Egocentric value maps of the near-body environment

⊕ Check for updates

Rory John Bufacchi ⓘ [1,2] ✉, Richard Somervail[2,3], Aoife Maria Fitzpatrick ⓘ [2,4], Yusuke Murayama ⓘ [1], Nikos Logothetis[1], Roberto Caminiti ⓘ [2] & Gian Domenico Iannetti ⓘ [2,3,5] ✉

Body-part-centered response fields are pervasive in single neurons, functional magnetic resonance imaging, electroencephalography and behavior, but there is no unifying formal explanation of their origins and role. In the present study, we used reinforcement learning and artificial neural networks to demonstrate that body-part-centered fields do not simply reflect stimulus configuration, but rather action value: they naturally arise from the basic assumption that agents often experience positive or negative reward after contacting environmental objects. This perspective successfully reproduces experimental findings that are foundational in the peripersonal space literature. It also suggests that peripersonal fields provide building blocks that create a modular model of the world near the agent: an egocentric value map. This concept is strongly supported by the emergent modularity that we observed in our artificial networks. The short-term, close-range, egocentric map is analogous to the long-term, long-range, allocentric hippocampal map. This perspective fits empirical data from multiple experiments, provides testable predictions and accommodates existing explanations of peripersonal fields.

Many neural and behavioral responses have a spatial component. Among the most illuminating examples is the activity of place and grid cells. These cells, located in the hippocampus and entorhinal cortex, explicitly encode an animal's allocentric position within its environment[1].

Our understanding of these cells has been substantially enriched by attempts to formally describe their activity patterns[1,2]. An example of particular relevance to this work comes from the exchange of ideas between neuroscience and reinforcement learning[2], which suggested that allocentric place and grid responses—responses with a reference frame anchored to the environment rather than to the individual—are not purely spatial. Rather, they might provide a more general description of the world in terms of its future states[3], to facilitate interactions with the environment at the scale of locomotion and navigation to distant objects[2]. It follows that allocentric spatial responses might provide an effective substrate for navigation in not only physical, but also abstract, mental environments, even when facing new tasks[3,4]. Empirical work has confirmed these predictions in humans and other animals[5]. Clearly, the substantial efforts to formally understand the computational mechanisms underlying allocentric responses have been enormously valuable.

In contrast, certain egocentric responses have been less studied, both empirically and theoretically. Anchored to body parts, and dependent on the spatial proximity between stimuli and those body parts, these measures form peripersonal response fields[6–8]. Peripersonal fields are recorded from macaque single neurons[7,9], human and nonhuman functional magnetic resonance imaging (fMRI)[10], electroencephalography (EEG)[11,12] and behavioral responses[13,14]. This prevalence

across scales and measures makes them excellent candidates for studying system-level neuroscience. Nevertheless, formal theories of the function and origin of peripersonal fields remain underdeveloped[15].

Current models suggest that peripersonal fields reflect diverse and sometimes vaguely defined concepts such as the sense of self, future self-location, impact prediction and multisensory representation of space[8,15,16]. The few attempts to formalize peripersonal fields have largely focused on perceptual aspects[17–21], leaving several features unexplained. For example, existing models often do not take into account the effects of environmental dynamics, motor repertoire and stimulus valence. A precise and unifying mathematical formalism reflecting a complete functional understanding of peripersonal fields is lacking.

In the present study, we demonstrate, both analytically and in silico, how a reinforcement learning perspective can explain the origin and function of peripersonal fields. These fields naturally emerge from action value, under the simple assumption that agents often experience positive or negative reward after contact with objects. The resulting action values—which can be thought of as the utility of potential motor plans—form body-part-centered fields that resemble biological peripersonal responses.

We further demonstrate that body-part-centered value fields can be the building blocks for an egocentric map of the world near the agent in terms of its short-term future states. This egocentric value-based mapping, a generalization of the concept of a successor representation, is analogous to the function of place cells, which provide an allocentric map of the more distant world for longer-term locomotion and navigation. The activity of a particular group of place cells suggests a trajectory of the agent through allocentric space[1], whereas activity of a group of peripersonal neurons would suggest a trajectory of an object through egocentric space, coded in terms of the actions that the object affords[22]. We show that such an egocentric value map offers a statistically robust fit for a wide spectrum of empirical data from multiple research groups and subsumes existing interpretations and formal models of peripersonal fields. Finally, we demonstrate that even simple reinforcement learning agents could successfully adopt egocentric maps, by capitalizing on the natural tendency of artificial neural networks (ANNs) to form task-specialized subnetworks.

Results

Body-part-centered fields emerge from rewarded contact
Animals navigating the world often experience reward or punishment as a consequence of their actions. Touching a hot stove hurts and grasping an apple is necessary for the pleasure of eating it. Thus, making or avoiding contact with objects often carries value (Fig. 1a and Extended Data Fig. 1). We contend that this simple fact fully accounts for the emergence of receptive fields centered around body parts. In this section, we demonstrate this possibility theoretically under a reinforcement learning framework (Supplementary Information 1 and 2) and then recreate canonical characteristics of peripersonal fields in ANNs.
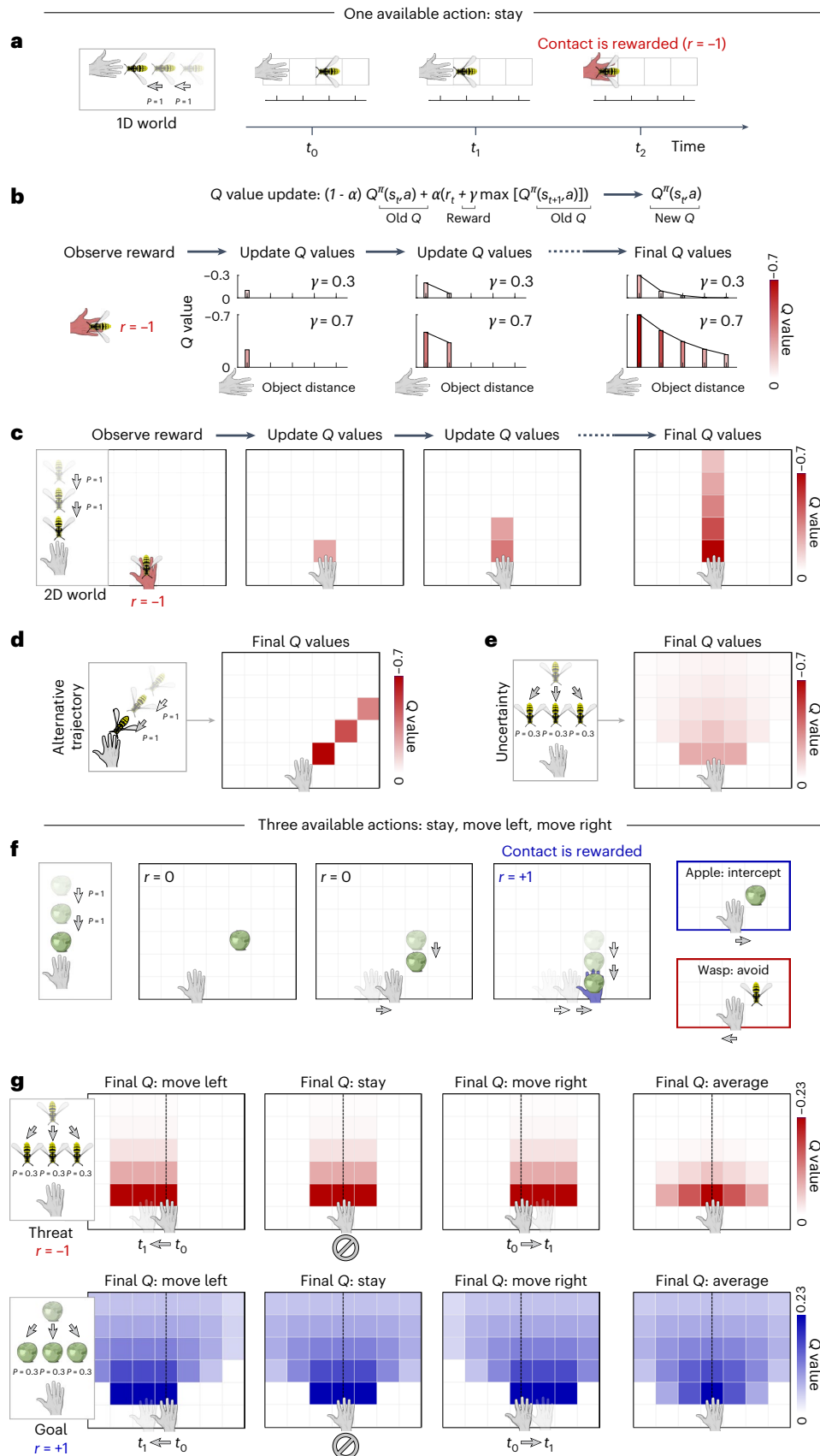
In a reinforcement learning model of an agent interacting with its environment, the values (Q) of actions that bring a body part into contact with certain objects correlate with the object's proximity to that body part. Plotting those Q values as a function of object position reveals graded fields surrounding the agent's limbs (Fig. 1b, Extended Data Fig. 1 and Methods, 'Tabular reinforcement learning for calculation of action values Q'). Three key factors create this correlation between action value and object proximity. Each of these factors can independently create body-part-centered fields, but all three co-occur in a reinforcement learning environment:

(1) Proximity in time. The temporal discount factor γ reduces the value of states n time steps away from object–body contact by γ^n (Fig. 1b; $\max_a |Q^\pi(s_t(|d_t| > 0), a)| < \max_a |Q^\pi(s_t(|d_t| = 0), a)|$, where a denotes an action and $|d_t|$ is the absolute distance between the body part and an object at time t). Given that object movement requires time, that is, object distance d_t is related to the number of time steps n before contact, the discount factor γ decreases the values of actions related to distant objects by γ^n. For example, immediate action is required if a wasp is near your hand to avoid the 'negative reward' of a sting, but the wasp can safely be ignored if it is meters away. In the simplest scenario, if the wasp—or any other object—always moves toward the hand (that is, $P(|d_{t+1}| < |d_t|) = 1$; Fig. 1a), action values are monotonically related to the distance between the object and the hand d_t (Fig. 1b), forming a body-part-centered receptive field.

(2) World dynamics. In a biologically more plausible case, objects do not always approach the body. For example, the wasp might follow some deterministic trajectory, potentially bypassing the hand. In this case, the probability of contact, and hence the value of a state, would depend on both the wasp's location $\mathbf{d}$ and its velocity $\dot{\mathbf{d}}$ ($P(|d_{t+1}| < |d_t|) = f(\mathbf{d}, \dot{\mathbf{d}})$ and $0 \leq f(\mathbf{d}, \dot{\mathbf{d}}) \leq 1$). Such dynamics create an asymmetrical fall-off of action value around the hand, because certain positions will not result in contact (Fig. 1c,d). The wasp might also move stochastically, introducing uncertainty in the environmental dynamics. In this case, contact is not guaranteed at any point in time and more time will elapse until contact is made. Therefore, the negative reward due to contact is weighted less strongly and action value will fall off more strongly as a function of distance to an object d_t along the direction in which it is traveling (Fig. 1e and Supplementary Information 2). Conversely, uncertainty about the wasp's movement and position also increases the number of states from which the wasp can contact the body, and hence laterally expands the value field (Fig. 1e). Perceptual uncertainties, such as impaired vision in a dimly lit room, similarly impact action values by reducing the accuracy of position estimates $\hat{\mathbf{d}}$ (henceforth, we will refer to perceptual and environmental uncertainties collectively as simply 'uncertainty'). World dynamics therefore also shape body-part-centered fields[20], yielding maximal action

---

Fig. 1 | Rewarded contact explains the origin and properties of body-part-centered receptive fields. a, Consider a simple one-dimensional environment in which an object (for example, a wasp) always moves toward a body part. Contact is rewarded (here with a negative reward; $r = -1$). **b**, Description of value calculation, demonstrating the effect of temporal discount in shaping body-part-centered value fields. The process of learning the value of the action 'stay still' (equation at the top). Agents incrementally update the values of the 'stay' action for each possible wasp position, resulting in the optimal description of the value of staying ('Final Q values', right). The value of staying in place forms a graded body-part-centered receptive field. Each row of line plots reflects a given amount of temporal discount (γ) for future rewards. **c**, World dynamics shape body-part-centered fields because they affect the positions from which an object can contact the body. We show results for $\gamma = 0.7$ in a 2D world. Color maps depict the Q value of the action 'stay still' as a function of object position relative to the

hand. **d**, Alternative object trajectories resulting in differently shaped body-part-centered fields. **e**, Sensory uncertainty, resulting form sensory noise and/or stochastic object motion, also changes the perceived positions from which the object can touch the body. This alters value fields, generally making them more spatially diffuse. **f**, An agent which has access to more actions than just staying still can maximize reward by moving toward objects offering a positive reward (apple) and away from objects offering a negative reward (wasp). **g**, The agent's motor repertoire also contributes to shaping body-part-centered fields. Columns show Q values for different actions. Having access to more actions generally results in smaller body-part-centered fields in response to negative-reward stimuli, because the agent can avoid them more effectively and hence expects the negative reward from fewer locations. Conversely, more available actions allow the agent to reach positive-reward stimuli more easily, thus expanding body-part-centered fields.

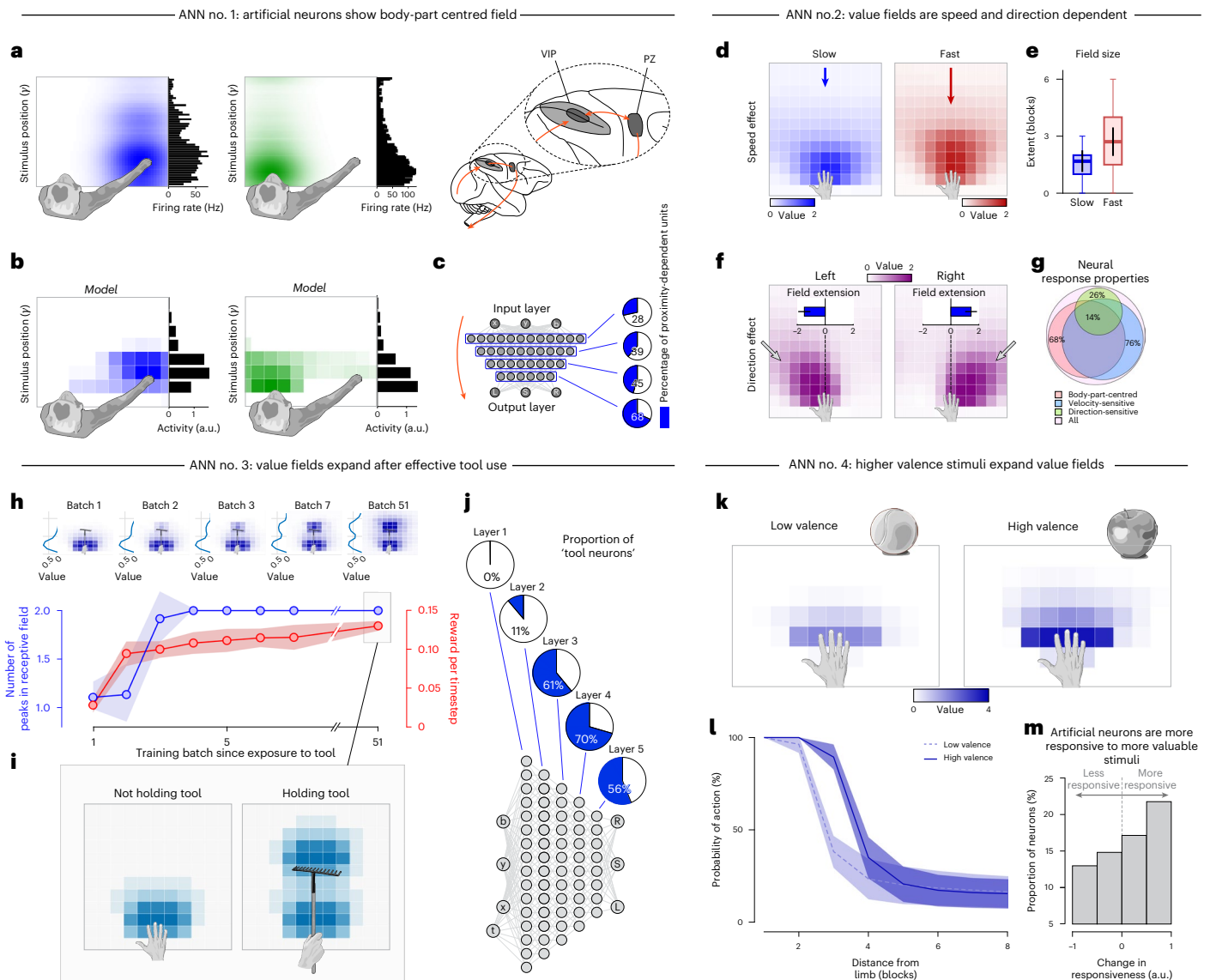Body-part-centered fields emerge naturally from action value

Fig. 2 | ANNs subserving interception and avoidance actions demonstrate canonical properties of peripersonal responses. a, Firing rates from example 'peripersonal' neurons with visual receptive fields around the hand and the face (left and center), prevalent in areas VIP and PZ (right; data from ref. 26). **b**, Activity for two example artificial neurons with receptive fields around different artificial body parts, from an ANN that simultaneously controlled two body parts and was rewarded for contacting an object. a.u., arbitrary units. **c**, Percentages reflecting the proportion of units showing body-part-centered activity (Supplementary Methods 3.2.1). **d**, Q values output by the ANNs forming larger body-part-centered fields in response to fast (red) rather than slow (blue) stimuli. **e**, This expansion is evidenced by the vertical distance at which the value of actions is closest to the mean ('field size'). The error bars, boxes and whiskers depict the s.d., 25th–75th percentiles and 5th–95th percentiles across stimulus x position, respectively. **f**, Artificial Q fields expanding in the direction of incoming stimuli (arrows). **g**, Most underlying artificial neurons display body-part-centered receptive fields, many of which are velocity sensitive (Supplementary Methods

3.2.2). **h**, An agent learning to use a tool progressively expands its Q field from just the limb, to eventually surrounding the tool tip. Color maps show Q values at various learning stages. Line plots to the left of each colormap are Q averages along the x dimension. The second row shows the number of Q-field peaks (blue) and the network performance (red) as a function of learning. Shaded areas show the s.d. across model instantiations. **i**, After learning, a Q-field peak appears around the tip of the tool, but only when the agent holds it. **j**, Many units in the underlying ANN show a similar field remapping after training. The pie charts indicate the proportion of tool-sensitive units (Supplementary Methods 3.2.3). **k**, Body-part-centered value fields are enhanced in response to stimuli of higher (right) versus lower (left) absolute valence. **l**, Actions creating or avoiding contact with a high-valence object are initiated at a further distance from the body. Shading depicts the s.d. across stimulus x position. **m**, Artificial neurons are more responsive to higher-valence stimuli. The histogram shows the difference in responsiveness to stimuli of two valence magnitudes (Supplementary Methods 3.2.4). Panels **a** (left) and **b** adapted from ref. 26, American Physiological Society.

value when the object is near a body part, where probability of contact is highest.

(3) Action repertoire. The agent's action repertoire also shapes body-part-centered fields, influencing the distance to an object (d_t; Fig. 1f,g). As Q values for current actions take into account potential actions that can be made in the future (Supplementary Information 2), it follows that body-part-centered fields can also

appear in response to static stimuli: even if the wasp is stationary, we can reach out to swat it away.

Importantly, there are many different body-part-centered fields that each reflect the value of different actions. For example, when faced with a distant wasp, we might not want to move our hand, but rather feel the urge to walk away. The value of walking away supersedes that

of moving a limb. The 'walk-away field' would be centered on the whole body and expand far into space, whereas the 'move hand out the way' field would be centered on the hand and expand less[6]. We provide a demonstration of the interactions between the values of different action types and their dependences on stimulus distance in this wasp scenario in Supplementary Fig. 1.

In summary, the interaction with objects that yield rewards, even indirectly (Extended Data Fig. 1), naturally leads to the formation of body-part-centered fields reflecting action values.

Body-part-centered fields in ANNs resemble their biological counterparts

Despite its simplicity, the principle that object–body contact is rewarded or punished accounts for many behavioral and neurophysiological results. In this section, we show that multiple canonical properties of peripersonal fields (Fig. 2a) emerge naturally in ANNs trained to make or avoid contact with environmental stimuli (Fig. 2b,c).

We trained a set of ANNs to move one or two body parts in a two-dimensional (2D) grid world. Precise details about these ANN models—each exposed to an environment designed to test several canonical properties of peripersonal fields—are provided in Supplementary Methods 3.2. Briefly, in all models an agent, controlled by an ANN, outputs Q values for three possible actions per limb: move left, stay still or move right (Fig. 2c; 'L', 'S', 'R' units, respectively). The agent performs the action with the highest Q value. The ANN inputs are the x position of its body parts and the x and y positions of external objects (Fig. 2c; 'b', 'x', 'y' units, respectively). These objects appear at the extreme end of the grid world opposite the agent and generally move toward the agent's side, with some additional random motion at each time step. The exact speeds and trajectories differ slightly between models (see Methods, 'ANN models' and Supplementary Information 3.2.2 for details). An object can be either a 'goal' or a 'threat'. Limb contact with a goal or a threat results in a positive or negative reward, respectively. For each model we ran multiple ANN architectures to test the consistency of experimental results across different network types (Methods 'ANN models'). This approach deliberately abstracts over physiological details, without making specific claims about what type of neurophysiological network mechanisms create biological peripersonal fields. Instead, our approach demonstrates that the results obtained in biological systems can be normatively explained from simple assumptions.

The ANNs recreated multiple foundational experimental findings describing canonical peripersonal fields:

(1) They are body-part centered, remaining anchored to a body part when it moves[6,7,9,10,15,23]
(2) They reshape depending on both stimulus speed and direction relative to the body part[9,24–26]
(3) They expand with tool use, but only after some initial experience with the tool[27,28]
(4) They expand in response to stimuli of stronger valence[13,29,30]

Artificial agents produce receptive fields anchored to a limb. In ANN model no. 1, we tested whether body-part-centered fields emerge from action value (Methods, 'Base ANN model' and Supplementary Information 3.2.1). Action values consistently correlated with limb–object proximity ($P \le 7.97 \times 10^{-3}$, all $|\rho| > 0.086$, mean $|\rho| = 0.509 \pm 0.171$ (s.d.); false recovery rate (FDR)-corrected Pearson's correlation tests). A substantial proportion of artificial neurons also showed body-part-centered receptive fields, regardless of network architecture and number of limbs ($51 \pm 9\%$; Fig. 2b,c and Extended Data Fig. 2; see Supplementary Methods 3.2.1 for a description of network architectures tested and their rationales). These units predominated in later network layers ('layer depth' increased their proportion by 36.0% (95% confidence interval (CI) = 26.0–46.1%), $P = 1.56 \times 10^{-9}$, $t = 7.081$, $\eta_p^2 = 0.249$, 62 degrees of freedom; linear mixed effects (LME) main

effect of layer depth, $68 \pm 21\%$ of units being body-part centric in the last layer; Fig. 2c). We found qualitatively identical results when learning was performed on-policy using SARSA (Supplementary Fig. 1).

Stimulus dynamics reshape artificial body-part-centered fields. In ANN model no. 2 we exposed agents to stimuli moving at different speeds and directions (Methods, 'Model variations' and Supplementary Methods 3.2.2). Action-value fields expanded when stimuli moved faster ($P \le 8.46 \times 10^{-3}$, all $\rho \ge 0.42$, mean $\rho = 0.494 \pm 0.042$; FDR-corrected Pearson's correlation tests; Fig. 2d,e), particularly in the direction from which stimuli approached ($P \le 3.24 \times 10^{-12}$, all $\rho \ge 0.85$, mean $\rho = 0.910 \pm 0.064$; FDR-corrected Pearson's correlation tests; Fig. 2f). The extent of receptive fields of individual artificial neurons similarly correlated with stimulus speed and direction (the receptive field extent of $26 \pm 6\%$ of units correlated with vertical movement speed, while $76 \pm 8\%$ of units expanded in the horizontal direction of movement and $20 \pm 6\%$ of unit activity correlated with both; Fig. 2g).

Tool use reshapes artificial body-part-centered fields. In ANN model no. 3 we gradually exposed agents to simulated tool use (Methods, 'Model variations' and Supplementary Methods 3.2.3). As agents gained more experience with the tool, the number of action-value field peaks increased from 1 to 2 ($P \le 7.57 \times 10^{-12}$, all $\rho \ge 0.272$, mean $\rho = 0.309 \pm 0.035$; FDR-corrected Pearson's correlation tests; Fig. 2h), with a new receptive field forming around the tool tip. After 50 training periods, the presence or absence of the tool strongly affected the value fields (1.5 ± 0.5 versus 2.0 ± 0.0 peaks in the value field: $P = 3.74 \times 10^{-5}$, $Z = -4.1$, effect size $r = -0.69$; approximate Wilcoxon's signed-rank test; Fig. 2i). A substantial proportion of units within the networks showed the same type of variation in their receptive fields as a function of tool use and training ($42 \pm 6\%$; Fig. 2j). This was especially true in deeper network layers, where units were closer to representing action value ('layer depth' increased the proportion of tool neurons by 69.0% (95% CI = 34.0–100.0%), $P = 8.81 \times 10^{-4}$, $t = 4.29$, $\eta_p^2 = 0.355$, 13 degrees of freedom.; LME main effect of layer depth; Fig. 2j).

Reward magnitude alters artificial body-part-centered fields. In ANN model no. 4 we investigated whether stimuli with higher valence elicit stronger responses in artificial units and yield larger body-part-centered fields. We varied the value of the reward r_t (Methods, 'Model variations' and Supplementary Methods 3.2.4), and found that all agents showed larger output Q values for higher reward stimuli (all $P \le 4.76 \times 10^{-9}$, all $|z\text{-score}| \ge 5.86$, mean $|z\text{-score}| = 11.3 \pm 7.0$, all effect size $|r| \ge 0.096$; FDR-corrected Wilcoxon's signed-rank test; Fig. 2k). Consequently, actions related to intercepting a higher reward stimulus were initiated when the stimulus was at greater distances from the body (all $P = 9.77 \times 10^{-4}$, all $|z\text{-score}| \ge 3.317$, mean $|z\text{-score}| = 3.51 \pm 0.17$, all effect size $|r| \ge 0.65$; FDR-corrected Wilcoxon's signed-rank tests; Fig. 2l). Individual units were more affected by the position of high-reward than low-reward stimuli ($P = 6.2 \times 10^{-3}$, $z\text{-score} = 2.738$, effect size $r = 0.299$; Wilcoxon's signed-rank test; Fig. 2m).

Multiple peripersonal fields can form an egocentric map
Having demonstrated that contact value creates body-part-centered value fields both theoretically (Fig. 1) and in ANNs (Fig. 2), we next integrated this knowledge with theoretical work on model-based reinforcement learning and generalized policy improvement[4,31,32]. Such theories address a common challenge in decision-making: the rapid changes of context and environment. Recomputing action values in response to each change is not only inefficient, but risks overwriting previously learned associations.

Collections of peripersonal neurons form successor features. One solution lies in using a predictive model of the world, rather than learning each task from scratch. This approach, known as model-based
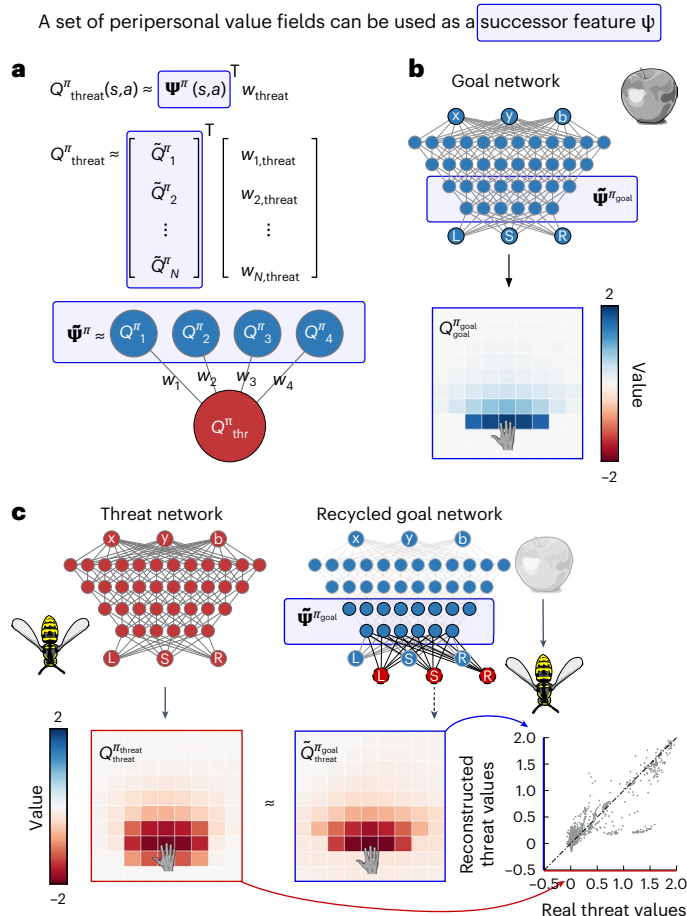
Fig. 3 | A collection of body-part-centered value fields can be used as a successor feature ψ^π. **a**, Successor features ψ^π used to create action values Q^π_{thr} for a new reward configuration (thr) and hence new task (Glossary, Supplementary Information 1). In this example the new task ('thr') consists of responding to an object offering negative reward: a threat. Top, basic equation describing linear feature recombination to create new action values Q^π_{thr}. Middle, a successor feature ψ^π can be composed of N Q-value approximations $[\widetilde{Q}^\pi_1, \widetilde{Q}^\pi_2, \dots, \widetilde{Q}^\pi_N]$. Bottom, this recombination can be envisioned as a one-layer ANN in which all N basis units contribute to the activity of an output unit representing Q value for the new task 'thr'. **b**, Approximate successor features $\widetilde{\psi}^\pi$ can be composed of ANN units with body-part-centered receptive fields. Although the formal theory behind successor features assumes that the N Q values represent N different reward configurations, here we instead used $\widetilde{\psi}^{\pi goal}$, a collection of neurons that offer multiple approximations of action value for a single reward configuration 'goal'. The reward configuration 'goal' is defined as the situation in which an object offers a positive reward. The optimal Q values for that reward configuration and the resulting task are shown in the colormap below as $Q^{\pi goal}_{goal}$. **c**, An ANN that has only learned to intercept positive-reward stimuli can be 'recycled' to approximate an appropriate value field for avoidance actions. The output of the red threat network ($Q^{\pi threat}_{threat}$) could be reconstructed by taking a weighted sum over $\widetilde{\psi}^{\pi goal}$, that is, a sum over the neural activities in the second half of the blue goal network. As this reconstruction uses successor features from a network trained to intercept only goals, we label the reconstructed action values as $\widetilde{Q}^{\pi goal}_{threat}$. The scatter plot shows the Q values calculated by the threat-specific network $Q^{\pi threat}_{threat}$ (x axis) plotted against the recycled Q values from the goal network $Q^{\pi goal}_{threat}$ (y axis).

reinforcement learning, enables agents to forecast expected rewards for new reward configurations by mapping them on to anticipated states and thereby enables estimation of appropriate action values. Among the solutions for implementing model-based reinforcement learning, the world model offered by 'successor features' (ψ^π; which are generalizations of the concept of a successor representation—see

Glossary in Supplementary Information 1) is particularly compelling, mainly as a result of its modular nature; individual successor features are building blocks that can be learned independently and later recombined to suit new reward configurations ($Q^\pi_i(s_t, a) = \psi^\pi(s_t, a)^T w_i$; Fig. 3a).

This principle of modular recombination should extend to peripersonal value fields, because collections of Q values can be used as successor features (that is, $\psi^\pi \approx \widetilde{\psi}^\pi = [\widetilde{Q}^\pi_1, \widetilde{Q}^\pi_2, \dots, \widetilde{Q}^\pi_N]^T$; Glossary in Supplementary Information 1)[31]. To demonstrate that such recombination is practically feasible and can be performed even with value-approximating units rather than full value functions, we trained two networks: network A with only positive-reward stimuli (Fig. 3b) and network B with only negative-reward stimuli (Fig. 3c). By linearly recombining activity from units in layers close to the output layer (that is, units likely to have body-part-centered receptive fields) from network A, we successfully approximated the threat value functions from network B ($P < 10^{-150}, \rho = 0.92$, FDR-corrected Pearson's correlation tests; Fig. 3c and Supplementary Methods 3.2.5). Hence, peripersonal neurons might constitute building blocks of a short-term predictive model of the world near the agent, reminiscent of hippocampal place cells[3] but working at shorter spatiotemporal scales.

Collections of varied successor features form an egocentric map. Successor features[31] also address a limitation of traditional successor representations by allowing for generalization beyond learned policies[4]. For instance, envision an agent facing new tasks (Fig. 4a) after having previously learned policies to both intercept and avoid objects (Fig. 4b). This agent can choose between these two learned policies and rescale its previously learned action values (that is, $a = \underset{A}{\arg\max}\ \underset{k}{\max}\ \widetilde{\psi}^{\pi_k}(s, A)^T w_i$; Fig. 4c,d). This immediately creates an effective set of action values for new tasks, such as intercepting a goal after a mandatory brief pause, evading a threat larger than those previously encountered, allowing an object to pass at a specific distance from its limb or dealing with an object that switches between offering reward and punishment based on the location in which it first appears (Fig. 4a, rows 1–4, respectively). Consequently, a comprehensive collection of body-part-centered value fields—especially those acquired from diverse tasks—can create a highly flexible model of the world near the body, an egocentric map ($\Psi = \{\widetilde{\psi}^{\pi_1}, \widetilde{\psi}^{\pi_2}, \dots, \widetilde{\psi}^{\pi_N}\}$; Fig. 4b).

Having demonstrated that a flexible egocentric map can be constructed from value fields, we now show that ANNs trained on multiple tasks naturally separate into neural architectures that are suitable for use in egocentric maps.

ANNs naturally separate into task-selective subnetworks. A comprehensive egocentric map is more effective if it can differentiate between tasks, prompting us to investigate under which circumstances task-specific modules appear within a network. By exposing 45 agents to 2 distinct tasks simultaneously (interception and avoidance; Supplementary Methods 3.2.6), we discovered that, after successful training, nearly all networks developed two spatially segregated subnetworks, favoring either positive- or negative-valence stimuli (units with similar stimulus preferences were closer to each other: $P \leq 0.05$ for 37 of 45 networks, mean t statistic = -5.1 ± 4.0, FDR-corrected main effect of interunit distance on unit similarity in an LME; Fig. 5a,b). Networks with more subnetwork structure also achieved greater total reward, suggesting that modularization is an integral and emergent aspect of successfully learning multiple tasks ($P = 0.021, \rho = -0.343$ (95% CI = -0.58 to -0.05); Pearson's correlation test between t statistic and average reward per time step; Fig. 5b). Functionally, almost all networks coded positive- and negative-valence stimuli in orthogonal subspaces ($P \leq 0.05$ for 40 of 45 networks, FDR-corrected bootstrap test with 1,000 samples,
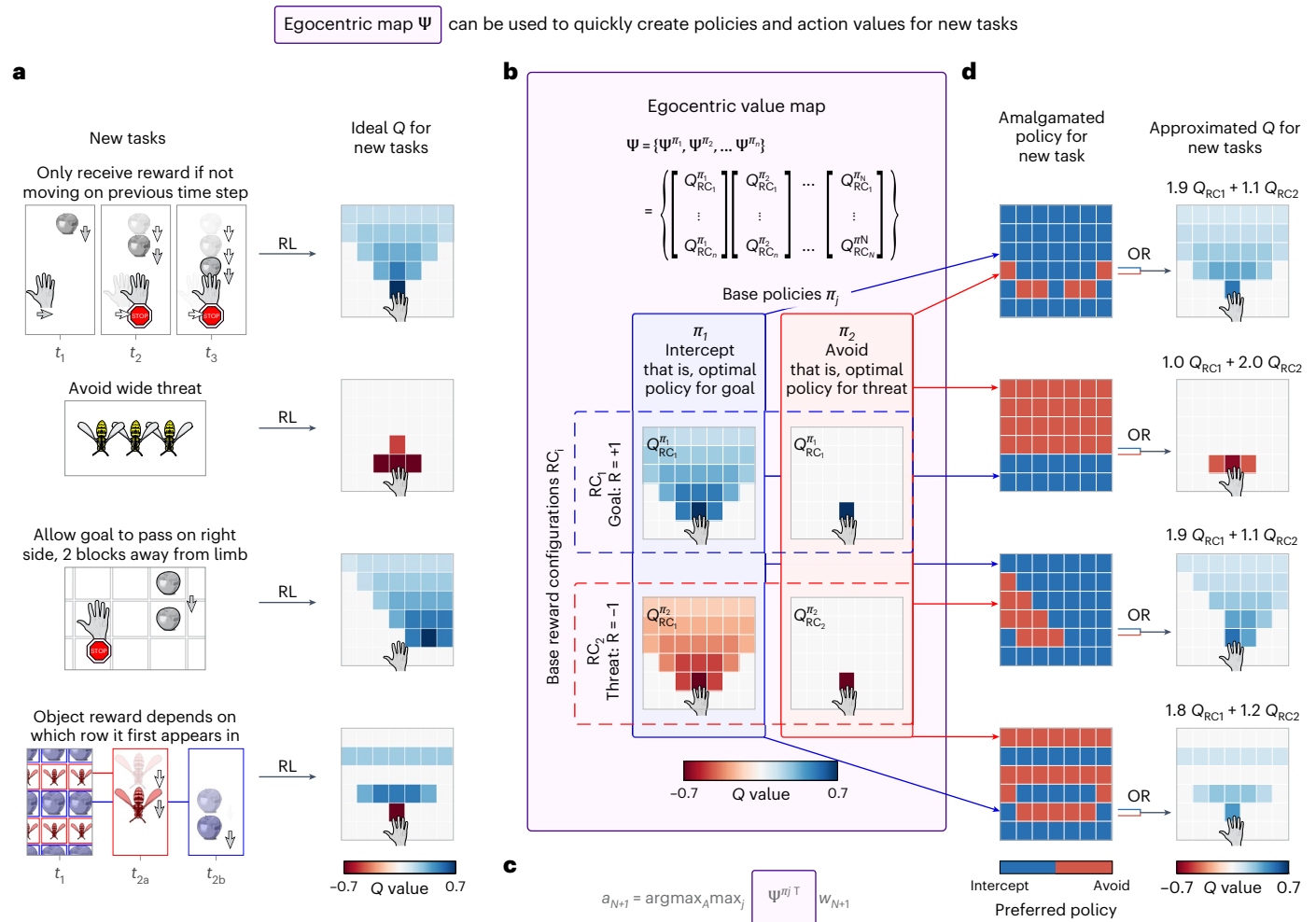
Fig. 4 | A collection of successor features can be used as an effective map of the world near the agent: an egocentric value map. a, Illustration of four new tasks, chosen to demonstrate the mechanism of an egocentric map. Left, task description. Right, the ideal Q values for each of the four new tasks (for the task at the bottom, we display action value only as a function of position for the specific timepoint at which the object appears). **b**, Scheme depicting an egocentric map Ψ, which is a collection of successor features. Specifically, Ψ is composed of N successor features $\boldsymbol{\psi}^{\pi_j}$, each defined by a particular policy π_j. The successor feature for a policy π_j itself consists of body-part-centered value fields for N base reward configurations $[Q_1^{\pi_j}, Q_2^{\pi_j}, \ldots, \tilde{Q}_N^{\pi_j}]$, as suggested by Barreto et al.[31]. This example has two successor features (columns) for two reward configurations (rows). The 'goal' policy (that is, 'intercept', π_1) is optimal policy when the agent is rewarded by contact ('goal' reward configuration RC_2). The 'threat' policy (that is, 'avoid', π_2) is optimal when the agent is punished by contact ('threat' reward

configuration RC_2). Color maps show the value of the action 'stay'. **c**, Mathematical formula for recycling an egocentric value map to face a new reward configuration (and hence task) RC_{N+1}. If the agent can estimate action values for multiple policies on its base reward configurations, and can also switch between policies π_j at each timepoint, it will be able to find the maximally valuable action for a new task. **d**, In each state of a new task, the agent selects the policy (left) which, combined with reweighting its task-specific value maps (right, equations), results in maximally effective actions. This process creates new Q values (right; color maps) that approximate the ideal Q values for each of the four new tasks described in **a**. We computed weights and preferred policies by minimizing the difference between the approximate (**d**) and ideal (**a**) Q values for each new task. In an agent, these weights are probably learned by a slower, but more realistic, process that doesn't require access to the optimal Q values[31].

against the null hypothesis that the angle between encoding directions diverged >15° from 90°, mean = 95.2° ± 10.6°; Fig. 5c).

The emergence of this subnetwork structure and orthogonal coding was not contingent on aspects of network design such as architecture, regularization and the transfer function of units composing the network: all network types that performed the tasks successfully showed subnetwork structure and orthogonal coding (Extended Data Fig. 3a,b,d,e). Furthermore, networks that performed the tasks better showed more structural and functional modularization (correlation between structural t statistic and reward per time step: $P = 4.53 \times 10^{-36}$, $\rho = -0.47$ (95% CI −0.52 to −0.41); correlation between task-coding orthogonality and reward per time step: $P = 1.70 \times 10^{-17}$, $\rho = -0.33$ (95% CI −0.38 to −0.26); Extended Data Fig. 3c,f). Finally, the more modular networks provided better successor features to reconstruct the optimal

Q values for the new tasks shown in Fig. 4a, even after controlling for the effects of total reward on the trained task (Extended Data Fig. 3g,h; partial correlation between structural t statistic and reconstruction quality: $P = 2.67 \times 10^{-10}$, $\rho = -0.13$ (95% CI −0.17 to −0.09); partial correlation between task-coding orthogonality and reconstruction quality: $P = 3.57 \times 10^{-8}$, $\rho = -0.11$ (95% CI −0.45 to −0.07)).

Thus, the necessary ingredients for egocentric maps emerge naturally even in simple ANNs.

Egocentric value maps robustly fit existing empirical data. To statistically assess the ability of egocentric value maps to explain empirical data, we created an approximate three-dimensional (3D) map around the upper body and fitted it to empirical results from 23 published experiments.
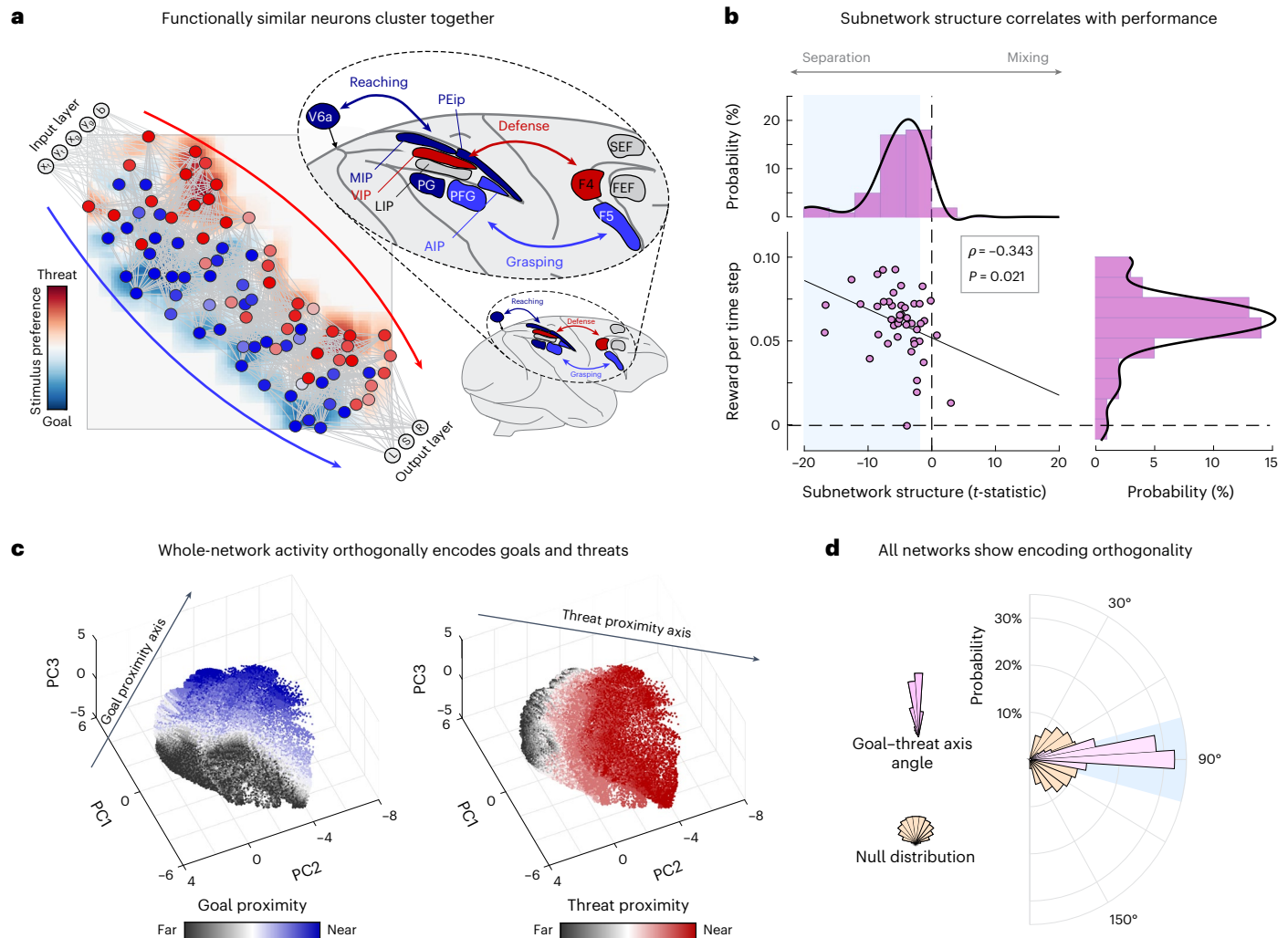
Fig. 5 | ANNs trained on multiple tasks show architectures conducive to egocentric maps. a. An ANN trained on simultaneous interception and avoidance tasks naturally adopting a modular structure, beneficial for use in an egocentric map (left, network graph). In this example ANN, individual neurons are classified as threat- or goal-preferring (red and blue, respectively). When the neuron-to-neuron distance is set to be the inverse of the neuron-to-neuron connection strength, clear threat- or goal-preferring subnetworks appear (red and blue backgrounds, respectively). This modular structure is reminiscent of the primate parieto-premotor system, where peripersonal neurons cluster together based on their behavioral function (right). **b**, More subnetwork structure results in better task performance. Data are from 45 networks trained on simultaneous interception and avoidance tasks. Most networks (37 of 45) showed above-chance subnetwork structure ($t < -2.5$, t statistic of a main effect of neuron similarity on neuron–neuron distance; top histogram, blue shaded area; Supplementary Methods 3.2.6). Right histogram indicates average reward. **c**, Functional modularity in the 45 networks described in **b**: activity from the first three principal components of an example network. Goal-proximity coding (black-to-blue scatter plot) was orthogonal to threat-proximity coding (black-to-red scatter plot). **d**, Most networks (38 of 45) show above-chance task coding orthogonality. Polar axes describe the direction of goal and threat coding. Angles within ±15° of 90° are shaded light blue. AIP, anterior intraparietal area; F4, premotor area F4; F5, premotor area F5; FEF, frontal eye field; LIP; lateral intraparietal area; MIP, medial intraparietal area; PEip, intraparietal part of area PE; PFG, parietal area PFG; PG, parietal area PG; SEF, supplementary eye field; V6a, visual area 6A; VIP, ventral intraparietal area. Panel **a** (right) adapted from ref. 9, Elsevier.

We constructed this egocentric map as the set of action values related to moving three individual body parts (head, torso, hand) when faced with two reward configurations: goals and threats (Fig. 6a; that is, π_1 is a collection of approximate Q values for that body part, for example $\widetilde{\psi}^{\pi}_{\text{hand}} = \left[\overline{Q}^{\pi}_{\text{hand up}}, \overline{Q}^{\pi}_{\text{hand down}}, \dots, \overline{Q}^{\pi}_{\text{hand stay}} \right]^{\text{T}}).$

Further details of the egocentric map construction and empirical data fitting are in Methods, 'Empirical data fitting' and Supplementary Methods 3.4. Briefly, we derived a single value representing egocentric map activity for each experimental condition by first extracting the maximal absolute Q value across actions and policies for each body part and then averaging these Q values across body parts:

$$S_{\text{exp cond}}(s, a) = \frac{1}{N'} \sum_{\text{body-part}=1}^{N'} \max_{\text{task}} \max_{a} \psi^{\pi_{\text{task}}}_{\text{body-part, exp cond}}(s, a)$$

(Fig. 6c).

Note that calculation of the maximal Q value across actions and policies also occurs when recycling the egocentric map for new tasks[31] and it can be conceptualized as the competition between affordances observed in real brains[33].

The approximated egocentric map successfully fit the data across all experiments ($P = 0.99$, $\chi^2 = 122$ (95% CI = 88–159); Fig. 7 and Extended Data Fig. 4; the P value obtained from comparing the χ^2 statistic with the χ^2 distribution is a measure of the confidence that the empirical data could have been generated by a process similar to the one described
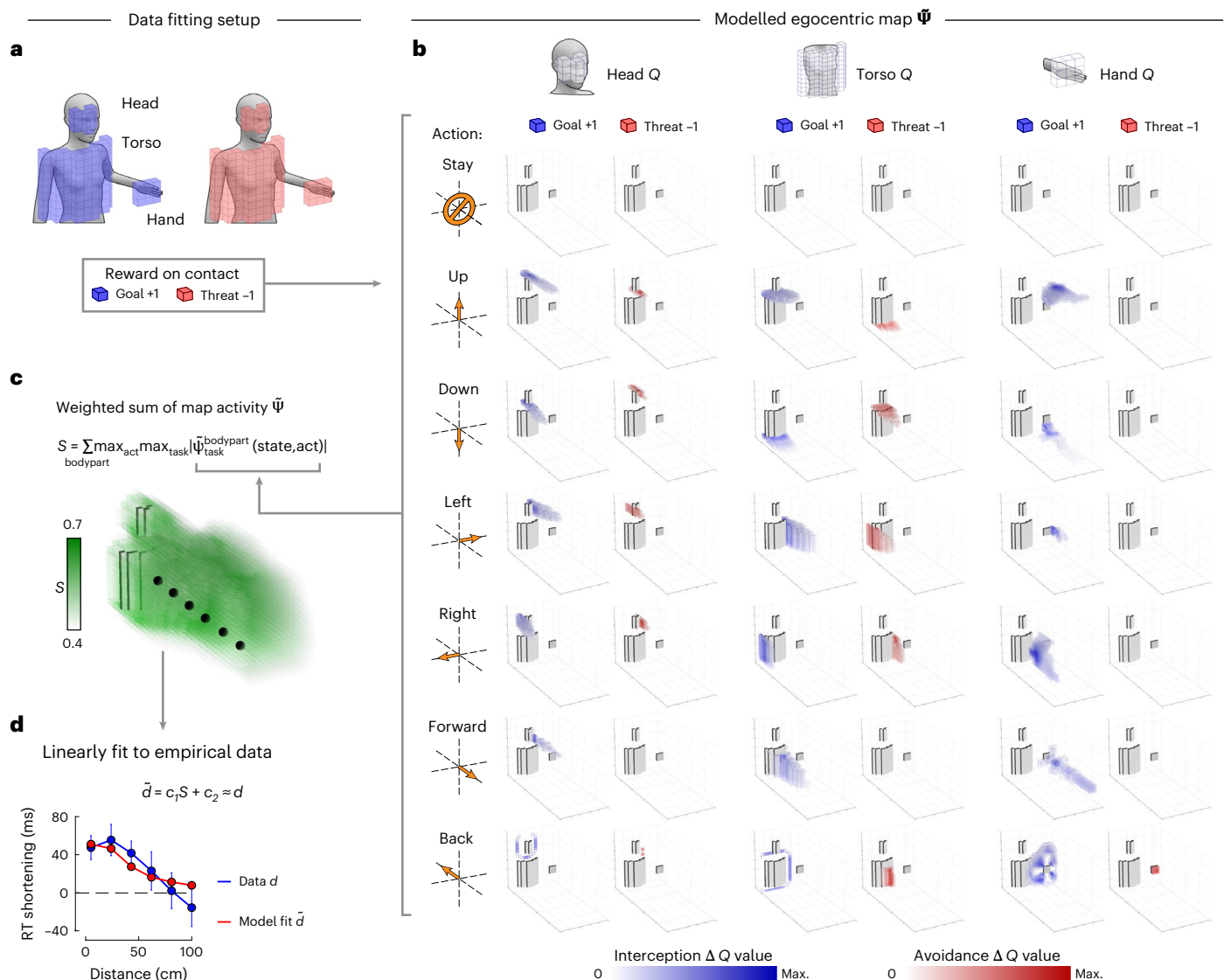
Fig. 6 | Fitting an egocentric map to empirical data: setup and method. a, Setup for approximating an egocentric map $\bar{\Psi}$ around the upper body, with the aim of fitting empirical data. Colored voxels indicate the body locations rewarded (blue) or punished (red) on contact. **b**, The approximated egocentric map included Q values for seven actions (rows: stay, up, down, left, right, forwards, backwards), for each of three body parts (columns: head, torso, hand) and for two tasks (subcolumns: intercept and avoid). Colored voxels indicate the stimulus locations where the value of each action was higher than any other (ΔQ values; we show values for optimal policies for both tasks). Colors are normalized within action and task. The lowest 20% of ΔQ values within action and task are not shown. **c**, Our framework posits that empirical peripersonal measures reflect a

single summed estimate of egocentric map activity (S). We extracted S by first taking the maximal absolute Q value across actions and tasks for each body part and every point in space, and then averaging these three values across body parts, as specified in the equation. Green voxels represent the resulting summed map activity. Spheres indicate locations of empirical measurements in an example experiment in which participants had to react as fast as possible to a tactile stimulus on the torso while task-irrelevant auditory stimuli moved toward their torso (see **d**). **d**, The summed egocentric map activity S linearly fits (red) the example empirical data (blue)[14], showing reaction time shortening when the auditory stimuli were closer to the torso ('RT shortening')[7]. The error bars show s.e.m. across participants. Max., maximum.

in our model. In other words, if the P value is very small, we must reject the null hypothesis that our model explains the data, but, if the P value is very large, we must accept that hypothesis as a possibility[21]). This fit, which involves a limited number of parameters, underscores the model's biological relevance and broad explanatory power. Only alternative models that also leveraged an egocentric map survived rigorous statistical tests (Extended Data Fig. 4), further affirming the biological validity of egocentric maps.

Discussion

We have demonstrated that body-part-centered fields naturally emerge from action value, under the simple assumption that agents often experience reward on contact with objects (Fig. 1). In artificial agents, such

body-part-centered value fields can be observed in both single-unit activity and behavior. These fields are sensitive to the same factors that modulate biological peripersonal fields: stimulus dynamics, tool use and valence (Fig. 2). Therefore, biological peripersonal fields probably index contact-action value, as originally hinted by electrophysiology[26] and later suggested theoretically[6].

Existing literature also indicates the importance of action value in creating peripersonal fields. First, canonical peripersonal neurons are predominantly located in sensorimotor regions such as the putamen, ventral intraparietal area (VIP) and dorsal F4 (refs. 7,9) (Figs. 2a and 5a). Second, chemically and electrically enhancing VIP and F4 activity increases the probability and vigor of defensive actions[7]. Third, parieto-premotor areas housing peripersonal neurons are more active
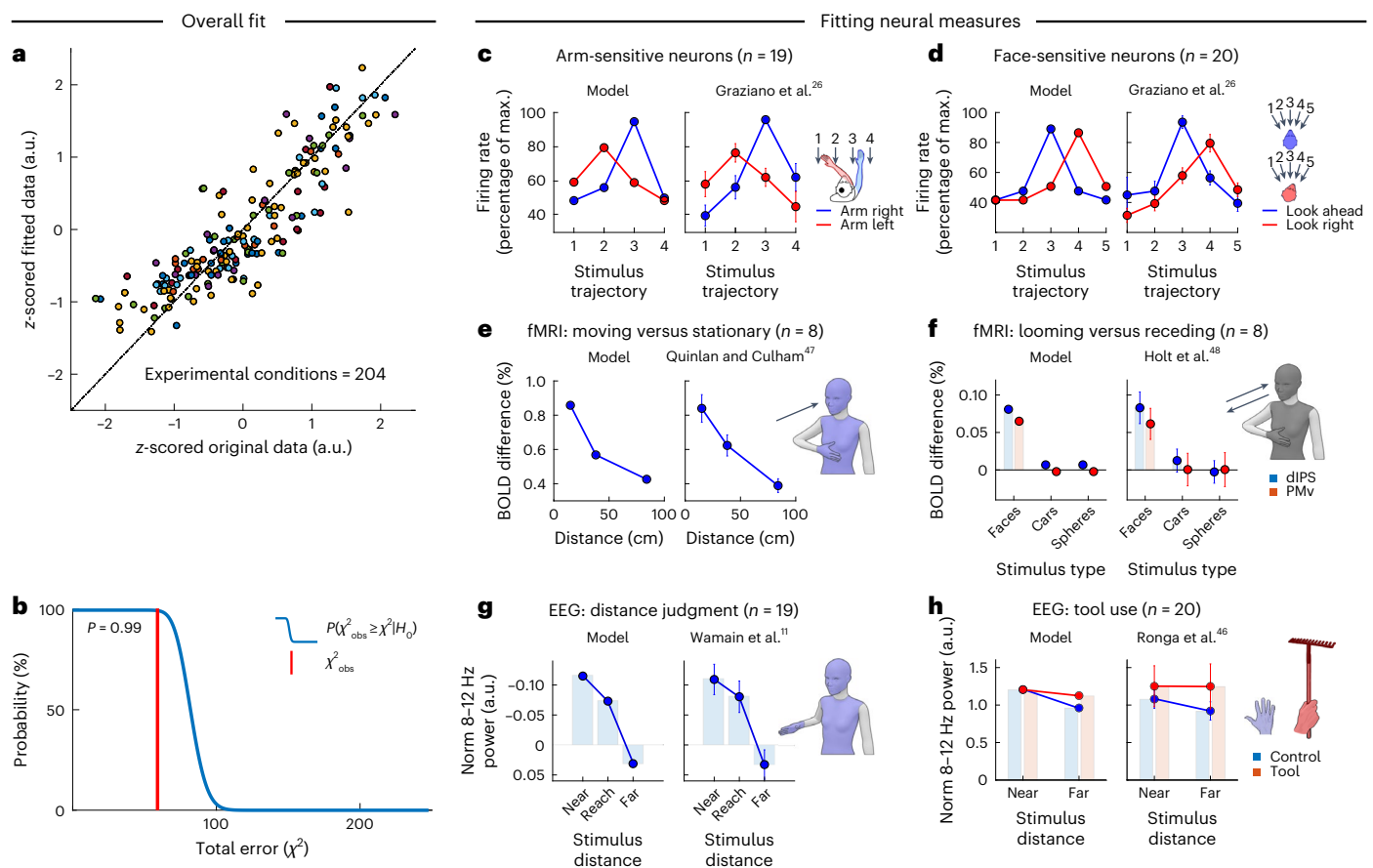
Fig. 7 | Fitting egocentric maps to empirical data, part 1: overall fit and neural measures. a, The distribution of real (x axis) and fitted (y axis) data along the identity line, demonstrating that an egocentric map robustly explains all experimental data. Twenty-three experiments across ten laboratories were included in total[11,14,25–27,46–48,49,50]. Colors indicate different experiments. For data fitting methodology, see Extended Data Fig. 4a–e, Methods, 'Empirical data fitting' and Supplementary Methods 3.4.1. **b**, The real total error (χ^2_{obs}; red line) superimposed on the probability of observing a theoretical total error χ^2 that is equally large or larger, under the null hypothesis H_0 that the model explains the data ($P(\chi^2_{obs} \geq \chi^2|H_0)$; blue curve). **c–h**, Empirical data (right in each panel) and model fits (left in each panel) from experiments in which neural peripersonal

measures were collected. For data and fits of behavioral measures, see Extended Data Fig. 4f–q. Colors indicate experimental conditions. The shaded surfaces on figurines show body parts included in the egocentric map. The error bars show the s.e.m. For a detailed description of each fitted experiment, see Supplementary Methods 3.4.1. **c**, See Supplementary Methods 3.4.1.1 for details. **d**, See Supplementary Methods 3.4.1.1 for details. **e**, See Supplementary Methods 3.4.1.2 for details. **f**, See Supplementary Methods 3.4.1.3 for details. **g**, See Supplementary Methods 3.4.1.4 for details. **h**, See Supplementary Methods 3.4.1.5 for details. Panels **c** and **d** adapted from ref. 26, American Physiological Society. Norm., normalized.

in response to stimuli perceived as rewarding[30]. Fourth, peripersonal fields expand when higher reward is offered by objects further from the body[34], a phenomenon that is also consistent with value coding (Supplementary Fig. 5 and Supplementary Information 3.4.2 and 4.2.2).

In contrast to our action-based model, which relies only on a simple a priori assumption about the value of contact, most alternative models require highly specific assumptions about encoded variables or neural architecture. For example, certain models have predefined multisensory integration pathways[19], whereas others presuppose explicit encoding of impact prediction[18,20] or multisensory integration[19], without explaining their necessity.

Despite some recent conceptual improvements of alternative models[20,35] (detailed in Supplementary Discussion 3.2), they still fail to fit the extensive combined dataset for which our action-based model offers a statistically robust explanation (Extended Data Fig. 5). Simple monotonous decay functions do not fit the data because they lack theoretical underpinnings and hence require additional parameters for each new experimental condition. The limitations of impact prediction and multisensory integration models, which rely on predicting touch under uncertainty, are particularly pronounced in 3D environments, in which fewer trajectories result in contact with the body compared

with lower dimensions. This important aspect was underappreciated in prior research, which typically modeled peripersonal fields in only one or two dimensions[17,18,20,21]. Nevertheless, uncertainty still probably plays a role in shaping peripersonal fields, as evidenced by the success of action value models incorporating it (Extended Data Fig. 5). Future work could explore frameworks like active inference, which explicitly account for sensory and perceptual uncertainty, potentially offering a more integrated perspective on peripersonal responses[36]. To provide physiological insight, such future models should also be tested against more granular single-cell data.

This article establishes the conceptual foundation for an egocentric value map: multiple body-part-centered value fields can be combined to form successor features that provide a model of the world near the agent in terms of its action values. Once diverse successor features and policies have been learned for a set of base tasks, they can effectively be recombined to face new reward configurations (Figs. 3 and 4). ANNs containing peripersonal units can therefore function as egocentric maps, capitalizing on two of their core features. First, their individual neurons provide the building blocks of an egocentric map because they approximate action value. Second, these ANNs naturally separate into task-specialized subnetworks

(Fig. 5 and Extended Data Fig. 3), providing a favorable architecture for a map that requires separation of task-specific action values.

The concept of an egocentric map offers multiple empirically testable predictions (Supplementary Discussion 5.1) and its validity is statistically supported by the robust fit of multiple empirical datasets (Fig. 7, Extended Data Figs. 4 and 5 and Supplementary Fig. 5). Egocentric value maps have further theoretical merit. They provide formal insight into previous qualitative theories, which posit that the perception of the space near the body is constructed from modular motor schemata[7,23], and support the view of posterior parietal areas as state estimators utilizing overlapping motor codes[37,38]. The egocentric maps that we propose are inherently modular constructs, being composed of successor features spanning tasks and policies. Each successor feature is in turn composed of body-part-centered value functions, which in practice can be sets of neurons that offer preliminary value approximations (Fig. 3). This modularity dovetails the modular structure of the ANNs trained on simultaneous interception and avoidance tasks (Fig. 5a). These task-specialized subnetworks not only mirror the modular anatomy of primate parieto-premotor areas containing peripersonal neurons[9,39], but also offer a simple physical foundation for egocentric maps, which are more effective when provided with Q values covering a broad range of tasks. Namely, if a network separates into task-specialized subnetworks, a reweighting process (Figs. 3 and 4) can access a broad range of tasks by simply taking inputs from a large physical space (Supplementary Discussion 5.3). Candidate neurophysiological mechanisms underlying such reweighting are the dopamine-mediated modulation of cortical outputs to the dorsomedial striatum[40] and the role of primary motor cortex in the integration of activity from parietal and prefrontal motor modules[39] (Supplementary Discussion 5.4). Note, however, that there is no strict one-to-one mapping between the reinforcement learning concept of reward or prediction error and any given dopaminergic pathway[32,40].

Our framework also has clear implications for the interaction between egocentric maps and their allocentric (or cognitive[3,5]) counterparts[41]. It posits a reciprocal relationship, where each map serves as a contextual input for the other. For instance, threatening objects near an animal's body promote escape behavior, whereas knowledge about a dead-end in its escape route biases it toward more vigorous defense from nearby threat. We suggest that cognitive maps indicate the currently active policy to egocentric maps. Conversely, egocentric maps serve as one-among-many building blocks for more complex cognitive maps, which recombine information about the world near the body with information from other brain areas, enabling more comprehensive action planning and evaluation. This view reconciles an apparent contradiction in map learning: although allocentric maps can form without explicit reward[3–5], egocentric value maps inherently depend on it. Allocentric maps are constructed from many components, including nonvalued features, because hippocampal areas are connected to large sections of the cortex. In contrast, egocentric maps directly extend from the brain's fundamental action-oriented functions, offering a more immediate representation of the agent's interaction with its environment.

It is important to highlight that immediate reward on all contact is not necessary for the emergence of body-part-centered receptive fields. Instead, emergence of value fields only requires some contact to be obligatorily rewarded. Pain, evolutionarily ancient and highly preserved, is a potent candidate for such signaling. Fittingly, pain serves three purposes that all rely on action: deploying immediate movements that prevent injury, learning to avoid those actions that result in contact with dangerous objects and enforcing inactivity to facilitate healing[42]. Indeed, pain is a fundamental reinforcement learning signal, resulting in strong activity in both the putamen[43], where peripersonal neurons can be found, and the nucleus accumbens, which is part of the mesolimbic circuitry underlying reinforcement learning[44]. Fitting the notion of pain as a reward signal that constructs egocentric value maps, certain pain-selective neurons in primate areas 7b also have visual body-part-centered receptive fields. Remarkably, their visual responses are most prominent when the approaching object is new or threatening (see Fig. 7 in ref. 45). In addition to the obligatory reward offered by pain, we showed that body-part-centered fields emerge even when contact leads to reward only indirectly (Extended Data Fig. 1). We also noted that an agent does not need to be acting, or even planning to act, for body-part-centered Q fields to exist: under reinforcement learning, the value of several potential actions is continuously calculated, although only a minority of them is actually performed. Movement is only necessary to learn the value of an action in the first instance. After learning, the value can be recalled without movement. Finally, body-part-centered fields should even be observable in response to neutral stimuli: as long as an animal has experienced enough valued contact in the past, its egocentric map will treat objects as if they offer reward on contact.

Thus, although complex, value-independent cognitive maps offer substantial advantages and are conserved once they have evolved, egocentric maps represent a more primal, value-driven form of spatial representation. Rooted in the brain's essential motor functions, these maps emerge naturally, providing a foundational framework for more advanced cognitive processes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-025-01958-7.

References

1. Moser, E. I., Kropff, E. & Moser, M.-B. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* **31**, 69–89 (2008).
2. Banino, A. et al. Vector-based navigation using grid-like representations in artificial agents. *Nature* **557**, 429–433 (2018).
3. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).
4. Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J. & Daw, N. D. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput. Biol.* **13**, e1005768 (2017).
5. Behrens, T. E. J. et al. What Is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* **100**, 490–509 (2018).
6. Bufacchi, R. J. & Iannetti, G. D. An action field theory of peripersonal space. *Trends Cogn. Sci.* **22**, 1076–1090 (2018).
7. Graziano, M. S. & Cooke, D. F. Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia* **44**, 845–859 (2006).
8. Noel, J.-P., Blanke, O. & Serino, A. From multisensory integration in peripersonal space to bodily self-consciousness: from statistical regularities to statistical inference. *Ann. N. Y. Acad. Sci.* **1426**, 146–165 (2018).
9. Cléry, J., Guipponi, O., Wardak, C. & Ben Hamed, S. Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: knowns and unknowns. *Neuropsychologia* **70**, 313–326 (2015).
10. Makin, T. R., Holmes, N. P. & Zohary, E. Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. *J. Neurosci.* **27**, 731–740 (2007).
11. Wamain, Y., Gabrielli, F. & Coello, Y. EEG μ rhythm in virtual reality reveals that motor coding of visual objects in peripersonal space is task dependent. *Cortex* **74**, 20–30 (2016).

12. Sambo, C. F. & Forster, B. An ERP investigation on visuotactile interactions in peripersonal and extrapersonal space: evidence for the spatial rule. *J. Cogn. Neurosci.* **21**, 1550–1559 (2009).

13. de Haan, A. M., Smit, M., Van der Stigchel, S. & Dijkerman, H. C. Approaching threat modulates visuotactile interactions in peripersonal space. *Exp. Brain Res.* **234**, 1875–1884 (2016).

14. Serino, A. et al. Body part-centered and full body-centered peripersonal space representations. *Sci. Rep.* **5**, 18603 (2015).

15. Hunley, S. B. & Lourenco, S. F. What is peripersonal space? An examination of unresolved empirical issues and emerging findings. *Cogn. Sci.* **9**, e1472 (2018).

16. Cléry, J. et al. The prediction of impact of a looming stimulus onto the body is subserved by multisensory integration mechanisms. *J. Neurosci.* **37**, 10656–10670 (2017).

17. Bertoni, T., Magosso, E. & Serino, A. From statistical regularities in multisensory inputs to peripersonal space representation and body ownership: insights from a neural network model. *Eur. J. Neurosci.* **53**, 611–636 (2021).

18. Roncone, A., Hoffmann, M., Pattacini, U., Fadiga, L. & Metta, G. Peripersonal space and margin of safety around the body: learning visuo-tactile associations in a humanoid robot with artificial skin. *PLoS ONE* **11**, e0163713 (2016).

19. Magosso, E., Zavaglia, M., Serino, A., di Pellegrino, G. & Ursino, M. Visuotactile representation of peripersonal space: a neural network study. *Neural Comput.* **22**, 190–243 (2010).

20. Straka, Z., Noel, J.-P. & Hoffmann, M. A normative model of peripersonal space encoding as performing impact prediction. *PLoS Comput. Biol.* **18**, e1010464 (2022).

21. Bufacchi, R. J., Liang, M., Griffin, L. D. & Iannetti, G. D. A geometric model of defensive peripersonal space. *J. Neurophysiol.* **115**, 218–225 (2016).

22. Pezzulo, G. & Cisek, P. Navigating the affordance landscape: feedback control as a process model of behavior and cognition. *Trends Cogn. Sci.* **20**, 414–424 (2016).

23. Rizzolatti, G., Fadiga, L., Fogassi, L. & Gallese, V. The space around us. *Science* **277**, 190–191 (1997).

24. Fogassi, L. et al. Coding of peripersonal space in inferior premotor cortex (area F4). *J. Neurophysiol.* **76**, 141–157 (1996).

25. Huijsmans, M., Haan de, M. A., Müller, B. C. N., Dijkerman, H. C. & Schie van, H. T. Knowledge of collision modulates defensive multisensory responses to looming insects in Arachnophobes. *J. Exp. Psychol. Hum. Percept. Perform.* https://doi.org/10.1037/xhp0000974.supp (2022).

26. Graziano, M. S., Hu, X. T. & Gross, C. G. Visuospatial properties of ventral premotor cortex. *J. Neurophysiol.* **77**, 2268–2292 (1997).

27. Holmes, N. P. Does tool use extend peripersonal space? A review and re-analysis. *Exp. Brain Res.* **218**, 273–282 (2012).

28. Iriki, A., Tanaka, M. & Iwamura, Y. Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport* **7**, 2325–2330 (1996).

29. Graziano, M. *The Spaces between Us: A Story of Neuroscience, Evolution, and Human Nature* (Oxford Univ. Press, 2017).

30. Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H. & Acuna, C. Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *J. Neurophysiol.* **38**, 871–908 (1975).

31. Barreto, A. et al. Transfer in deep reinforcement learning using successor features and generalised policy improvement. In *35th International Conference on Machine Learning*, Vol. 2, 844–853 (ICML, 2018).

32. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).

33. Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* **33**, 269–298 (2010).

34. Gigliotti, M. F., Soares Coelho, P., Coutinho, J. & Coello, Y. Peripersonal space in social context is modulated by action reward, but differently in males and females. *Psychol. Res.* **85**, 181–194 (2021).

35. Bertoni, T. et al. The self and the Bayesian brain: testing probabilistic models of body ownership through a self-localization task. *Cortex* **167**, 247–272 (2023).

36. Friston, K., Sajid, N., Heins, C., Pavliotis, G. A. & Parr, T. The free energy principle made simpler but not too simple. *Physics Rep.* **1024**, 1–29 (2022).

37. Wolpert, D. M., Goodbody, S. J. & Husain, M. Maintaining internal representations: the role of the human superior parietal lobe. *Nat. Neurosci.* **1**, 529–533 (1998).

38. Medendorp, W. P. & Heed, T. State estimation in posterior parietal cortex: distinct poles of environmental and bodily states. *Prog. Neurobiol.* https://doi.org/10.1016/j.pneurobio.2019.101691 (2019).

39. Bufacchi, R. J., Battaglia-Mayer, A., Iannetti, G. D. & Caminiti, R. Cortico-spinal modularity in the parieto-frontal system: a new perspective on action control. *Prog. Neurobiol.* **231**, 102537 (2023).

40. Arber, S. & Costa, R. M. Networking brainstem and basal ganglia circuits for movement. *Nat. Rev. Neurosci.* **23**, 342–360 (2022).

41. Burgess, N. Spatial memory: how egocentric and allocentric combine. *Trends Cogn. Sci.* **10**, 551–557 (2006).

42. Melzack, R. & Wall, P. D. *The Challenge of Pain* (Penguin, 1988).

43. Seymour, B. et al. Temporal difference models describe higher-order learning in humans. *Nature* **429**, 664–667 (2004).

44. Baliki, M. N. et al. Corticostriatal functional connectivity predicts transition to chronic back pain. *Nat. Neurosci.* **15**, 1117–1119 (2012).

45. Dong, W. K., Chudler, E. H., Sugiyama, K., Roberts, V. J. & Hayashi, T. Somatosensory, multisensory, and task-related neurons in cortical area 7b (PF) of unanesthetized monkeys. *J. Neurophysiol.* **72**, 542–564 (1994).

46. Ronga, I. et al. Seeming confines: electrophysiological evidence of peripersonal space remapping following tool-use in humans. *Cortex* **144**, 133–150 (2021).

47. Quinlan, D. J. & Culham, J. C. fMRI reveals a preference for near viewing in the human parieto-occipital cortex. *Neuroimage* **36**, 167–187 (2007).

48. Holt, D. J. et al. Neural correlates of personal space intrusion. *J. Neurosci.* **34**, 4123–4134 (2014).

49. Ferri, F., Tajadura-Jiménez, A., Väljamäe, A., Vastano, R. & Costantini, M. Emotion-inducing approaching sounds shape the boundaries of multisensory peripersonal space. *Neuropsychologia* **70**, 468–475 (2015).

50. Taffou, M. & Viaud-Delmon, I. Cynophobic fear adaptively extends peri-personal space. *Front. Psychiatry* **5**, 3–9 (2014).

Methods

Tabular reinforcement learning for calculation of action values Q

Action values Q reported in this article were calculated either by ANNs (described below and in Supplementary Methods 3.2), or through tabular learning. Tabular learning has, under specific circumstances, the potential to speed up the Q-value calculations.

To simplify the tabular calculation of action values, we: (1) assumed that objects follow roughly linear trajectories, allowing objects only to move either forward or backward relative to the agent (that is, the y axis in 2D, which is the anteroposterior axis when fitting empirical data in 3D) within one simulation; (2) calculated action value separately for approaching and receding objects; and (3) calculated action value separately for each body part, in body-part-centered coordinates (unless stated otherwise).

These assumptions ensured that no object's position could be re-visited before re-starting a simulation. This allowed us to loop through the moving dimension as an outer loop and calculate Q only once for each possible object position, without having to consider the effects of object movement in the other (one or two) dimensions. Therefore, we were able to update Q fully with a learning rate of 1 (rather than <1), thereby substantially speeding up the calculations and making tabular learning feasible.

We calculated the body-part-centered value fields using pseudocode shown in Supplementary Methods 3.1. Briefly, starting from the space behind each body part, we looped through depth, width and height and updated, for each action, the Q values in each voxel as a function of the voxels that an object could reach in one time step (through object movement or actions available to the agent, or both). In the case of environments where object dynamics have some degree of uncertainty along the y axis, we repeated this process 15× to account for probabilistic world dynamics and the possibility for a state to be re-visited multiple times. For the 2D case, only 1 voxel was used in the z-dimension and the procedure was otherwise identical.

ANN models

Base ANN model. The agent's environment was a grid world of 13×14 blocks (width (x) × depth (y)), with recurring infinite boundary conditions on the left and right edges. In this world, an ANN controlled the x position of its 'limb', located at $y = 3$. At each time step, the ANN could either keep the limb still or move it left or right. Simultaneously, an object moved in the world. When the object came into contact with the limb, the agent received a positive reward (+2).

The object spawned at a random x coordinate at the edge of the world, opposite the agent ($y = 14$) and moved toward the agent's side at a speed of one block per time step. In addition, in the same time step, the object had a 50% chance to additionally move one block left, right, up or down—a process simulating kinematic noise and perceptual uncertainty. The kinematic noise was therefore different at every time step. After an object reached the agent's side of the grid world, a new object was spawned at the opposite end. Moving left or right had a cost of −0.001 (that is, negative reward), to disincentivize random movement, whereas staying still did not entail any reward (unless the object touched the limb).

The ANN received world information through 'proprioceptive' input, reflecting the limb's x position and 'visual' input reflecting the object's x and y positions. The network consisted entirely of artificial neurons (units) with a hyperbolic tangent (*Tanh*) transfer function, except for the final layer, which was composed of pure linear units. In Extended Data Fig. 3, we also considered networks composed of other unit types. The network output a Q value for each possible action. Q values were learned through Q learning with experience replay. Each simulation lasted 4,000,000 time steps, each of which was stored as a state transition. At the end of the simulation, the network was further trained on stored state transitions to ensure near-optimal fitting of the value function, using 100 batches of 10,000 time steps each. The reason for using such a long training period was that the scope of this article was not to create an algorithm that optimally learned particular Q values. Therefore, the parameters used to learn Q values were not rigorously optimized with respect to computational speed or performance. The used parameters provided a balance between satisfactory performance and time invested in parameter tuning.

Model variations. Most of the in silico experiments described in the main text entail some variations to the base model environment described above. The exact variations are detailed in the description of each experiment, in Supplementary Methods 3.2.1–3.2.8, as are the statistical tests applied to each model. Here we list only the parameters that were varied:

(1) Number of limbs. In some environments, we allowed the model to control an additional limb, placed one block below the base limb (that is at $y = 2$). This limb could move left and right independently of the first limb.

(2) Object velocity. In some environments, objects with different velocity could spawn. Each time an object was spawned, its y velocity was set randomly between 1 and 3 (blocks per time step). Similarly, its x velocity was set randomly between −2 and +2 (where negative and positive indicate moving left and right, respectively). Object velocity only changed when a new object was spawned, and was applied in addition to the kinematic noise, which instead changed at every time step.

(3) Input time steps. In the environments where the objects could have different velocities, the network was provided with a 'memory' input: proprioceptive and visual information from the preceding time step. This allowed the network to infer object velocity.

(4) Reward offered by objects. In some environments, the reward consequent to contact with an object was set to either −2 or +4 instead of the +2 of the base model. Note that throughout the manuscript we use the term 'reward' in the most general sense. Thus, the agent can receive a negative reward, which can be understood as a punishment.

(5) Presence of a 'tool'. In some environments, we provided the agent with a 'tool', the effective part of which was located four blocks above the limb and measured 1×1. The tool moved when the limb moved. When an object came into contact with the tool tip (that is, the effective part of the tool), the agent experienced reward as if its limb had contacted the object.

(6) Number of objects in the environment. In some environments, two objects were simultaneously present. In such environments, we set the rewards to +2 for one object and −2 for the other, thus making the objects either 'goals' or 'threats', respectively. The network received separate inputs specifying the x and y coordinates of each object.

An egocentric value map to solve new tasks

To demonstrate how an egocentric map can be used to solve new tasks, we first created a map consisting of two successor features (Glossary, Supplementary Information 1) and hence two reward configurations: goals and threats (Fig. 4b). We used the tabular learning approach described in the pseudocode in Supplementary Methods 3.3.

Next, we created Q values for the four new tasks. We used the pseudocode in Supplementary Methods 3.1 to generate the Q maps for new tasks 1, 2 and 3 ('Only receive reward if not moving on previous time step', 'Avoid a wide threat', and 'Allow the goal to pass on right side', respectively). More specifically, for new task 1, we set the reward to 1 only if the agent chose the action 'stay still' just before contact was made. For new task 2, we set the reward to −1 if the object was anywhere within 1 block (in the x direction) from the limb. For new task 3, we set

the reward to 1 only when the object was 2 blocks to the right of the limb. For new task 4 ('The object reward depends on location of first appearance'), we interleaved the optimal Q values for threat and reward in a row-by-row fashion.

Finally, for each new task, we calculated the optimal policies for each state and the optimal weights for the successor features, by minimizing the squared error between the reconstructed Q values and the task-optimal Q values.

Empirical data fitting

In this section, we describe the overall approach used to fit empirical data from multiple experiments using an egocentric map. For the specific details of each experiment, see Supplementary Methods 3.4.1.

Empirical data. We extracted mean values and standard errors from 23 experiments published across 10 articles, encompassing 426 (180 + 45 + 148 + 20 + 19 + 19 + 15 + 8 + 8) human participants and 3 macaques. In Supplementary Methods 3.4.1, we describe in detail which data was extracted from each article.

Model design. We created body-part-centered Q fields for moving the hand, trunk and head in all six directions (that is, for any possible movement) and fitted the combined map $\bar{\Psi}$ to 23 previously published experimental results. The base model design for each individual experiment is described here. When we made changes to more accurately reflect certain experimental conditions, these are described in Supplementary Methods 3.4.1.1–3.4.1.8.

World setup. To create an approximate egocentric map χ^2_{obs}, we simulated an area of space around the upper body. Relative to the center of the chest, this area spanned from −50 cm to +250 cm along the anteroposterior (y) axis, −60 to +90 cm along the rostrocaudal (z) axis and −50 to +80 cm along the mediolateral (x) axis, with a $5 \times 5 \times 5\,cm^3$ voxel resolution.

Within this space, we defined surfaces of each of the three body parts as separate sets of voxels that lead to reward on contact with objects. The surfaces of each body part are shown in Supplementary Fig. 2a–g.

Stimulus dynamics. We assumed that objects near the body follow roughly linear trajectories, with two possible object velocities: +20 and −20 cm s^{-1} along the anteroposterior axis. We used time steps of 1 s. We also gave the object some nondeterministic movement: every time step it could move 1 voxel (5 cm) away from its deterministic location in each direction (we used a gaussian with $\sigma = 5$ cm as spatial displacement probability).

Action specification. We allowed the head and the trunk to move at a speed of 1 voxel s^{-1} (~5 cm s^{-1}) and the hand at up to 5 voxels s^{-1} (~25 cm s^{-1}), in any direction. These parameters approximate the speed of movements naturally chosen during a long, seated experiment. Nevertheless, our results are robust to changes in assumed movement speeds.

Calculating Q values. We calculated tabular Q values in 3D as described in Methods above, and in Supplementary Methods 3.1, separately for +1 and −1 rewards on contact.

Creating approximate 3D egocentric value maps. For each experimental condition, we created an approximate egocentric map $P(\chi^2_{obs} \geq \chi^2 | H_0)$; as the set of action values related to moving three individual body parts (head, torso, hand) during tasks entailing reward and punishment on contact (Fig. 4b) (that is, $\bar{\Psi}^\pi$ where $\bar{\psi}^\pi = \{\bar{\psi}^\pi_{torso}, \bar{\psi}^\pi_{head}, \bar{\psi}^\pi_{hand}\}$ and $\bar{\psi}^\pi_{body-part}$ is a collection of approximate Q values for that body part, for example, $\bar{\psi}^\pi_{hand} = [\bar{Q}^\pi_{hand\,up}, \bar{Q}^\pi_{hand\,down}, \cdots, \bar{Q}^\pi_{hand\,stay}]^T$).

We next transformed the egocentric map into approximations of the data $\hat{\mathbf{d}}$. First, we found the maximal absolute action value for each body part and reward magnitude. Second, we averaged this absolute action value across all body parts. This gave us one value S for the overall egocentric map activity of each experimental condition:

$$S_{exp\,cond}(s,a) = \frac{1}{N'} \sum_{body-part=1}^{N'} \max_{task} \max_a \left| \bar{\psi}^{\pi_{task}}_{body-part,exp\,cond}(s,a) \right|$$

Note that calculating the maximal Q value across actions and policies also occurs when selecting actions as we described for the recycling of the egocentric map for new tasks[31] and it can also be conceptualized as the competition between affordances observed in real brains[33].

Furthermore, taking the absolute value is mathematically justified because the Q value for the task 'reward' on the policy 'punishment' is simply the negative of the optimal Q value for the task 'punishment' (Fig. 4c). Similarly, the Q value for the task 'punishment' on the policy 'reward' is the negative of the optimal Q value for the task 'reward'. Therefore, the absolute prevents us from having to explicitly calculate these redundant Q values.

Fitting egocentric maps to empirical data. For each experimental condition of the empirical data d, we extracted the egocentric map magnitude $S_{exp\,cond}$ as described above. We then linearly fitted the map magnitude to the data: we multiplied S by a separate parameter A for each different grouping of body parts and response types and added an offset for each experimental session:

$$\hat{d}_{exp\,cond} = A_{body-part\,group,response}\, S_{exp\,cond} + B_{exp\,sess}$$

We optimized A and B by minimizing the squared error between the linearly transformed egocentric map $\hat{d}$ and the empirical data $\mathbf{d}$.

Statistics. To statistically assess the goodness of fit, we first calculated the χ^2 statistic, a measure of total error (Supplementary Fig. 2j, top left):

$$\chi^2 = \sum_{i=1}^{N} \frac{(d_i - \hat{d}_i)^2}{var(d_i)}$$

where d_i is the mean data for a given experimental condition i, $\hat{d}_i$ is the fitted estimate of that mean and $var(d_i)$ is the estimated variance of the mean d_i, that is, the s.e.m.

Next, we compared the χ^2 statistic with a χ^2 distribution with k degrees of freedom, where $k = N - np$, N is the total number of conditions and np the number of parameters used to fit the data. This distribution represents the expected χ^2 distribution if the process generating d is the same as that generating $\hat{d}$.

It follows that the P value obtained from comparing the χ^2 statistic with the χ^2 distribution with k degrees of freedom is a measure of the confidence that the empirical data could have been generated by a process like the one described in our model. In other words, if the P value is very small, we must reject the null hypothesis that our model explains the data. Conversely, if the P value is very large, we must accept that hypothesis as a possibility[21]. For this test, the total number of observations was equal to the total number of fitted empirical conditions: 204. To report confidence intervals for the χ^2 statistic, we bootstrapped the data 10^6 times.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All used empirical data was extracted from publicly available figures[11,14,25–27,46–48]. Generated data can be recreated using the code at

https://github.com/rorybufacchi/EgocentricValueMaps. Source data are provided with this paper.

Code availability

All analyses were performed in MATLAB (2020a and 2022b). All code is available at https://github.com/rorybufacchi/EgocentricValueMaps.

Author contributions

R.J.B. and G.D.I. conceived the project. R.J.B. did the modeling and analysis. R.J.B. and G.D.I. created the figures. R.J.B., R.S., A.M.F., R.C. and G.D.I. planned the analysis. R.J.B. and G.D.I. wrote the original draft. R.J.B., R.S., A.M.F., Y.M., N.L., R.C. and G.D.I. reviewed and edited the MS. Y.M., N.L. and G.D.I. acquired the funds.

Competing interests

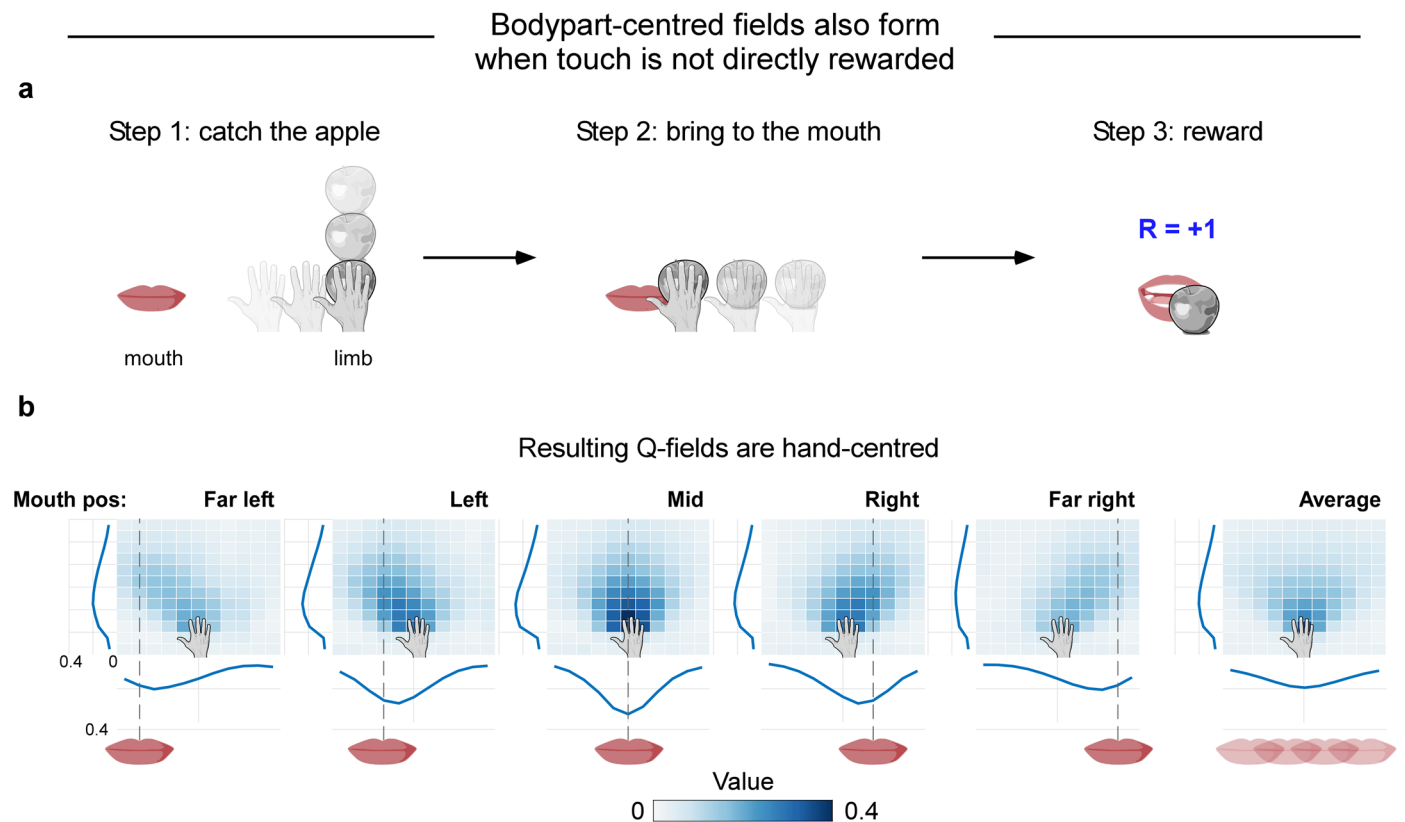The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41593-025-01958-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-025-01958-7.

Correspondence and requests for materials should be addressed to Rory John Bufacchi or Gian Domenico Iannetti.
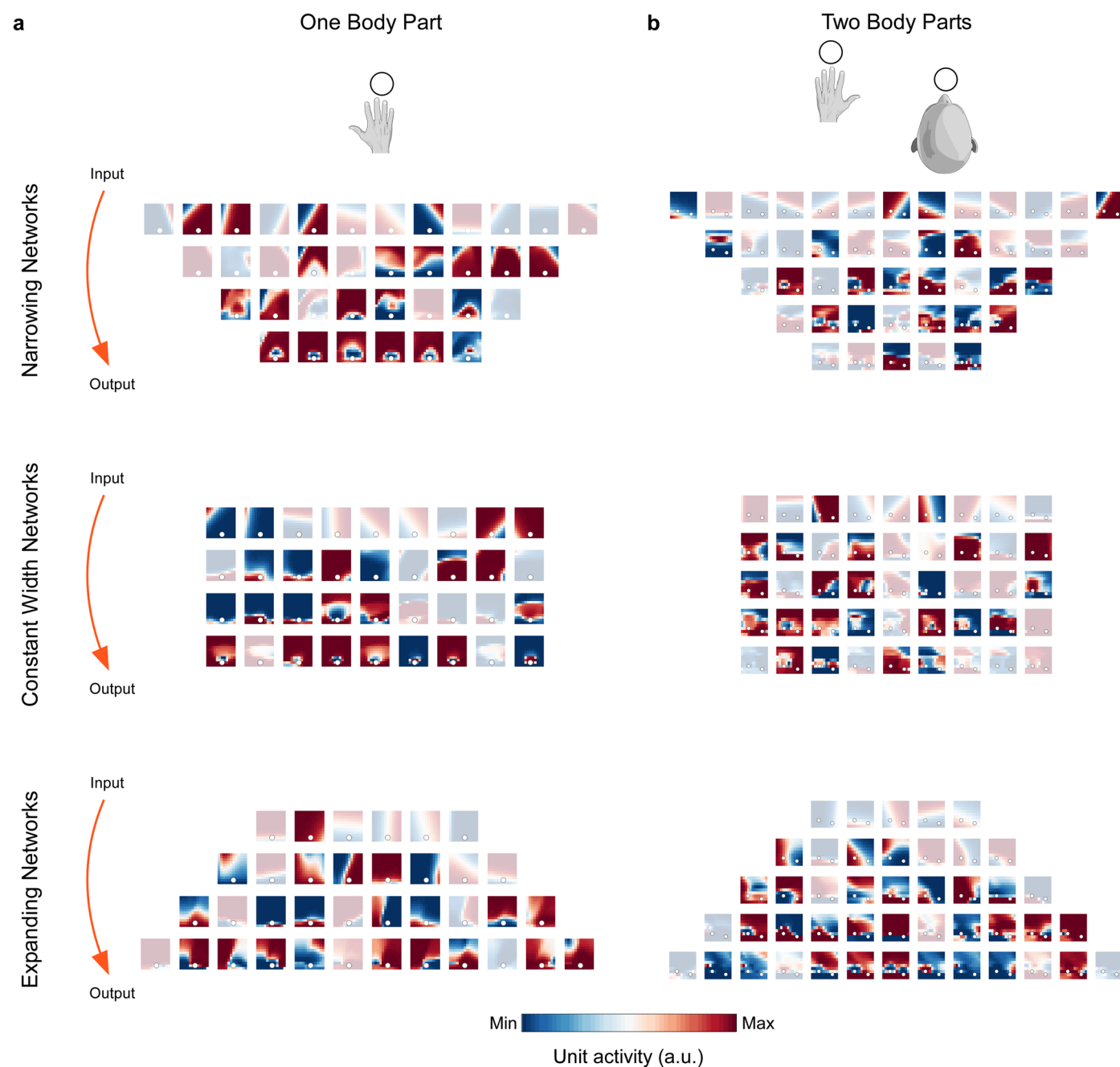
Peer review information *Nature Neuroscience* thanks Andrea Serino and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
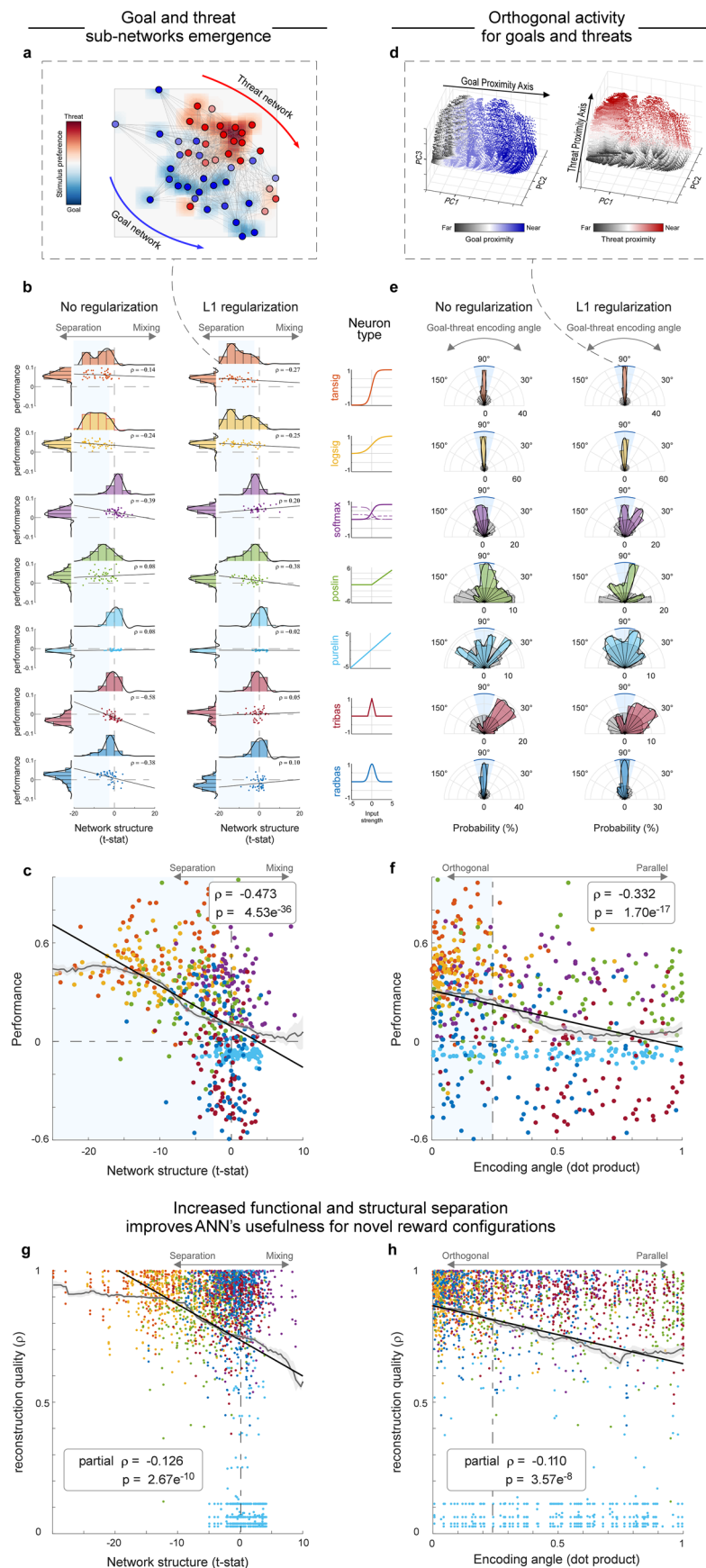
Extended Data Fig. 1 | Bodypart-centred value fields arise even if touch is not directly rewarded. a) In the main text we assumed that touch resulted in immediate reward, for the sake of simplicity. However, bodypart-centred fields can also arise if touch itself is not directly rewarded, but is a prerequisite to obtain reward at a later stage. To demonstrate this, consider the situation in which an agent must first catch an object (here, an apple) with its limb (step 1) before bringing it to its mouth (step 2) and finally obtaining the reward by 'eating' it (step 3). **b)** Value fields as a function of object and mouth position. Each column shows Q-values for a specific mouth position. Right-most column shows the average Q-values across all mouth positions. Line plots to the bottom and left of each heatmap are average Q-values along the y and x axes, respectively. In this environment, value fields still emerge around the limb, because contact is a step along the path to reward. The field shape and magnitude also depend on the position of the mouth, because intercepting the 'apple' when the mouth is near the hand leads to more immediate reward, and is therefore more valuable.

a One Body Part

b Two Body Parts

Narrowing Networks

Constant Width Networks

Expanding Networks

Input

Output

Min [====] Max

Unit activity (a.u.)

Extended Data Fig. 2 | Individual neural response fields in all architectures used for ANN model 1. a) Each colour map shows the activity of a single network unit as a function of the stimulus spatial position relative to the body part (body-part location is indicated as white circles). The three main rows indicate different network architectures. Full-colour plots show units classified as bodypart-centred; greyed-out plots show units not classified as bodypart-centred. Within each ANN, the proportion of bodypart-centred units per layer increased as a function of layer depth. **b)** Same as (**a**), but for ANNs that controlled two body parts instead of 1.
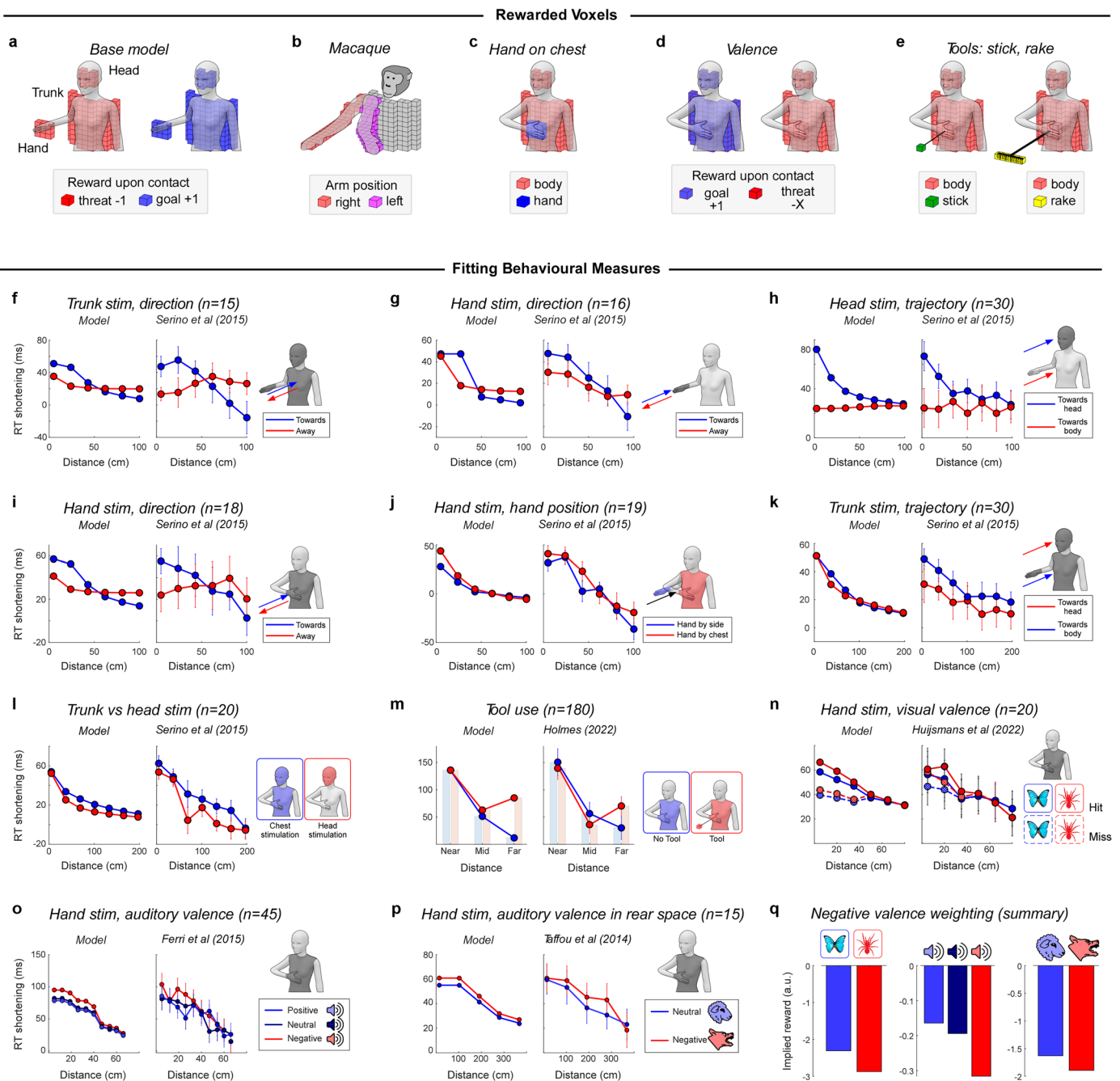
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Modularity emerges only when multiple tasks are learned successfully, regardless of artificial neuron type. a) An ANN trained on simultaneous interception and avoidance tasks naturally adopts a modular, task-specialised structure. This example ANN consists of fewer neurons than the modular ANNs reported in the main text (see Fig. 5 and Supplementary Methods 3.2.6). Individual units are classified as threat- or goal-preferring (red and blue, respectively). **b)** Structural modularity in ANNs, split by the neuron types that compose the ANNs (see Supplementary Methods 3.2.6 and Supplementary Results 4.1.2). Rows indicate neuron type, while columns indicate regularization method. The histograms above each line plot show the amount of sub-network structure (indexed by a t-statistic; see Supplementary Methods 3.2.6). Histograms to the left of each line plot show network performance (indexed by expected reward per unit time). Light blue shaded area indicates above-chance sub-network structure. Network types with above-chance performance also had above-chance sub-network structure, regardless of regularization method. **c)** Across neuron types and regularization methods, the amount of sub-network structure (x-axis) predicted the performance of the network (y-axis). Colours indicate neuron types. Dashed line indicates zero. Light blue shaded area indicates above-chance sub-network structure. Rho and p-value result from a Pearson's correlation test. **d)** The same ANN as in (**a**) also naturally encodes task-specific variables in orthogonal spaces. Scatter plots show activity from the first 3 principal components. Information related to goal-proximity (black-to-blue scatter plot) is orthogonal to information about threat-proximity (black-to-red scatter plot). **e)** Orthogonality of goal and threat proximity coding in the same networks described in (**b**). Coloured histograms show the angle between goal- and threat-coding in the 1st 3 PCs. Light blue areas indicate angles within 15° of 90°. Grey histograms indicate null distributions (1000 permutations). **f)** The degree of task-encoding orthogonality in the first 3 principal components (x-axis) predicted the performance of the network (y-axis). Dashed line indicates the absolute dot product equivalent to a 75° angle difference. Light blue shaded area indicates angles within 15° of 90°. Rho and p-value result from a Pearson correlation test. **g)** Across neuron types and regularization methods, the amount of sub-network structure (x-axis) is predictive of the ability of the network to reconstruct Q-values for novel tasks (y-axis; correlation coefficient between original and reconstructed Q-values; see Supplementary Methods 3.2.6). The novel tasks were the same as those described in Fig. 4a of the main text. Rho and p-value result from a Partial correlation test, which factors out the effects of performance (reward per unit time) and task-encoding orthogonality. **h)** Across neuron types and regularization methods, the amount of task-encoding orthogonality (x-axis) is also predictive of the ability of the network to reconstruct Q-values for novel tasks (y-axis). Colours indicate neuron types. Rho and p-value result from a Partial correlation test, which factors out the effects of performance (reward per unit time) and network structure.
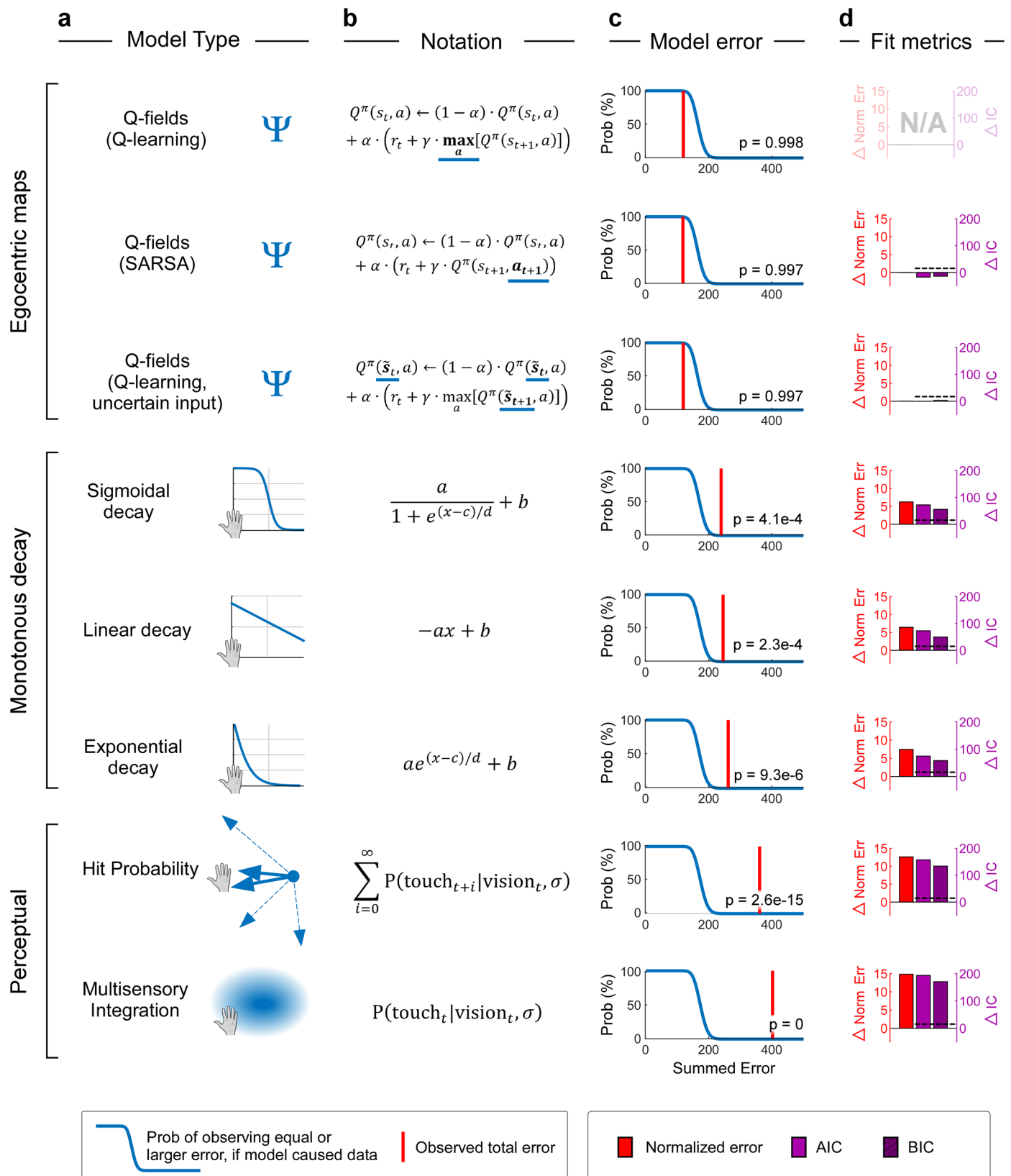
Extended Data Fig. 4 | Fitting egocentric maps to empirical data, part 2: setup and behavioural measures. a) To create a model egocentric map around the upper body, we created a default set of voxels on the surface of the hand, head, and trunk, which offered reward upon contact (coloured cubes). Experiment-dependent variations of these rewarded voxels are shown in b-e. For extended data fitting methodology, see Methods 'Empirical data fitting' and Supplementary Methods 3.4.1. **b)** To model arm-centred peripersonal neurons in macaques (Fig. 7a), we created two sets of voxels, each simulating one of the two arm positions during the experiment. **c)** In a subset of human experiments (**I,j,l,m,n,o,p**; Fig. 7e–h), the hand was held in front of the chest, instead of to the side. **d)** To model experiments with variable stimulus valence (**n,o,p**), we additionally optimised the negative reward offered by contact (Methods, 'Empirical data fitting' & Supplementary Methods 3.4.1). The optimised reward values implied by the stimuli of differing valence are shown in **q. e)** To model tool use (**m**; Fig.7h), we also rewarded contact with the tip of a tool. The two tool-related experiments used a stick and a rake, which we respectively modelled with

1 (green) or 5 (yellow) voxels. **f-p)** Empirical data (right in each panel) and model fits (left in each panel) from experiments in which behavioural peripersonal measures were collected. For data and fits of neural measures, see Fig. 7c–h. Colours indicate experimental conditions. Shaded surfaces on figurines show body parts included in the egocentric map. Error bars show SEM. For a detailed description of each fitted experiment, see Supplementary Methods 3.4.1. **f)** See Supplementary Methods 3.4.1.6 for details. **g)** See Supplementary Methods 3.4.1.6 for details. **h)** See Supplementary Methods 3.4.1.6 for details. **i)** See Supplementary Methods 3.4.1.6 for details. **j)** See Supplementary Methods 3.4.1.6 for details. **k)** See Supplementary Methods 3.4.1.6 for details. **l)** See Supplementary Methods 3.4.1.6 for details. **m)** See Supplementary Methods 3.4.1.7 for details. **n)** See Supplementary Methods 3.4.1.8 for details. **o)** See Supplementary Methods 3.4.1.9 for details. **p)** See Supplementary Methods 3.4.1.10 for details. **q)** Best-fitting negative reward magnitudes for the stimuli of differing valence from experiments displayed in **n**,**o**, and **p**.

Only egocentric map models fit the empirical data

Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Comparison to alternative models. a) We fitted three model families to an empirical dataset combined from 23 published experiments across 10 different research groups. The 'Egocentric maps' family (top three models) is the main topic of this paper, and the 'Q-fields (Q-learning)' model is the specific model described in the main text. The 'Monotonous decay' family (middle three models) contains purely empirical models that attempt to describe the data, but without having a theoretical a-priori reason for being appropriate models. The 'Perceptual models' (bottom two models) have previously been used to fit individual datasets, and are largely based on the notion that peripersonal fields arise due to uncertainty in visual and auditory input, while estimating the probability that the source of the visual input makes contact with the body. We calculated all quantities for each $5 \times 5 \times 5$cm voxel around the upper body, and fit them to the data with at least the same number of parameters as we used for Q-value fitting. The exponential and linear falloffs required two additional parameters, to fit the size and slope of the receptive fields. We parametrised the uncertainty necessary for the perceptual models by taking the same values as reported in[17,20]. **b)** Mathematical description of each model. For the 'Egocentric maps' models family, we display the update equation for the Q values, and underline the part of the equation that is unique to each of the three models. **c)** Summed error when each model is fitted to the empirical data (red line). The error of all models other than egocentric maps is larger than the error expected from a model that appropriately describes the generative mechanism behind the data (blue distribution). Models with a summed error corresponding to $p < 0.05$ (using the variant of chi-squared goodness of fit testing described in Methods 'Empirical data fitting') can be confidently rejected as explanations of the data. **d)** Metrics of fit quality relative to the 'Q-fields (Q-learning)' model. The normalised error (left y axis, red) is the summed error from (**c**) scaled by the variability of the data. Purple bars show the difference between the AIC and BIC (Akaike and Bayes Information Criterion, respectively) of each model vs the 'Q-fields (Q-learning)' model (right y axis, purple). A difference of >10 for AIC and BIC (indicated by dashed black lines) is commonly taken to indicate that the considered model can be rejected in favour of the reference model.

nature portfolio

Corresponding author(s):   Rory Bufacchi

Last updated by author(s):   Feb 12, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All data collection was performed in-silico on custom-written code in Matlab 2020a and 2022b, using the default neural network toolbox.
Data analysis	All data analysis was performed on custom-written code in Matlab 2020a and 2022b, using the default statistics and neural network toolboxes. We used (sometimes modified) Q-learning and SARSA algorithms to perform reinforcement learning. All code is available at https://github.com/rorybufacchi/EgocentricValueMaps.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All used empirical data was extracted from figures in publicly available journals, and is additionally made available as statistical source data for figures 7 and Extended Data Figure 5 (in excel format). Generated data can be recreated using code the at https://github.com/rorybufacchi/EgocentricValueMaps, and statistical source data required to generate the figures is available for download alongside the manuscript.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](). See also policy information about [sex, gender (identity/presentation), and sexual orientation]() and [race, ethnicity and racism]().

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf]()

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The study was entirely in-silico. While no explicit sample size calculations were performed a-priori, in order to ensure that ANN results were not due to flukes of network architecture, all statistics were performed on at least three, separately trained networks. Given the exceedingly low p-values encountered, adding more networks would not have changed the results
Data exclusions	Given that the study was entirely computational, no data was excluded.
Replication	Given that the study was entirely computational, it can be be re-run indefinitely. All code is available for anyone to do so
Randomization	Given that the study was entirely computational, no randomization was necessary, other than as a null-distribution for permutation testing
Blinding	Given that the study was entirely computational, blinding was not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology and archaeology
☒	Animals and other organisms
☒	Clinical data
☒	Dual use research of concern
☒	Plants

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A