored areas with decreasing probability of such exploration (Fig. 5*g*). As the prior probability of encountering a wall during locomotion was generally small, this gradient parallels a gradient of decreasing novelty density. This analysis is unique to this type of environment and task (i.e., a round arena with no external rewards), since it has many types of novelty zones. In other commonly used tasks (e.g., mazes with food), there are fewer types of novelty zones (only walls) and other confounding motivational factors.

Another possible confound for the last point arises from the possibility that the exit area was explored first not because of novelty management but because of distance management, that is, exploration from nearby to distant areas. While this hypothesis cannot be fully rejected, two findings are inconsistent with distance being managed before novelty. First, mice preceded circumference exploration with exploration of the center of the arena, often against the distance order. Second, within the exit area mice managed novelty, as shown by the randomization control (Fig. 5*g*).

These results suggest that our rodents managed their novelty input across and within excursions in an attempt to maximize novelty signal-to-noise ratio, during an entire exploration episode. They explored the environment in a descending order of novelty density starting with the zone of the highest novelty density, maximized novelty accumulation during the outbound phase, and retreated upon encountering high novelty flow.

First whisker touch induces locomotive retreat

We next combined whisking and locomotive explorations into a single cohesive model (Fig. 1*b*). The whisker-based wall perception, performed via novelty-managed whisker exploration, depended on body position and orientation within the arena, and affected locomotion behavior. Locomotive exploration dynamics was based on whisker perception in that the mouse changed its position only after accumulating whisker-dependent novelty above a certain threshold w_{th} (measured in bits of accumulated information gain), which indicated the existence of a wall for the locomotive system. Thus, the first time that the whisker-based novelty exceeded the threshold, the triggered actor was the locomotive retreat primitive, since the locomotive novelty passed its threshold level. The next time that the whisker-based novelty of the same object was above threshold, the locomotion system did not necessarily trigger retreat. Hence, the combined model only assumes how multimodal novelty is connected and not which actions are triggered.

The model thus predicts that novelty induced by whisker contact results in locomotive retreat, and accordingly that whisker contact should show a gradual buildup in duration and extent. We tested these hypotheses by recording and analyzing the whisker and locomotive behavior of 11 mice that encountered a novel object (vertical metal pole) for the first time in their lives (Fig. 6*a*). The integrative whisker–locomotive behavior exhibited a clear manifestation of novelty management, as follows: the initial contact episode, which ended with a retreat from the pole, was significantly ($p < 0.02$, paired t test) shorter than the following contact episode (Fig. 6*b*).

To test the hypothesis that whisker-based object perception affects locomotive exploration via a novelty management principle, we concatenated whisker contacts from consecutive episodes. For each mouse, contacts slowly increased in both duration and extent (single whisker, multiple whiskers, and then nose) as contact events continued (Fig. 6*b*,*c*). Contact duration as a function of contact sequence was fitted with an exponential, $ae^{n/b}$, where n represents contact number, a represents the initial duration, and b represents the contacts constant (i.e., the number of contacts by which contact duration increased by e). Interestingly, the extracted contacts constant ($b = 7.37$ contacts) is comparable to the mean number of contacts before the mouse retreats from the pole (7.54 ± 1.87 contacts). A direct model comparison to these results requires modeling the entire whisker pad, which is beyond the scope of the current model. Nevertheless, this study suggests that mice use a novelty management principle both in whisking and locomotive explorations, in accordance with our generic model.

Predictions related to brain circuitry

Recent advances in understanding whisking control allow specific predictions related to the circuitry of the vibrissal system and its development. An important support to our distinction of protraction and retraction actors comes from the finding that, during whisking, protraction is controlled by a circuit that includes the vIRT, whereas retraction is controlled by a circuit that includes the Bötzinger complex (Moore et al., 2013). According to our model, these two actors operate at different levels; the retraction actor is the default actor (Actor 0; Fig. 1) and the protraction actor (during whisking) is the actor of loop 2 (Actor 2; Fig. 1). Their selection by the NMU should follow different rules and should develop differently. As natural candidates for the implementation of the NMU are the action-selection circuits of the basal ganglia (BG; Redgrave, 2007), our model predicts that the connectivity from the BG to the Bötzinger complex is fixed from birth, while that to the vIRT (or related nuclei) is modified according to vibrissal experience. As Actor 1 (Fig. 1) should also change with experience, the model predicts that the influence of the BG on the facial nucleus, a natural candidate for a twitching-related actor (Herfst and Brecht, 2008; Simony et al., 2010), will also change with vibrissal experience.

A reasonable mapping of the locomotion actors to brain circuits is that the actor of locomotion loop 1 is implemented by the mesencephalic locomotor region—spinal cord circuits that control walking and turning (Musienko et al., 2012; Ryczko and Dubuc, 2013). The effect of BG on these circuits is thus expected to be modified by locomotion experience. In contrast, the effects of BG on a locomotion retreat circuit (implementing Actor 0; Fig. 2)

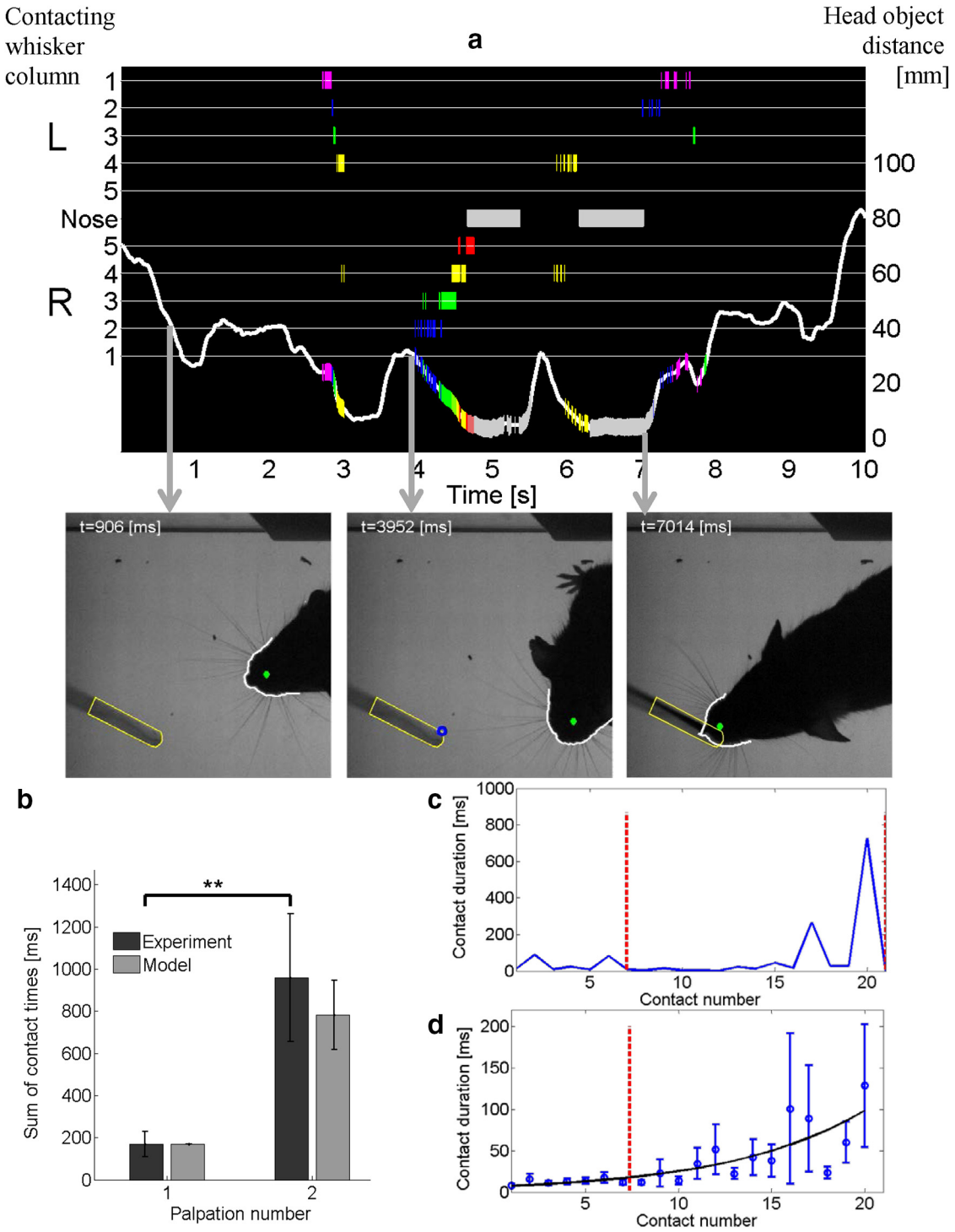

Figure 6. Behavior upon first touch in life with a vertical metal pole during an exploration excursion out of the home cage. ***a***, Top, An example of 3 successive palpations. Numerals and colors denote which whisker column, either on the left (L) or right (R) side, touched the pole, where gray (midline) denotes contact with the nose. The white curve represents the distance of the center of the head from the object as a function of time, overlaid with the contact events displayed above. Bottom, Example images from the recorded films at different times. Snout (white contour) is automatically tracked, whereas the stationary object (green contour) is manually marked. Left, The mouse is distant from the pole and does not touch it. Middle, The mouse is lightly touching the pole (blue circle). Right, The mouse is touching the pole with its nose. ***b***, Comparison of the sum of contact durations between experiment and model during the first and second palpations. Experimental results present the average over 11 mice (error bars denote SEM, $^{**}p < 0.02$), model results averaged over 10 runs with accumulated novelty w_{th} uniformly drawn from the range of [8,9] bits and $\sigma_{whisker} = 0.001$. Model times were normalized by first palpation duration. ***c***, The first sequence of contact durations of a mouse as a function of contact number. The dashed red vertical line denotes the end of the first palpation episode. ***d***, Same as ***c*** averaged over all mice ($n = 11$). Black line denotes exponential fit, $ae^{n/b}$ ($a = 6.54$ ms, $b = 7.37$ contacts), where dashed red vertical line represents b.

should not change. As locomotion retreat must involve navigation capabilities, it can be implemented by circuits that are capable of reverse replay of locomotion (Foster and Wilson, 2006; Diba and Buzsáki, 2007) or path integration (McNaughton et al., 2006).

Finally, our model predicts experience-dependent multimodal integration of novelty via connections from the relevant perceivers to the NMU. Reasonable networks for implementing vibrissal and locomotion perceivers are the thalamocortical net-

work (Curtis and Kleinfeld, 2009; Yu et al., 2013) and the entorhinal–hippocampal network (Zhang et al., 2014), respectively. Thus, connections from these networks to the BG are expected to be modified in a modality-specific, experience-dependent manner. Manipulating these connections should impair multimodal novelty integration, such as that demonstrated here (Fig. 6).

Discussion

Information has previously been suggested as a constant source of "attraction" in exploration, a behavior called "infotaxis" (Vergassola et al., 2007). On the other hand, qualitative analysis of the exploration of an animal has shown that there is a balance between neophilia and neophobia (File, 2001; Elliot, 2006; Hughes, 2007). We incorporate these observations in a minimal model based on an intrinsic motivation (Harlow, 1950; Schmidhuber, 1990; Oudeyer et al., 2007; Singh et al., 2010; Baldassarre, 2011) reinforcement learning scheme in which novelty serves as an intrinsic reward (Gordon et al., 2014). In other words, the novelty disclosed by an action becomes intrinsically rewarding. We modeled exploration via hierarchical curiosity (i.e., novelty rewarded) loops (Gordon and Ahissar, 2012), which concurrently perceive the environment and learn how to act to optimize the perceptual process. Switching between curiosity loops occurs at a higher level, where within-excursion behavior is governed by novelty-seeking exploration followed by the novelty-aversive primitive, which is innate and governed by a retreating (Moldovan and Abbeel, 2012) behavior to the most familiar states. Analytical convergence proof of this model is beyond the scope of this work; thus an analytical derivation of maximization of the novelty signal-to-noise ratio was not addressed. Nevertheless, our model and quantitative analysis suggest that novelty is managed across excursions so as to be high (via efficient exploration motor primitives) and stable (via retreats). This mechanism increases SNR by precluding novelty values that are too small or too large. Retreats might be implemented via mechanisms shared with withdrawal reflexes (Andersen, 2007; Waldenström et al., 2009), in which they are triggered by novelty rather than by external events. Consistent with our model, our results also suggest that in structured environments novelty is further controlled by exploring along a sequence of descending novelty density, from the most novel zone toward the least novel zone.

The concept of rewarding novelty (Schmidhuber, 1990; Redgrave and Gurney, 2006) is widely used in developmental robotics (Weng, 2004) and has also been addressed recently by several information-theoretic approaches that attempt to either maximize (Polani, 2009; Tishby and Polani, 2011) or minimize (Friston, 2010) novelty. Our new quantitative findings suggest that curious animals do not attempt to maximize or minimize novelty, but rather maintain a constant flow of novelty by switching between behaviors that increase or reduce it. Our model, which has only a few free parameters, is based on novelty management (Gordon et al., 2014) and suggests that a single neophobic primitive governs the reduction of novelty flow, whereas neophilia is mediated by a hierarchy of exploration motor primitives; each of these primitives converges, via the same basic intrinsic reward mechanism, to an efficient exploratory behavior of a specific feature of the environment. The hierarchical structure, based on neuroanatomical evidence of the mammalian brain (Kleinfeld et al., 2006), enables the ontological timescale emergence of efficient exploration of complex environments. Furthermore, it can augment current implementation of neuro-inspired exploration strategies of biomimetic robots (Prescott et al., 2006; Caluwaerts et al., 2012; Pearson et al., 2013).

Our model uses similar information-theoretic principles as those used in previous approaches (Tishby and Polani, 2011; Still and Precup, 2012), yet in a different setting that enables sequential learning of convergent exploration primitives. Convergence, in our approach, occurs due to interactions with different instances of the same environment type. In contrast to previous models (Aston-Jones and Cohen, 2005; Der and Martius, 2012), which exhibit a single nonstationary policy, our model results in curiosity loops converging to different exploration primitives, each perceiving a different set of features of the environment. We have shown that combining curiosity-driven emergent behaviors, multiple-loop architecture, and novelty management results in a satisfactory fit to experimental behavioral data in different modalities. The model makes several new testable predictions in both whisker and locomotion domains and can be easily extended to incorporate richer dynamics. While we presented evidence of novelty management in the behaviors of rodents, similar principles were recently shown to occur in purely perceptual tasks with human subjects, wherein optimal control of perceptual novelty explained a whisker-based localization behavior (Saig et al., 2012), thus suggesting that these principles are ubiquitous in the mammalian kingdom.

The framework predicts novel neural circuitry during development. To facilitate rewarding information gain, there should be a strong input connectivity to the rewarding system from internal model areas [e.g., cerebellum (Lalazar and Vaadia, 2008; Shadmehr and Krakauer, 2008) and sensory perception areas, such as primary sensory cortices (Matyas et al., 2010; Bastos et al., 2012; Feldmeyer et al., 2013)]. The framework predicts that this connectivity should be stronger during development to allow convergence of the stereotypical exploratory behaviors apparent in adult rats. Furthermore, the information conveyed in these connections should code prediction error signals. The anatomical and functional circuitry of the developing pup is mostly unknown, yet the underlying infrastructure for the proposed curiosity loops should be evident to corroborate the proposed framework.

Within each curiosity loop there are several internal variables that play a critical role in the framework. The first is the average reward, which determines the novelty threshold of that loop (i.e., novelty greater than it instigates retreat). It was suggested that average reward is related to opportunity costs and latency between action switching (Niv et al., 2007; Cools et al., 2011). Furthermore, average reward was suggested to be related to tonic dopamine levels in nucleus accumbens. Our model predicts that this parameter converges with time, with respect to the exploration motor primitive. Novelty-aversive behavior is suggested to be tightly connected to fear-related regions due to the hypothesis that it relates to anxiety (Misslin and Cigrang, 1986; Fonio et al., 2009). However, according to our formal framework, it should be related to perceptual learning regions (e.g., place cells in the hippocampus in an arena exploration scenario), since they must be accessible when determining the next action. In other words, novelty-aversive behavior is similar to goal-directed policy, whose goal is to return to a known safe state. This requires information that has already been learned about the environment and is to be contrasted with other fear-related behavior such as freezing, which does not. The novelty management unit suggests a centralized region that receives novelty as an input and switches between the novelty-seeking and novelty-aversive behaviors, which is reminiscent of an action selection mechanism. Hence, the basal ganglia are good candidates for the location for novelty management processing (Redgrave, 2007).

The qualitative and quantitative results presented here suggest that our novel model may explain exploratory behavior in more general settings. Along the time axis, we have shown that the model explains behavior both during development and in adulthood, thus covering almost the entire span of rodent life. Along the complexity axis, the hierarchical curiosity loop architecture enables the emergence of exploratory behavior of increasingly complex features of the environment. For example, whisker exploration can be extended beyond self-motion and object localization to object compliance and texture. Along the modality axis, the model can be extended beyond the two modalities explored here to include other modalities, such as vision and olfaction. The model predicts that novelty is coordinated across modalities. We demonstrate here, using a novel "first-touch-in-life" experiment, that this is indeed the case between the whisking and locomotion modalities. The model allows the hierarchical cascading of exploratory behaviors, enabling an open-ended construction of complex multimodal exploration. We thus propose that the model and its underlying principles are ubiquitous and can account for exploratory behaviors of animals and humans; infants and adults.

References

Ahissar E, Knutsen PM (2008) Object localization with whiskers. Biol Cybern 98:449–458. CrossRef Medline

Andersen OK (2007) Studies of the organization of the human nociceptive withdrawal reflex. Focus on sensory convergence and stimulation site dependency. Acta Physiol (Oxf) 189:1–35. CrossRef Medline

Aston-Jones G, Cohen JD (2005) An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci 28:403–450. CrossRef Medline

Bahar A, Dudai Y, Ahissar E (2004) Neural signature of taste familiarity in the gustatory cortex of the freely behaving rat. J Neurophysiol 92:3298–3308. CrossRef Medline

Baldassarre G (2011) What are intrinsic motivations? A biological perspective. Paper presented at IEEE International Conference on Development and Learning (ICDL), Frankfurt am Main, Germany, August.

Barnett SA (1958) Exploratory behaviour. Br J Psychol 49:289–310. CrossRef Medline

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76:695–711. CrossRef Medline

Benjamini Y, Fonio E, Galili T, Havkin GZ, Golani I (2011) Quantification of behavior Sackler Colloquium: quantifying the buildup in extent and complexity of free exploration in mice. Proc Natl Acad Sci U S A 108:15580–15587. CrossRef Medline

Berg RW, Kleinfeld D (2003) Rhythmic whisking by rat: retraction as well as protraction of the vibrissae is under active muscular control. J Neurophysiol 89:104–117. CrossRef Medline

Berger-Tal O, Nathan J, Meron E, Saltz D (2014) The exploration-exploitation dilemma: a multidisciplinary framework. PLoS One 9:e95693. CrossRef Medline

Berlyne DE (1960) Conflict, arousal, and curiosity. New York: McGraw-Hill.

Bhatnagar S, Sutton R, Ghavamzadeh M, Lee M (2007) Incremental natural actor-critic algorithms. Paper presented at Twenty-First Annual Conference on Advances in Neural Information Processing Systems, Vancouver, BC, Canada, December.

Caluwaerts K, Staffa M, N'Guyen S, Grand C, Dollé L, Favre-Félix A, Girard B, Khamassi M (2012) A biologically inspired meta-control navigation system for the psikharpax rat robot. Bioinspir Biomim 7:025009. CrossRef Medline

Cohen JD, McClure SM, Yu AJ (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos Trans R Soc Lond B Biol Sci 362:933–942. CrossRef Medline

Cools R, Nakamura K, Daw ND (2011) Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology 36:98–113. CrossRef Medline

Curtis JC, Kleinfeld D (2009) Phase-to-rate transformations encode touch in cortical neurons of a scanning sensorimotor system. Nat Neurosci 12:492–501. CrossRef Medline

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876–879. CrossRef Medline

Der R, Martius G (2012) The playful machine. New York: Springer.

Deutsch D, Pietr M, Knutsen PM, Ahissar E, Schneidman E (2012) Fast feedback in active sensing: touch-induced changes to whisker-object interaction. PLoS One 7:e44272. CrossRef Medline

Diamond ME, von Heimendahl M, Knutsen PM, Kleinfeld D, Ahissar E (2008) "Where" and "what" in the whisker sensorimotor system. Nat Rev Neurosci 9:601–612. CrossRef Medline

Diba K, Buzsáki G (2007) Forward and reverse hippocampal place-cell sequences during ripples. Nat Neurosci 10:1241–1242. CrossRef Medline

Dvorkin A, Szechtman H, Golani I (2010) Knots: attractive places with high path tortuosity in mouse open field exploration. PLoS Comput Biol 6:e1000638. CrossRef Medline

Elliot AJ (2006) The hierarchical model of approach-avoidance motivation. Motiv Emot 30:111–116. CrossRef

Fanselow EE, Sameshima K, Baccala LA, Nicolelis MA (2001) Thalamic bursting in rats during different awake behavioral states. Proc Natl Acad Sci U S A 98:15330–15335. CrossRef Medline

Feldmeyer D, Brecht M, Helmchen F, Petersen CC, Poulet JF, Staiger JF, Luhmann HJ, Schwarz C (2013) Barrel cortex function. Prog Neurobiol 103:3–27. CrossRef Medline

File SE (2001) Factors controlling measures of anxiety and responses to novelty in the mouse. Behav Brain Res 125:151–157. CrossRef Medline

Fonio E, Benjamini Y, Golani I (2009) Freedom of movement and the stability of its unfolding in free exploration of mice. Proc Natl Acad Sci U S A 106:21335–21340. CrossRef Medline

Foster DJ, Wilson MA (2006) Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature 440:680–683. CrossRef Medline

Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat Neurosci 12:1062–1068. CrossRef Medline

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138. CrossRef Medline

Gao P, Bermejo R, Zeigler HP (2001) Whisker deafferentation and rodent whisking patterns: behavioral evidence for a central pattern generator. J Neurosci 21:5374–5380. Medline

Gordon G, Ahissar E (2012) Hierarchical curiosity loops and active sensing. Neural Netw 32:119–129. CrossRef Medline

Gordon G, Fonio E, Ahissar E (2014) Learning and control of exploration primitives. J Comput Neurosci. Advance online publication. Retrieved August 7, 2014. doi:10.1007/s10827-014-0500-1. CrossRef Medline

Grant RA, Mitchinson B, Fox CW, Prescott TJ (2009) Active touch sensing in the rat: anticipatory and regulatory control of whisker movements during surface exploration. J Neurophysiol 101:862–874. CrossRef Medline

Grant RA, Mitchinson B, Prescott TJ (2012) The development of whisker control in rats in relation to locomotion. Dev Psychobiol 54:151–168. CrossRef Medline

Harish O, Golomb D (2010) Control of the firing patterns of vibrissa motoneurons by modulatory and phasic synaptic inputs: a modeling study. J Neurophysiol 103:2684–2699. CrossRef Medline

Harlow HF (1950) Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. J Comp Physiol Psychol 43:289–294. CrossRef Medline

Herfst LJ, Brecht M (2008) Whisker movements evoked by stimulation of single motor neurons in the facial nucleus of the rat. J Neurophysiol 99:2821–2832. CrossRef Medline

Hughes RN (2007) Neotic preferences in laboratory rodents: issues, assessment and substrates. Neurosci Biobehav Rev 31:441–464. CrossRef Medline

Humphries MD, Khamassi M, Gurney K (2012) Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. Front Neurosci 6:9. CrossRef Medline

Kaplan F, Oudeyer PY (2007) In search of the neural circuits of intrinsic motivation. Front Neurosci 1:225–236. CrossRef Medline

Kawato M (1999) Internal models for motor control and trajectory planning. Curr Opin Neurobiol 9:718–727. CrossRef Medline

Kleinfeld D, Ahissar E, Diamond ME (2006) Active sensation: insights from the rodent vibrissa sensorimotor system. Curr Opin Neurobiol 16:435–444. CrossRef Medline

Knutsen PM, Ahissar E (2009) Orthogonal coding of object location. Trends Neurosci 32:101–109. CrossRef Medline

Knutsen PM, Pietr M, Ahissar E (2006) Haptic object localization in the vibrissal system: behavior and performance. J Neurosci 26:8451–8464. CrossRef Medline

Lalazar H, Vaadia E (2008) Neural basis of sensorimotor learning: modifying internal models. Curr Opin Neurobiol 18:573–581. CrossRef Medline

Little DY, Sommer FT (2013) Learning and exploration in action-perception loops. Front Neural Circuits 7:37. CrossRef Medline

Matyas F, Sreenivasan V, Marbach F, Wacongne C, Barsy B, Mateo C, Aronoff R, Petersen CC (2010) Motor control by sensory cortex. Science 330:1240–1243. CrossRef Medline

McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser MB (2006) Path integration and the neural basis of the "cognitive map." Nat Rev Neurosci 7:663–678. CrossRef Medline

Misslin R, Cigrang M (1986) Does neophobia necessarily imply fear or anxiety? Behav Processes 12:45–50. CrossRef Medline

Mitchinson B, Martin CJ, Grant RA, Prescott TJ (2007) Feedback control in active sensing: rat exploratory whisking is modulated by environmental contact. Proc Biol Sci 274:1035–1041. CrossRef Medline

Moldovan TM, Abbeel P (2012) Safe exploration in Markov decision processes. Paper presented at International Conference on Machine Learning (ICML). Edinburgh, Scotland, UK, June.

Montague PR, Dolan RJ, Friston KJ, Dayan P (2012) Computational psychiatry. Trends Cogn Sci 16:72–80. CrossRef Medline

Moore JD, Deschênes M, Furuta T, Huber D, Smear MC, Demers M, Kleinfeld D (2013) Hierarchy of orofacial rhythms revealed through whisking and breathing. Nature 497:205–210. CrossRef Medline

Musienko PE, Zelenin PV, Lyalka VF, Gerasimenko YP, Orlovsky GN, Deliagina TG (2012) Spinal and supraspinal control of the direction of stepping during locomotion. J Neurosci 32:17442–17453. CrossRef Medline

Nicolelis MA, Baccala LA, Lin RC, Chapin JK (1995) Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. Science 268:1353–1358. CrossRef Medline

Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology 191:507–520. CrossRef Medline

Oudeyer PY, Kaplan F, Hafner VV (2007) Intrinsic motivation systems for autonomous mental development. IEEE Trans Evolutionary Comput 11:265–286. CrossRef

Pearson MJ, Fox C, Sullivan JC, Prescott TJ, Pipe T, Mitchinson B (2013) Simultaneous localisation and mapping on a multi-degree of freedom biomimetic whiskered robot. In: ICRA '13 IEEE, pp 586–592. New York: Institute of Electrical and Electronics Engineers.

Perkon I, Kosir A, Itskov PM, Tasic J, Diamond ME (2011) Unsupervised quantification of whisking and head movement in freely moving rodents. J Neurophysiol 105:1950–1962. CrossRef Medline

Polani D (2009) Information: currency of life? HFSP J 3:307–316. CrossRef Medline

Prescott TJ, Montes González FM, Gurney K, Humphries MD, Redgrave P (2006) A robot model of the basal ganglia: behavior and intrinsic processing. Neural Netw 19:31–61. CrossRef Medline

Redgrave P (2007) Basal ganglia. Scholarpedia J 2:1825. CrossRef

Redgrave P, Gurney K (2006) The short-latency dopamine signal: a role in discovering novel actions? Nat Rev Neurosci 7:967–975. CrossRef Medline

Ryczko D, Dubuc R (2013) The multifunctional mesencephalic locomotor region. Curr Pharm Des 19:4448–4470. CrossRef Medline

Saig A, Gordon G, Assa E, Arieli A, Ahissar E (2012) Motor-sensory confluence in tactile perception. J Neurosci 32:14022–14032. CrossRef Medline

Schmidhuber J (1990) A possibility for implementing curiosity and boredom in model-building neural controllers. In: From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior (Complex Adaptive Systems), pp 222–227. Paris: MIT.

Semba K, Szechtman H, Komisaruk BR (1980) Synchrony among rhythmical facial tremor, neocortical 'alpha' waves, and thalamic non-sensory neuronal bursts in intact awake rats. Brain Res 195:281–298. Medline

Shadmehr R, Krakauer JW (2008) A computational neuroanatomy for motor control. Exp Brain Res 185:359–381. CrossRef Medline

Simony E, Bagdasarian K, Herfst L, Brecht M, Ahissar E, Golomb D (2010) Temporal and spatial characteristics of vibrissa responses to motor commands. J Neurosci 30:8935–8952. CrossRef Medline

Simpkins A, Todorov E, de Callafon R (2008) Optimal trade-off between exploration and exploitation. Paper presented at the American Control Conference, 2008, Seattle, WA, June.

Singh S, Lewis RL, Barto AG, Sorg J (2010) Intrinsically motivated reinforcement learning: an evolutionary perspective. IEEE Trans Auton Ment Dev 2:70–82. CrossRef

Soibam B, Goldfeder RL, Manson-Bishop C, Gamblin R, Pletcher SD, Shah S, Gunaratne GH, Roman GW (2012) Modeling *Drosophila* positional preferences in open field arenas with directional persistence and wall attraction. PLoS One 7:e46570. CrossRef Medline

Still S (2009) Information-theoretic approach to interactive learning. Europhys Lett 85:28005. CrossRef

Still S, Precup D (2012) An information-theoretic approach to curiosity-driven reinforcement learning. Theory Biosci 131:139–148. CrossRef Medline

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.

Szwed M, Bagdasarian K, Ahissar E (2003) Encoding of vibrissal active touch. Neuron 40:621–630. CrossRef Medline

Tchernichovski O, Benjamini Y, Golani I (1998) The dynamics of long-term exploration in the rat. Part I. A phase-plane analysis of the relationship between location and velocity. Biol Cybern 78:423–432. CrossRef Medline

Tishby N, Polani D (2011) Information theory of decisions and actions. In: Perception-action cycle, pp 601–636. New York: Springer.

Towal RB, Hartmann MJ (2008) Variability in velocity profiles during free-air whisking behavior of unrestrained rats. J Neurophysiol 100:740–752. CrossRef Medline

Vergassola M, Villermaux E, Shraiman BI (2007) 'Infotaxis' as a strategy for searching without gradients. Nature 445:406–409. CrossRef Medline

Waldenström A, Christensson M, Schouenborg J (2009) Spontaneous movements: effect of denervation and relation to the adaptation of nociceptive withdrawal reflexes in the rat. Physiol Behav 98:532–536. CrossRef Medline

Welker WI (1964) Analysis of sniffing of the albino rat. Behaviour 22:223–244. CrossRef

Weng J (2004) Developmental robotics: theory and experiments. Int J Humanoid Robotics 1:199–236. CrossRef

Yu C, Horev G, Rubin N, Derdikman D, Haidarliu S, Ahissar E (2013) Coding of object location in the vibrissal thalamocortical system. Cereb Cortex. Advance online publication. Retrieved August 7, 2014. doi:10.1093/cercor/bht241. CrossRef Medline

Zhang SJ, Ye J, Couey JJ, Witter M, Moser EI, Moser MB (2014) Functional connectivity of the entorhinal–hippocampal space circuit. Philos Trans R Soc Lond B Biol Sci 369:20120516. CrossRef Medline